Cover page

Modern Imperatives:

Essays on Education and Health Policy

ISBN: 978 90 361 0574 3

This book is no. 747 of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Modern Imperatives:
Essays on Education and Health Policy

Moderne Imperatieven:
Essays over Onderwijs- en Gezondheidsbeleid

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Friday, September 27, 2019 at 11:30 hours

by

MAX HARRIS COVENEY

born in Perth, Australia

Erasmus University Rotterdam

Doctorate Committee

Promotors:	Prof. dr. E.K.A van Doorslaer
	Prof. dr. T.G.M Van Ourti
Other members:	Prof. dr. H.D. Webbink
	Dr. T.M. Marreiros Bago d'Uva
	Dr. P.R. Dias
Co-promotor:	Dr. P. García-Gómez

Acknowledgments

It's hard for me to pinpoint the exact moment in late 2013 that I finally decided to apply for a PhD position at Erasmus University. Nevertheless, looking back six years later, this is one decision I can say with certainty that I do not regret. Between then and now I not only produced the thesis that you are now reading, but have come away with countless fond memories, new friendships, and a great deal of professional and personal growth. There are a number of people that I need to thank for making all this possible.

This thesis would not exist if it were not for my supervisors: Tom Van Ourti, Pilar García-Gómez and Eddy van Doorslaer. I was lucky enough to find myself with three supervisors who are not only brilliant academics, but also wonderful people.

Tom, I have learned a lot from the way you approach and conduct research. I hope at least some of your precision and sense of ethics has rubbed off on me. Pilar, thank you for always giving me honest advice, both job related and otherwise. Your encouragement gave me the confidence to further develop my research interests into new areas. Eddy, thank you for your wisdom, experience, and for always keeping my spirits up. You made many long meetings far more enjoyable than they should have been. To all three I would like to express the deepest gratitude for giving me plenty of encouragement and praise even when I probably didn't deserve it, and giving me constructive and gentle criticism when I probably deserved a lot more.

One of the most valuable parts of my PhD experience was being surrounded by so many talented colleagues. Thank you to all the fellow PhDs I met on the 8th floor of the Tinbergen building and elsewhere. Many thanks also to the support staff, the postdocs and professors, and everyone else in the Applied Economics department.

I would especially like to express my gratitude to two people I met during the first months of my PhD: Arash and Matthijs. Arash, I could always count on you to be a calm and cheerful presence even during the most difficult phases of the PhD, in particular during the first year courses. Matthijs, your contagious enthusiasm for research has made working with you on the final chapters of my thesis a real pleasure. Thank you both for being such good friends during these years.

I am also deeply grateful to my wonderful group of friends outside the university, both in the Netherlands and elsewhere. While there is not enough space to thank you all individually, special mention goes to my good friends Juan, Marta, Martin, and João; the fact that I can look back fondly on my last six years in the Netherlands is largely down to you. Will and James, thanks for visiting me in Europe and hosting me without hesitation on my trips back to Australia. I'm happy that our friendship remains strong despite the long absences and timezone differences. Lilly, thank you for being a constant source of love, support, and inspiration. You all gave me a welcome distraction during the more difficult parts of my PhD, and helped remind me of the important things in life.

Finally, and most importantly, I have to acknowledge my family. Claudia, it's been such a pleasure this past year to have been reunited in Rotterdam with my best friend. I have often heard that the final year of a PhD is the most difficult and lonely one. Thanks to your presence, it was definitely my favourite. Mum and Dad, there's no way I can properly express my gratitude towards you in words. I can only say thank you for everything you have given me, and for your unwavering belief and support in everything I do.

Contents

Chapter 1

Introduction

The spheres in which government policies perhaps most obviously present themselves in individuals' everyday lives are those of education and health. In the area of education, vast numbers of policy decisions govern when children are required to start their schooling, what they learn (and don't learn), and innumerable other aspects of their educational experiences. Similarly, health systems in most countries are heavily shaped and supported by governments. Medical licenses determine who can provide healthcare, policy makers decide on the availability and price of medication, who gets access at what cost, and where these costs are borne.

As well as being salient, the policies in these spheres have deeply important and far reaching consequences for individuals themselves. Both the formal curriculum as well as the socialization process that occurs during schooling — the so-called hidden curriculum — are vitally important in shaping individuals and preparing them for adult life. It goes without saying that health, and the policies governing health care, also have important individual consequences. Good health and well-being are not only fundamentally enjoyed in their own right, but are also a prerequisite for taking part in nearly all other aspects of life.

The fundamental importance and the empowering nature of health and education mean that they are usually placed at the forefront of governments' agendas. Indeed, Amartya Sen's "capabilities approach" (Sen, 1979) advocates a prioritization of policies aimed at maximizing individuals' capabilities, and good health and a quality education are often presented as among the most fundamental measures of such capabilities. According to this approach, their delivery should therefore be considered as a basic obligation of modern governments to their citizens.

At the same time, and as economists will be quick to point out, we cannot devote unlimited attention and materials to promoting health and education. Resources are limited, and a euro or dollar spent on one program is a euro or dollar not spent elsewhere. Faced with this scarcity, and also with the fundamental importance of health and education, it is imperative that all policies in these spheres are informed by evidence coming from careful and detailed scientific research. The chapters of this thesis present the work that I have undertaken (with my co-authors), all with the basic theme of providing such evidence in an effort to inform education and health policy.

My focus has been on two distinct and important subtopics under the broad umbrella of health and education. The first of these, addressed in Chapters 2 and 3, is concerned with a crucial part of students' educational experiences: their relationship with their peers. Indeed, the friends, acquaintances, and classmates encountered during school arguably leave the longest lasting impressions from our primary and secondary educational experiences. In the previous decades a substantial economic literature has emerged studying the effect of these peers on a range of outcomes. Peer composition has been shown to influence a person's academic performance, behaviours inside and outside of school, and current and future attitudes and beliefs.

The second of these subtopics, addressed in Chapters 4 and 5, is the phenomenon of income-related health inequalities (IRHI). Such inequalities describe the pervasive difference in health by income whereby, in almost every context including the European one, richer individuals live longer and healthier lives than poorer individuals. Inequalities have become one of the most contentious and widely discussed political issues today. Because of their importance for social cohesion and solidarity, as well as basic ethical and fairness concerns, health inequalities are an increasingly important part of this debate.

Taken together, education and health address some of the most pressing priorities that jurisdictions face. Thus, there is a moral onus not only on governments to commit to organize enabling circumstances for good health and education, but also for decision makers to design and refine policies in these public spheres wisely and based on scientific evidence. The chapters contained in this thesis, based on rigorous economic analysis, go some way to providing such evidence.

Chapter 2

In the study of the effects of peers in education the question that has received most attention is a simple one: does having smarter peers in school result in better grades? As well as being interesting from a purely scientific perspective, this question has important policy implications. Depending on the answers, it may be possible to improve educational outcomes simply by rearranging students between classrooms. For instance, if the presence of smart students in a classroom helps other smart students, but hurts low achieving students, then gathering all smart students in their own classroom will be academically beneficial to everyone.

Motivated by this, a large number of studies have now provided evidence of small yet meaningful positive effects on individuals' grades of the academic ability of their peers in a wide range of educational contexts. Although the specifics do vary somewhat, it can be said that - on average - having smarter peers seems to increase students' performances (Sacerdote, 2011). However, despite the vast literature, researchers have not yet reached the stage at which they can advise policy. Attempts to harness and reproduce educational peer effects via interventions have failed, even resulting in detrimental effects (see Carrell et al. (2013) for an example).

How can the literature move forward to realizing the goal of implementing reliable and predictable ability peer effect policies? Perhaps the largest obstacle is that the channels through which academic peer effects occur in the classroom are not well understood. Insights into this black box would allow policy makers to craft interventions focussing on the crucial mechanisms underlying peer effects, thereby maximizing their probability of success. In this chapter, we aim to pin down these mechanisms.

We make a distinction between two broad and exhaustive channels that have been suggested by the existing literature. The first is peer effects occurring due to social interaction between peers (peer-to-peer teaching, collaborative studying, etc.), the second is peer effects occurring through the classroom environment, independent of social interaction between students (a "superstar" student posing good questions, a disruption-free classroom environment, teachers responding to the ability composition of the class, etc.).

We test for the importance of each channel by exploiting the structure of year-long tutorial groups at a large European university. Upon arrival, all first year students are randomly allocated to not only one of these tutorial classrooms – within which exercises and assignments are completed - but also to one of two subgroups within their classroom. These subgroups meet frequently in the first months of university, with the aim of creating bonds and friendships to support students during their transition to university life. In essence, this system ensures not only that students are randomly allocated to a classroom, but also that their group of likely friends within this classroom is also randomly allocated.

Taking advantage of this, we examine how the ability (measured by high school grades) of a student's peers within each of these groups influences their subsequent academic performance in the first year. If only the ability of a student's classroom friend matters for their grades, then this would imply that peer effects work through social interaction. On the other hand, if the ability of relative strangers in the classroom has an influence on student's grades, this would suggest that classroom-level effect are at work. Our results find a role only for the social interaction channel of peer effects.

Chapter 3

The effects of university peers are not isolated to grades. Friends can be influential in much more fundamental ways: for instance, by changing our attitudes, beliefs and even how we view the world (Sacerdote, 2014).

These more radical effects of peers have been central to recent debates surrounding the increased admission of foreign students. Especially in the Netherlands, the vast number of international students admitted to universities has become controversial; some claim that internationals take resources and university places away from native students. In response, proponents of the trend towards internationalization cite the positive effect that interaction with foreigners can have on locals, effects which are very much in the spirit of the broader peer effects mentioned above. In essence, contact with foreigners is said to make native students more rounded and globally minded individuals.

However, despite the debate in the Netherlands and elsewhere, there is little evidence on the actual friendship patterns of native and foreign students, nor evidence on the degree to which such friendships can be encouraged by universities themselves. If there is little meaningful contact between these groups, then many of the purported benefits of having an internationally diverse campus will be absent.

In an effort to better inform this important educational policy debate, this chapter investigates the actual occurrences of friendships between native and foreign students at a large European university. To do so, we use a novel technique to elicit friendship ties. Students at this university must register to study groups with their fellow students. Given that friends prefer to be in study groups together, we use students' choice of study groups as a signal from which we uncover the actual ties between classmates. Our results point to a notable degree of segregation between native and foreign students.

We go on to investigate the degree to which universities may be able to encourage native-foreign friendships by forcing students to share a close personal space. We study how forcing a native and foreign student into the same tutorial classroom for a full academic year affects their subsequent probability of friendship. Our results suggest that while forced exposure may promote native-foreign friendship and interaction, this depends heavily on the characteristics of both students.

Chapter 4

Disparities in health by income are pervasive and persistent. Drawing on the large, multi-disciplinary literature concerned with measuring and understanding these disparities, one observes - across virtually all contexts, measures of health, and socio-economic status - sizeable socio-economic differences in health in favour of the richer, wealthier and better educated individuals (for instance see Van Doorslaer et al. (1997) and Mackenbach et al. (2008)).

The attention to these inequalities is based on ethical concerns; many find the concentration of good health amongst the rich to be objectionable. The reduction of these inequalities has thus long been a policy goal in many countries, and is considered a vital part of progress towards achieving greater social inclusion and cohesion in the EU. To the extent that society favours a more equitable distribution of health, these inequalities must be considered when designing and implementing policies that may influence this distribution.

The Great Recession has led to a renewed focus on health inequalities. In particular, EU policy makers have expressed concerns that socio-economic disparities in health may have been exacerbated, given that the negative effects of economic crises tend to disproportionately affect the most vulnerable members of society (European Commission, 2009).

Are these concerns warranted? Despite the attention of governments and policy makers, and the justified alarm about the potential for deepened disparities during this period, evidence on how IRHI actually evolved during the Great Recession is missing. In this chapter, my co-authors and I seek to shed some light this evolution.

In order to investigate the link between health inequalities and the Great Recession, we focus on Spain between 2004 and 2012. Spain was one of the European countries to experience the most severe consequences of the recession, and therefore serves as an interesting case study. Using a concentration index to calculate these inequalities over time, we reach a perhaps surprising finding: while the inequalities in health by income were rising before 2008, they subsequently reduced during the crisis.

To explain this counter-intuitive finding, we apply a decomposition method in order to shed further light on the source of these trends. We find that the unequal effects of the crisis by age, and the fact that the elderly's incomes were largely protected due to the "sticky" nature of pensions, were crucial to the evolution of IRHI. The trends appear to be the logical consequence of the income-reducing effects of the crisis being concentrated amongst the youngest – and therefore healthiest – groups.

Chapter 5

While the insights from the previous chapter into how and why IRHI evolved during the economic crises are informative for the Spanish context, the findings need not hold for other EU countries. The impacts of the crisis unfolded in different ways and degrees between countries. Greece, for instance, was obliged to implement harsh austerity measures as a result of the crisis, while effects in countries like Austria were much smaller in comparison. Institutions, welfare policies and initial economic conditions also differed across Europe, which could all have implications for the evolution of IRHI during this period.

Motivated by this, chapter 5 continues our investigation into how IRHI responds to economic conditions by expanding the countries and time frame under investigation. Specifically, we compute IRHI trends between 2004 and 2013 for Spain, Portugal, Italy, France, Belgium, Austria and Greece. These trends reveal a distinct pattern separating the so-called “crisis countries”, countries harshly affected by the crisis, and other European countries, where effects were less severe. In general, the former countries experienced a drop in IRHI post 2008, while the trends in the latter countries continued on their pre-crisis trajectory.

To explore these trends, we develop and apply a novel decomposition method based on the findings from the previous chapter suggesting that government transfers play an important role in the evolution of IRHI. Specifically, our new decomposition seeks to isolate the separate roles of market incomes (e.g. wages) and government transfer (e.g. pensions) in determining changes in IRHI. The variation across countries in the effects of the crisis, the existing government policies, and the responses to the crisis allows a deeper look at how IRHI is affected by economic conditions. Our conclusions point to a pro-cyclical pattern of IRHI that is primarily driven by the interplay between market and government transfer income and their distribution across age groups.

Chapter 2

What Drives Ability Peer Effects?

Joint work with Matthijs Oosterveen

2.1 Introduction

Economists' ongoing interest in classroom peer effects is not hard to justify; simply by reorganizing peer groups, and without additional resources, it may be possible to increase aggregate student performance. Taking into account important methodological advances (Manski, 1993), the past decade of empirical research includes many well-identified studies in primary, secondary, and tertiary education (Sacerdote, 2014). While these studies have to a large extent confirmed the existence of small peer effects in the classroom, little to no credible evidence exists on the mechanisms through which these effects operate. For instance, it remains unclear whether students benefit from better peers because of social interaction with these peers, or because the quality of teacher instruction improves in a classroom with better students, or through another potential mechanism.

This paper is the first to exploit random group assignment to empirically test between two exhaustive and policy-relevant channels driving ability peer effects. Based on the current literature, we distinguish between the following two channels; social proximity and classroom-level effects. Social proximity relates to the degree of familiarity between classroom peers (Foster, 2006), and this channel captures spillovers that arise due to friendship, bonding, and student-to-student interaction between classroom peers. Classroom-level effects capture spillovers that stem from the classroom environment, which are independent of the social proximity between students, *e.g.* teacher response to the ability composition of the classroom. The context in which we study these two channels is the first year of an economics undergraduate program across six cohorts at a large public university in the Netherlands.

We exploit the institutional manipulation of the social proximity between students and their classroom peers. Students are randomly assigned to a tutorial group of approximately 26 students and one of two subgroups of 13 students within their tutorial group. The university encourages interaction, bonding, and friendship within, and not between, these subgroups during the first weeks of the academic year via several informal meetings. From the perspective of one student, the *close* peers are the subset of their tutorial peers with whom social proximity is encouraged, whereas their *distant* peers belong to the adjacent subset with whom social proximity is not encouraged. For each student, her close and distant peers together form her tutorial group whom she follows classes with throughout the first year. By exploiting the differences between these two types of peers, we are able to disentangle the two broad mechanisms driving ability peer effects. We use high school GPA - which includes the nationwide final exams before entering university - as a pre-treatment indicator of own and peer ability. This allows us to avoid problems related to reflection and common shocks. Moreover, Stinebrickner and Stinebrickner (2006) show that high school GPA (relative to *e.g.* university and college entrance exams) is a comprehensive measure of peer quality.

Exploiting the novel within-classroom random assignment we find that peer effects are solely driven by a student's close peers; the subset of peers within the classroom with whom students are socially proximate. We find no role for distant peers. This implies that meaningful social interaction drives peer effects, whereas classroom-level effects are unimportant. The point estimate from our linear model implies that a one standard deviation increase in close peer GPA causes student performance to increase with 0.026 standard deviations. Using student evaluations we provide suggestive evidence that students with better close peers change their study behavior by substituting lecture attendance for collaborative self-study with their close peers at university. Examining heterogeneity in spillovers by ability, we find that high and low ability students benefit (suffer) from social proximity with high (low) ability close peers. These spillovers, however, diminish over time, and are completely absent by the end of the first year.

Having shown that peer effects arise due to social proximity, the evolution of the social proximity between students and their assigned close peers, and the degree to which new friendship are formed, is of major importance to group assignment policies. We study how students cluster by daily tutorial attendance in first year and find some evidence that the social proximity between assigned close peers gradually diminishes. Analysing tutorial choice in second year we confirm that students largely sort themselves out of their close peer groups. We also show that they sort into new self-chosen peer groups, which are based on shared characteristics such as gender and ethnicity. We do not find evidence that students sort on ability, though our estimates suggest this could be academically beneficial. Overall, we believe this sorting behaviour shows that students have strong preferences dictating with whom they become socially proximate. The erosion of social proximity between assigned close peers provides an intuitive explanation for the short-lived spillovers on student performance, though we cannot provide causal evidence to confirm this intuition.

Our study has three main implications for group assignment policies aiming to exploit spillovers. First, our results suggest that such policies should focus on fostering social proximity within student groups. As it stands, attempts to implement alternative group assignment policies using estimates of peer effects under one particular assignment policy do not lead to predictable results. A well-known example of this is the study by Carrell et al. (2013), in which the authors use credible estimates of spillovers to construct "optimal" peer groups at the United States Air Force Academy. They find that low ability students whom they intended to help with this group assignment policy actually performed worse than untreated low ability students.[1] The importance of social proximity and the absence of classroom-level effects implies that it may be insufficient to simply place students together in a classroom. Our results suggest group assignment policies could be more successful if social proximity within peer groups was fostered. Additionally, such fostering could result in larger spillovers than those previously observed. Our estimated spillovers in the linear-in-means model are more than twice the size of those found in very similar contexts, where manipulation of social proximity is absent (Booij et al., 2017; Feld and Zölitz, 2017).

Second, our results imply that social proximity between diverse assigned peers can indeed be manipulated by a relatively simple intervention, consisting of several informal meetings.[2] However, the persistence of these bonds in the longer run, especially among students of different backgrounds, may be low.

Third, given the importance of social proximity to ability peer effects, our results imply that long-run effects on student performance from group assignment policies may be difficult to sustain. Individuals have strong homophilic preferences, and over time tend to experience diminishing social proximity with their assigned peers as they sort into new peer groups based on these preferences.

With respect to the literature on peer effects more broadly, Sacerdote (2014) highlights the large degree of heterogeneity in the magnitudes of spillovers across the current studies. The findings of this paper may to some extent help explain this heterogeneity. Given that peer effects crucially depend on the degree of social proximity, the study-to-study variation in peer spillovers may partly be explained by the degree that social proximity was present, or perhaps even encouraged.

[1] In Carrell et al. (2009), data based on ability mixing (natural random variation) suggested that low ability students would benefit from being mixed with high ability students, were high ability students would not suffer from being paired with low ability students. Carrell et al. (2013) then create optimal squadrons that consisted of low- and high ability students (bimodal squadrons) and squadrons with middle ability students only (homogeneous squadrons).

[2] The analysis on voluntary sorting shows that a student's close peers are more strongly related to her first-year tutorial attendance and second-year tutorial registration than distant peers.

Our results may also provide some suggestions for the literature on theoretical models of peer effects, which in turn might generate new insights for empirical work. Most of the well-known models of educational peer effects imply that they take place at the classroom level. Lazear (2001) argues that a classroom can be considered as a public good, where one disruptive student may impose negative externalities on *all* students. The taxonomy of models on peer effects by Hoxby and Weingarth (2005) also encapsulates this idea, whereby *e.g.* one superstar student can increase the grades for the rest of the class. Our results imply more nuanced versions of these existing models; a model which focuses on social interaction would more realistically capture the processes driving peer effects in tertiary education.

Apart from their importance for understanding peer effects, the patterns on voluntary sorting behaviour of students also provide a rare insight into how friendship formation occurs at university, a question that has been asked independently by Marmaros and Sacerdote (2006) using data on email exchanges between students. The exogenous allocation of first year students to close peer groups allows us to analyse the importance of "manipulated social proximity" against other factors like ethnicity and gender. These results are of interest because of the recent emphasis on the importance of diversity in the education process both by European and American universities.[3] To this end, our results show that the intervention did little to promote long-lasting diversity on campus. We cannot rule out, however, that a more sustained and focused intervention would deliver larger effects.

2.1.1 Related Literature and Channels.

Based on the empirical literature, we distinguish between two broad and exhaustive channels driving peer effects; social proximity and classroom-level effects.

- **Social Proximity**: peer effects driven by meaningful social interactions between classroom peers. Peer effects from this channel are restricted to peers who are socially proximate; those for whom bonds exist and social interactions occur.

- **Classroom-Level Effects**: peer effects that stem from the overall classroom environment and are independent of the social proximity between students. They potentially originate from and have an impact on all students in a classroom, even between students that do not explicitly interact.

[3]In the U.K., the former Prime Minister David Cameron and the Universities and Colleges Admissions Service (UCAS) announced applications to be name-blind from 2017 onward, after which several institutions introduced pilots. In the U.S., many leading American institutions, such as MIT and University of Chicago, filed an amicus brief in November 2015 with the U.S. Supreme Court in *Fisher v. University of Texas*. This brief stressed the role of government in diversity of higher education, of which race and ethnicity are components.

The social-proximity channel would, for instance, include having a high ability peer in the classroom with whom a student discusses material. This could potentially happen both inside or outside class. Alternatively, an example of a classroom-level effect is teachers responding to the composition of students in the classroom. Having many high ability students in a class might induce teachers to change the level of their instruction. A student posing an insightful question in class that benefits all other students is another example of a classroom-level effect.[4]

Several papers rely on social proximity, and thus interaction between peers, as the main explanation for spillovers. Booij et al. (2017) and Feld and Zölitz (2017) use voluntary course evaluation data and find that students with better tutorial peers reported better interactions with other students. In attributing the negative results of their experiment to voluntary sorting, Carrell et al. (2013) implicitly argue that peer effects are generated via the social proximity of peers.[5]

Other researchers attribute their findings to classroom-level effects. Duflo et al. (2011) argue that the resulting peer effects of a student tracking experiment can be explained by changes in teaching behavior based on the ability composition of the class. Lavy et al. (2012a) and Lavy and Schlosser (2011) explore potential channels using a student survey and find that a higher proportion of low ability students has negative effects on the quality of student-teacher relationships, on teachers' pedagogical practices, and increases classroom disruptions.[6]

The strategies used in the empirical literature thus far to explore potential channels is to (i) search for heterogeneity in the data that supports or refutes certain peer effect channels or (ii) look at additional outcomes using secondary data sources, such as student evaluations.[7] The results using the first strategy are, however, mostly circumstantial and unable to definitively rule out other competing explanations. An example of this is Carrell et al. (2009), who looks at the heterogeneity of peer effects between courses to find suggestive evidence of study partnerships as a driver of peer effects. With the second strategy researchers must often attribute their results to other unobserved factors (see *e.g.* Feld and Zölitz (2017)). In both cases, these strategies involve looking for an explanation after the fact. Researchers have rightly been cautious in interpreting the findings derived from these strategies.

[4]Because classroom-level effects are defined as the complement of social proximity, together they are exhaustive. Though our main distinction is between these two broad channels, we also use supplementary data to hint at finer channels such as those listed by Sacerdote (2011). We find suggestive evidence that spillovers revolve around collaborative self-study and peer-to-peer teaching.

[5]Other papers that attribute their results to the social-proximity channel include Garlick (2018); Brunello et al. (2010); Carrell et al. (2009); Stinebrickner and Stinebrickner (2006); Arcidiacono and Nicholson (2005).

[6]Other research relying on a classroom-level explanation are Oosterbeek and Van Ewijk (2014); Burke and Sass (2013); Lyle (2009); Foster (2006); Hoxby and Weingarth (2005).

[7]For strategy (i) see, among others, Garlick (2018); Oosterbeek and Van Ewijk (2014); Duflo et al. (2011); Brunello et al. (2010); Carrell et al. (2009); Lyle (2009); Foster (2006); Arcidiacono and Nicholson (2005); Hoxby and Weingarth (2005); Hoxby (2000). For strategy (ii) see, for example, Booij et al. (2017); Feld and Zölitz (2017); Lavy et al. (2012a); Lavy and Schlosser (2011); Stinebrickner and Stinebrickner (2006).

The definition of what constitutes a peer group varies substantially in the literature. It includes entire schools (Lavy and Schlosser, 2011), classes (Feld and Zölitz, 2017), dorms (Garlick, 2018) and dorm roommates (Sacerdote, 2001; Zimmerman, 2003), students in the same group during university orientation week (Thiemann, 2017), students that share more than a certain number of classes (De Giorgi et al., 2010), and students who sit next to each other in class (Lu and Anderson, 2014; Hong and Lee, 2017). It may be that different types of peers deliver spillovers via different channels. The manipulation of social proximity allows us to cleanly separate two broad and exhaustive channels in the same context. Furthermore, our results may be of more general interest than many of the studies mentioned above, as opportunities to manipulate classroom peers arise in almost every educational setting, while contexts where universities or schools can assign dorm mates or students' seating arrangements are far more infrequent.

Finally, it is worth noting that the relative importance of the two different channels might vary across different levels of education. Our focus is on university students and tutorial peer groups, which are mostly taught by senior students and PhDs. Because of the inexperience of these teachers, one might reason that teacher response is unlikely. However, evidence from a similar public Dutch university suggests academic rank of instructors is unrelated to student performance; Feld et al. (2018) show that full professors are not significantly more effective in tutorial teaching than students or PhDs. Moreover, since future employment at the university depends largely on their performance in student evaluations, teaching assistants (TAs) have incentives to teach well and put forth effort. Similarly, one might argue that disruptive students are not present at the university level. However, personal experience and interviews with TAs suggest otherwise. Notably, every TA at the university of our study undergoes a one-day training, part of which teaches them to deal with disruptive student behaviour through role-playing.[8] Thus, we believe that there is a priori little reason to dismiss the presence of either channel in the university setting, and that our results are not necessarily uninformative for other education contexts.

2.2 Context

2.2.1 Institutional Setting.

Our setting for studying peer effects is the economics undergraduate program at a large public university in the Netherlands. Every year the economics program experiences approximately 400 newly

[8]A web search reveals that many other universities also provide advice to their teaching staff on how to deal with disruptive students, indicating that the phenomenon is not absent in higher education. For example, see the following resource page from Stanford University: `https://teachingcommons.stanford.edu/resources/teaching-resources/interacting-students/classroom-challenges`.

enrolled first-year students. During the first two undergraduate years the program is identical for every student, as they follow the same twenty courses across the two years, covering basic economics, business economics, and econometrics. Come the third year, students must choose their own courses. The program only admits Dutch students. The admission requirement is based on a having a pre-scientific high school diploma.

The three academic years are divided into five blocks of eight weeks each (seven weeks of teaching and one week of exams).[9] Students in the first- and second year have one light and one heavy course per block, for which they can earn four and eights credits respectively. Sixty credits account for a full year of study.[10] In the first- and second year, courses consist of both lectures and tutorial sessions. The heavy courses have three large-scale lectures per week, while light courses have two. Heavy courses have two small-scale tutorials per week, while light courses have one. Lectures and tutorials both last for 1 hour and 45 minutes. While attendance at lectures is voluntary, first-year students have to attend at least 70 percent of the tutorials per course. Students who fail to meet the attendance requirement are not allowed to take the final exam for their course and must wait a full academic year before they can take the course again.

During tutorial sessions a teaching assistant (TA) typically works through question sets based on the materials covered in the lectures. Roughly 10 percent of the TAs are PhDs, with some exceptions the remaining 90 percent are senior students. Unlike lectures, the tutorial sessions often require preparation and active participation from the student, *e.g.* via discussion of assignments or related materials. First-year students follow the tutorials with the same group throughout the whole first year. To verify whether the 70 percent attendance requirement is met, TAs register attendance at the start of each session. The requirement ensures that students experience a sizable degree of exposure to tutorials and their tutorial peers, and are not able to voluntarily attend different groups during the first year. Appendix Table A.2.1 gives an overview of the first-year courses, their characteristics, and an accompanying tutorial description. We investigate peer effects originating from these first-year tutorial peer groups.

Grading is done on a scale that ranges from 1 to 10. Students fail a course if their grade is below 5.5. Most of the courses in first- and second year are (partly) multiple choice and therefore graded without interference by the instructor or TAs. For exams with open questions, instructors disallow TAs from grading their own groups.

[9] At the end of the academic year, at the start of summer, there is a resit period. During two weeks first- and second-year students have the opportunity to resit a maximum of three courses.

[10] In this institution credits are measured through ECTS, which is an abbreviation for European Transfer Credit System. This measure for student performance is used throughout Europe to accommodate the transfer of students and grades between universities. The guidelines are that one ECTS is equivalent to 28 hours of studying.

2.2.2 Close and Distant Peers.

A key institutional feature of the economics program is that each first-year tutorial group is divided into two subgroups. The university induces social proximity, and thus student-to-student interaction, only within these subgroups of students. For a student we term *close* peers to be the group with whom bonds are encouraged, where *distant* peers are the adjacent group of peers in the tutorial group with whom interaction is not encouraged. This means that if student $S1$ and $S2$ are in the same tutorial group but in different subgroups, the close peer group of student $S1$ will be the distant peer group for student $S2$ and vice versa.

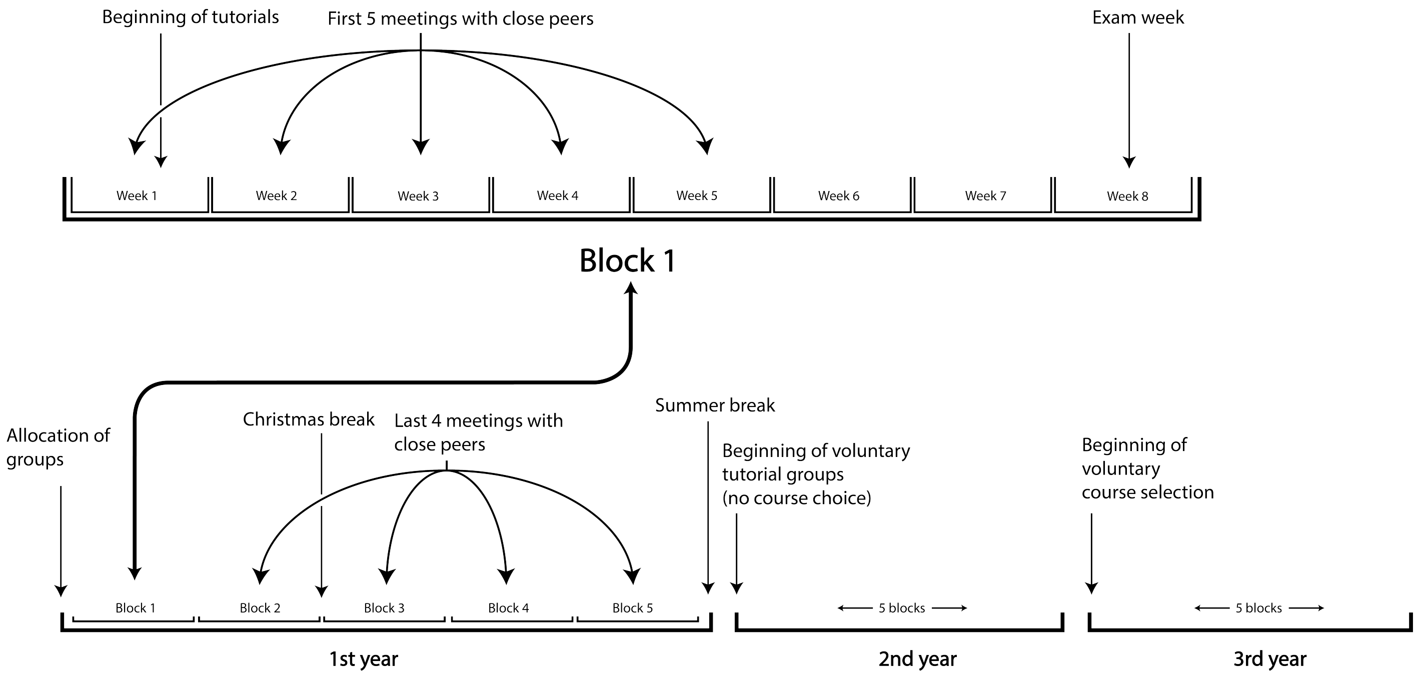
The main purpose of the close peer group is to facilitate the formation of social ties to help students adjust to, and get acquainted with, life at university. These ties are primarily facilitated via five compulsory close peer group meetings during the first block.[11] As discussed in more detail below, these meetings revolve around discussion and active student participation, which the university aims to foster via the smaller subgroups. The first close peer group meeting is in the first week of university, before any lectures or tutorials have taken place. As well as meeting each other in the subsequent tutorial sessions, which also include the set of distant peers, there are weekly close peer group meetings up until week five. During the first five weeks close peers see each other 20 times; 5 times at the close peer meetings and 15 times at the regular tutorials. There are four remaining meetings with the close peer groups that are evenly spread out across the year (one per block). An overview of the first block and the whole undergraduate program can be found in Figure 2.1.

The university assigns senior students as discussion leaders to guide the close peer meetings. The subjects and the setting of these meetings are less formal than the tutorial groups. The first close peer meeting is a get-to-know-you session, where students have to introduce themselves to the group. The subsequent four sessions in the the first block consist of group discussions of the use of study timetables, exam preparation, fraud and plagiarism, teamwork, and plans concerning the future of their studies, among other topics. There is an emphasis on active participation of all students during these discussions. Importantly, course material is not discussed during these meetings.

Given the timing and the nature of their introduction, the close peer groups serve as the first plausible group of fellow students that a new student will interact with and form friendships with. Our empirical evidence presented later on implies that the close peer meetings resulted in substantial social proximity between close peers, at least initially. Conversely, the structure of the program resulted in comparatively much less, if any, meaningful bonding with members of distant peer groups.

[11] While the students do not get any credits for these meetings, according to the Teaching and Examination Regulations students must attend all of these meetings in order to pass the first year. Our administrative attendance data reveals students attend on average 94 percent of the sessions of the group they have been assigned to.

Figure 2.1: An overview of the characteristics of the undergraduate Economics program relevant to our study



2.2.3 Assignment of Students to Groups.

During the final year of students' pre-scientific education, and before the start of the academic year, students must preregister for the economics program. Those who have done so are requested to come to campus on the first day of the academic year to confirm their registration.[12] This is done by means of approximately 10 to 15 administrative personnel, who add students' numbers and names to an electronic register.

A list containing the information of all students who confirmed their registration is sent to an administrative worker. This list is then sorted by a randomly assigned ID and group membership is determined on a rotating basis. The first student on the list is allocated to tutorial group 1, close peer group $1A$; the second student is allocated to tutorial group 2, close peer group $2A$; the third student is allocated to tutorial group 3, close peer group $3A$, and so forth. The allocation continues until the maximum tutorial group has been reached, after which the rotation begins again by allocating the next unassigned student to tutorial group 1, close peer group $1B$, the next student to tutorial group 2, close peer group $2B$, and so forth. The university uses this allocation method to ensure that students are exposed to new peers and that the groups are roughly of equal size.[13]

[12]In this way the university avoids, to a large extent, taking into account no-shows when forming the first-year groups.

[13]We conducted numerous interviews with the administrative worker and university administrators, and received accompanying documentation, in order to confirm that the allocation process occurred as described. The same administrative worker has been in charge of this process across the six cohorts we study. The allocation process is done with BusinessObjects BI and Microsoft Excel software.

Figure 2.2: A graphical representation of the allocation to tutorial and close peer groups for a hypothetical cohort

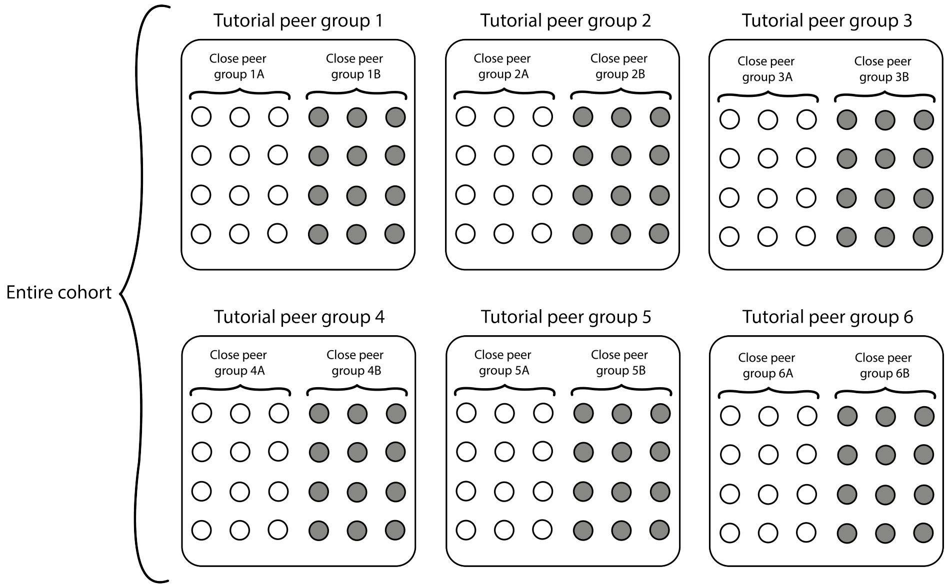


Figure 2.2 clarifies the structure of the tutorial and close peer groups for a hypothetical cohort. The 144 students, represented by dots, are distributed across 6 tutorial groups and 12 close peer groups. For a student in close peer group $1A$, her distant peers are those students belonging to close peer group $1B$, and vice versa.

A student who wants to follow the program, but did not show up at the first day of the year, is allocated to a group at the discretion of the administrative worker. Reallocating a student to a different group only happens in case of special circumstances, such as when a student practices top sports, has special needs, or has some otherwise unresolvable scheduling conflicts. Again, the groups to which these students are reallocated to is at the discretion of the administrator. Our data does not allow us to observe which student registered late or ended up in their group via a reallocation. According to the administrative worker these cases are rare, but may result in slightly different variation in peer ability and class size than would have been observed when strictly following the allocation procedure described above. We present balancing tests in Section 2.4 that cannot reject the final allocation results in a random assignment of students to tutorial, close, and distant peer groups.

2.3 Data

Our main source of data is the administrative database of the university between the academic years 2009-10 and 2014-15. This database includes the complete history of student outcomes and choices at university; grades of all courses followed by the student, first-year tutorial attendance, and second-year tutorial choice. Additionally we observe a rich set of student characteristics; gender, age, residential address, high school GPA and zip code, and the groups students have been assigned to in their first year. Our baseline results are based on almost 19,000 first-year grades from 2,300 students.[14] This sample only includes a student's first attempt at completing a course. Although we also observe resits, which are taken at the end of the academic year at start of summer, we do not include them in our analysis as they do not require preparation via tutorials.

High school GPA is a 50-50 weighted average of grades obtained during the last three years of high school and on the nationwide standardized exams at the end of high school (before entering university) across all courses. We use high school GPA as a comprehensive proxy for the latent ability of students and their peers. In case of classical measurement error, our estimate for spillovers would be attenuated as students are randomized into groups (Feld and Zölitz, 2017).[15]

2.3.1 Attendance and Student Evaluations.

In the first year all students are required to attend at least 70 percent of the tutorials per course. To verify whether the attendance requirements are met, TAs register attendance at the start of each tutorial. This attendance is then uploaded to the university portal and verified at the end of the block by the exam administration. We merge this attendance data with the administrative database, which allows us to observe attendance at the student-tutorial-course level for 98.5 percent of the student-course observations.[16]

At the end of the course, students are invited by email to fill in student evaluations. A set of 20 questions are asked covering 9 characteristics of the course, which are detailed in Appendix Table A.2.2. Merging the student evaluations to the administrative data gives a response rate of roughly 30 percent. Column (1) of Appendix Table A.2.8 reveals that participating in the course evaluation is selective. Students with a better high school GPA are more likely to respond. However, column (1) also shows the absence of a relationship between the high school GPA of a student's close peers and their response rate. Results using the course evaluations should be interpreted with caution, and we use them to provide supplementary evidence on the channels of peer influence.

[14]This sample excludes some students. For 227 students we do not observe high school GPA (225 students) or one of the main control variables (2 students). Furthermore, to ensure that peer GPA consists of an appropriate number of students, we dropped fourteen tutorial groups (215 students) for whom we observe less than ten students' GPA in at least one of the two close peer groups. Our results are completely robust to the inclusion of these groups. Note that these groups occurred because of missing data on high school GPA and because some students were reallocated after the initial assignment.

[15]There are two potential sources of measurement error in our measure of ability. First, for 50 percent high school GPA is determined via unstandardized school exams. It should be noted, however, that the Dutch Inspectorate of Education pays strong attention to schools where the grades on school exams deviate more than 0.5 points from grades on the nationwide standardized exams (DUO, 2014). Second, although students have followed the same level of education in high school (pre-scientific), entering the last three years of high school students must choose one of four tracks. Though these tracks share compulsory courses (such as Dutch), some courses between tracks differ. For a subsample we can show that over 70 percent of our students followed the same track.

[16]For our grade-analysis we use the whole sample. Results are identical for the sample that is matched to the attendance data. We verified that peer high school GPA cannot explain whether a student is matched.

2.3.2 Descriptive Statistics.

Table 2.1 shows the descriptive statistics by cohort. Panel A provides an overview of the student characteristics. Panel B does the same for student outcomes. All student characteristics show similar values across cohorts. The percentage of women fluctuates somewhat around 20 percent, the students are on average 19.5 years old halfway into their first year, and their high-school GPA is close to the nationwide average of 6.7 (scale from 1 to 10, a 5.5 is sufficient). Appendix Figure A.2.1 shows histograms of student's own high-school GPA, the leave-out mean for the tutorial- and close peer group, and the mean for the distant peer group. Notice that, in contrast to the leave-out mean for the close peer group, the mean for the distant peer group takes upon identical values for everybody in the same subgroup. This explains the somewhat more discrete nature of this figure. A histogram of the leave-in mean for the close peer group is similar to the mean for the distant peer group.[17]

Table 2.1 further shows that the size of the close peer group fluctuates between 12 and 14 students. In 2009 the groups where somewhat larger due to an unexpectedly high number of enrolled students. University grades seem to gradually increase, also reflected by the increase in the number of credits earned. This is most likely the consequence of stricter academic dismissal policies introduced halfway in our sample. Course dropout occurs if a student does not attend the final exam for that particular course. Across cohorts, 8 to 19 percent of the students dropped out of both courses in block 5, the final block of the first year. We refer to this as student dropout.

[17] Angrist (2014) shows that using leave-in means, rather than leave-out means, would only change the peer-effects estimate for close peer high school GPA by a factor of $N_g/(N_g-1)$, where N_g is the size of close peer group g. Therefore, this distinction has little to no importance for our results.

Table 2.1: Descriptive Statistics per Cohort

	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Panel A: Student Characteristics						
Female	0.21 (0.41)	0.21 (0.41)	0.22 (0.42)	0.22 (0.42)	0.21 (0.40)	0.23 (0.42)
Age	19.54 (1.65)	19.62 (1.29)	19.61 (1.28)	19.57 (1.57)	19.67 (1.34)	19.48 (1.42)
Distance to University (km)	21.77 (26.37)	24.03 (30.96)	21.62 (26.20)	22.56 (26.32)	26.39 (31.96)	18.08 (20.32)
Own High School GPA	6.72 (0.54)	6.60 (0.48)	6.63 (0.49)	6.62 (0.47)	6.68 (0.56)	6.68 (0.47)
Tutorial High School GPA	6.72 (0.09)	6.60 (0.10)	6.63 (0.10)	6.62 (0.10)	6.68 (0.13)	6.68 (0.09)
Close Peer High School GPA	6.72 (0.12)	6.60 (0.12)	6.63 (0.16)	6.62 (0.13)	6.68 (0.17)	6.68 (0.14)
Distant Peer High School GPA	6.73 (0.11)	6.60 (0.11)	6.63 (0.15)	6.62 (0.13)	6.69 (0.17)	6.68 (0.14)
Tutorial Group Size	35.30 (1.52)	26.84 (2.80)	22.31 (1.15)	22.08 (1.32)	26.12 (1.29)	24.17 (1.63)
Close Peer Group Size	17.72 (1.35)	13.51 (1.85)	11.19 (0.88)	11.08 (0.94)	13.12 (1.10)	12.12 (1.06)
Number of Students	458	371	356	308	442	361
Panel B: Student Outcomes						
Grades	5.98 (1.76)	5.91 (1.71)	6.38 (1.55)	6.06 (1.64)	6.21 (1.68)	6.35 (1.41)
Attendance	0.89 (0.16)	0.89 (0.12)	0.89 (0.10)	0.88 (0.11)	0.88 (0.11)	0.89 (0.10)
Number of Student-Grades Obs.	3598	2999	3098	2580	3462	2999
Number of Student-Att. Obs.	3433	2955	3094	2577	3436	2950
Number of Credits per Student	29.74 (1.62)	30.06 (18.88)	37.96 (18.12)	33.53 (20.59)	32.39 (21.61)	40.00 (20.69)
Number of Courses per Student	8.49 (2.48)	8.79 (2.26)	9.29 (1.84)	8.76 (2.25)	8.43 (2.55)	8.94 (2.24)
Dropout	0.18 (0.38)	0.16 (0.37)	0.08 (0.27)	0.15 (0.36)	0.19 (0.39)	0.12 (0.33)

Notes:
1. Table shows the mean and standard deviation per cohort of student characteristics (Panel A) and student outcomes (Panel B). Panel B is further divided into student-course level outcomes (first section) and student level outcomes (second section).
2. Age is evaluated on January 1^{st} in the academic year that the cohort started. Distance to University refers to the number of kilometers from a student's registered address to the university. High school GPA and university grades are unstandardized, measured on a scale from 1 to 10.
3. Dropout is the fraction of students who did not write an exam in the last block of the first year (block 5).

2.4 Empirical Specification

To derive our empirical model we start with the canonical specification for peer effects as laid out by Manski (1993):

$$Y_{igct} = \alpha_0 + \alpha_1 \overline{Y}_{(-i)gc} + \alpha_2 \overline{GPA}_{(-i)g} + \alpha_3 GPA_i + \mu_{gct} + \epsilon_{igct}$$

Where Y_{igct} is the grade at university of student i in tutorial group g on course c of cohort t. GPA_i is the average grade obtained in high school and the variables $\overline{Y}_{(-i)gc}$ and $\overline{GPA}_{(-i)g}$ are leave-out means for tutorial group g for student i of university grades and high school GPA respectively. Everything else that is common to tutorial group g is captured by μ_{gct}.

In the terminology of Manski (1993), α_1 measures the endogenous effect of peers' outcomes on the outcome of student i, α_2 captures the exogenous effect of pre-determined peer characteristics, and μ measures the correlated effects capturing, for example, common shocks such as a good TA. The distinction between α_1 and α_2 reveals little about the channels, but it does have different implications for policy, as endogenous effects might generate a social multiplier.[18] However, identification of α_1 is obscured, mostly due to the well-known reflection problem; did the peers affect student i, or did student i affect her peers? As such we follow most of the previous peer effects literature and solve for the reduced form.

2.4.1 Reduced-Form Peer Effects.

The standard linear-in-means reduced form specification is given by:

$$Y_{igct} = \beta_0 + \beta_1 \overline{GPA}_{(-i)g} + \alpha_3 GPA_i + \beta_2 \mu_{gct} + \tilde{\epsilon}_{igct} \tag{2.1}$$

Where $\beta_1 = \frac{\alpha_2 + \alpha_1 \alpha_3}{1 - \alpha_1}$. Subsequently a test for whether β_1 is different from zero is a test for the presence of peer effects, may they be exogenous and/or endogenous.

The institutional manipulation of the social proximity between students and their tutorial peers allows us to extend this standard model. We make a distinction between the leave-out mean of the close peer group $\overline{GPA\,Close}_{(-i)g}$ and the mean of the distant peer group $\overline{GPA\,Distant}_g$. To identify the separate potential channels we replace $\overline{GPA}_{(-i)g}$ in Equation (2.1) by the following expression:

$$\overline{GPA}_{(-i)g} = \frac{N^C - 1}{N^C + N^D - 1} \overline{GPA\,Close}_{(-i)g} + \frac{N^D}{N^C + N^D - 1} \overline{GPA\,Distant}_g$$

[18] When referring to the social multiplier, Manski (1993) uses the example of a tutoring program. If such a program is provided to only one half of the student population, it might indirectly help the other half of the students as well, as peers' outcomes affect each other.

Where N^C and N^D are the total number of students in the two subgroups within a tutorial group. In practice, $N^C = N^D = 13$. This substitution allows us to arrive at the following specification:

$$Y_{igct} = \beta_0 + \beta_1^C \overline{GPA\,Close}_{(-i)g} + \beta_1^D \overline{GPA\,Distant}_g + \alpha_3 GPA_i + \beta_2 \mu_{gct} + \tilde{\epsilon}_{igct} \qquad (2.2)$$

Estimates of this equation allow us to separate the two peer effect channels possibly at work. Equation (2.2) tests the restriction of Equation (2.1) that the spillovers β_1 from close and distant peers are identical. Recall that the only distinction between an individual's close and distant peers is that social proximity was induced with the former, whereas no social proximity exists with the latter.[19] Hence, the difference between β_1^C and β_1^D captures peer effects through the social proximity channel. If β_1^C and β_1^D are approximately equal, this indicates that peer effects work solely through classroom-level effects.[20]

Consistent with their definitions, the two channels are presented as being substitutes in the production of student grades. However, to capture possible complementarity between social proximity and classroom-level effects, some specifications will also include an interaction between close and distant peer ability.

The peer group meeting intervention that encouraged social proximity permits the investigation of the mechanisms underlying peer effects. In order for our results to be generalizable however, we must assume that the intervention itself does not alter the nature of the mechanisms through which peer effects operate in the classroom. In the counter-factual scenario in which social proximity between close peers was not encouraged, we think our finding of no classroom-level effects would hold. It seems unlikely that a non-invasive intervention of little duration would comprehensively change the nature of classroom peer effect channels. Instead, our findings suggest that without the intervention the spillovers from tutorial peers would be smaller than what we observe, and would diminish at a faster rate.

2.4.2 Balancing Tests.

As the average high school grade is a predefined measure, we avoid the reflection problem and the estimates for β_1 are unlikely to be biased by common shocks. The main identifying assumption,

[19] In practice, we cannot rule out ex-ante that some social proximity exists between a student and her distant peers. If this was the case, we would overestimate the importance of classroom-level effects and underestimate the importance of social proximity. Our finding of zero for β_1^D implies that there was no meaningful social proximity between students and their distant peers.

[20] In fact, because the mean GPA from the distant peer group contains one more student than the leave-out mean of the close peer group, if the spillovers from close and distant peers are identical then $\beta_1^C = \beta_1^D(\frac{12}{13})$. We confirm this in a simulation, in which we arbitrary re-allocate existing tutorial peer groups into placebo close peer groups 1,000 times. Estimating Equation (2.2) and taking the average of the estimates we verify that $\overline{\hat{\beta_1^C}} \approx \overline{\hat{\beta_1^D}}(\frac{12}{13})$. For practical testing purposes we deem this as sufficiently close to equality.

however, is that peer high school GPA is uncorrelated with other characteristics that might determine a student's grade. As we are not able to observe all other characteristics that might be important for grades, we need the covariance between $\overline{GPA}_{(-i)g}$ and (μ_{gct}, $\tilde{\epsilon}_{igct}$) to be zero. Random assignment of students to groups makes this identifying assumption likely to hold.

We test this identifying assumption in several ways. First, we analyze whether the treatment, in the form of assigned peer ability, can be explained by background characteristics (X_i) or high school GPA:

$$\overline{GPA}_{(-i)g} = \gamma_0 + \gamma_1 X_i + \gamma_2 GPA_i + T_t + \epsilon_{igt}$$

We include cohort fixed effects (T_t) as randomization into groups takes place cohort-by-cohort. Estimates of γ_1 or γ_2 that are different from zero most likely violate the identifying assumption mentioned above. Table 2.2 shows the results of this test, where column (1) to (3) take tutorial, close, and distant peer high school GPA as outcome variables respectively. Across the three specifications we find all student characteristics to be individually and jointly insignificant.[21] This stands in stark contrast to the joint significance of student characteristics in a regression where first-year GPA at university is taken as an outcome variable (p-value$<$0.000).

Our second balancing test is more flexible. We regress background characteristics - student number, gender, age, and distance to university - and high school GPA on close peer group dummies and cohort fixed effects. Next, in a separate model we regress the student characteristics upon cohort fixed effects only and perform a F-test on the small versus big model. This test would reveal if students with certain characteristics cluster together in certain groups. Appendix Table A.2.3 shows the F-test does not reject the null hypothesis for all student characteristics. In other words, a small model with cohort fixed effects only is favored above a model that also includes close peer group dummies.

We perform a similar analysis per cohort. We regress each student characteristic on a set of close peer group dummies separately for each cohort. Appendix Figure A.2.2a plots the histogram of the p-values of the close peer group dummies obtained from these regressions. As expected under randomization, the p-values are roughly uniformly distributed, where for instance roughly 10 percent of the p-values are below 0.10. Figure A.2.2b shows the results for this analysis are identical if close peer group dummies are replaced with tutorial group dummies. A Kolmogorov-Smirnov equality of distribution test does not reject the null-hypothesis of a uniform distribution in both cases; the p-

[21] If we regress student high school GPA on peer high school GPA we reach identical conclusions. Guryan et al. (2009) argue this balancing test should also control for the mean high school GPA of all peers that can be matched with student i in group g. In our case this control would be the leave-me-out mean GPA of her cohort. This is infeasible as there is no variation in the group that student i can be matched too. Indeed, GPA_i is related to the mean GPA of her cohort $\overline{GPA}_t$ and the leave-me-out mean GPA of her cohort, $\overline{GPA}_{(-i)t}$, by the following identity: $GPA_i = N \times \overline{GPA}_t - (N-1) \times \overline{GPA}_{(-i)t}$.

Table 2.2: Balancing Tests for Peer Ability

	Tutorial Peer GPA	Close Peer GPA	Distant Peer GPA
	(1)	(2)	(3)
Student Number	-0.0157	-0.0187	-0.0077
	(0.0410)	(0.0451)	(0.0401)
Female	-0.0339	-0.0319	-0.0212
	(0.0376)	(0.0457)	(0.0504)
Age	-0.0081	-0.0024	-0.0100
	(0.0220)	(0.0232)	(0.0191)
Distance to University	-0.0132	0.0022	-0.0227
	(0.0145)	(0.0173)	(0.0151)
Own GPA	0.0076	-0.0171	0.0285
	(0.0281)	(0.0283)	(0.0255)
Observations	2296	2296	2296
Adjusted R^2	0.151	0.085	0.098
F-test	0.25	0.26	0.77
p-value	0.938	0.933	0.570

Notes:
1. All regressions also include cohort fixed effects.
2. Peer GPA refers to the leave-out mean of high school GPA for the tutorial- and close peers, and to the mean for distant peers. All dependent and independent variables are standardized except for the female dummy.
3. The F-test, and corresponding p-value, refer to a test for the joint significance of all the independent variables shown in the table.
4. Standard errors in parentheses, clustered on the tutorial level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

values are equal to 0.86 and 0.60 for the histograms belonging to the close- and tutorial peer group dummies respectively.

Allocation of teaching assistants to tutorial groups is done for each course by the instructor of that specific course. Our analysis would still be compromised if instructors base the TA assignment upon tutorial group ability. Instructors are unaware of the GPA composition of the tutorial groups and base the assignment of the TAs upon scheduling restrictions. To confirm this, we code the gender of the TA and whether he or she was a PhD. If coordinators base their decisions on the difficulty of groups, they might, for example, assign PhD's to low GPA groups. Regressing TA type on tutorial peer GPA, however, shows that coordinators do not base TA assignment on class composition (see Appendix Table A.2.4). The same assignment method is used for the discussion leaders that guide the close peer group, though we cannot confirm this empirically as we do not observe these discussion leaders in our data.

We conclude that we are able to identify reduced-form peer effects and estimate Equation (2.1) and (2.2) without controlling for μ_{gct}. Throughout all specifications we will, however, include course-cohort fixed effects and background characteristics; student number, gender, age, and distance to university. The baseline results are identical when we do not control for background characteristics. We cluster standard errors at the tutorial level, which nests the close-peer-group level cluster. Own GPA, peer GPA, and the outcome variables (when suitable) are standardized over the estimation sample, such that the estimates can be interpreted in terms of standard deviations.

2.5 Baseline Results

Before presenting the baseline results for grades and passing rates, we discuss the extent to which course- and student dropout could potentially bias our estimates. Table 2.1 shows that the student dropout rate at the end of first year is relatively low; between 8 and 19 percent across the six cohorts. In Section 2.5.3 we will show that average peer high school GPA has no impact on the number of courses a student attends the final exam for nor on whether the student dropped out by the end of first year. We can show, but omit for brevity, that these null-results for number of courses and student dropout extend to the non-linear model used in Section 2.5.5. Selection bias therefore does not contaminate the following baseline peer effects estimates.

2.5.1 First-Year Grades and Passing Rates.

Table 2.3 shows our baseline results, where panel A regresses first-year grades upon average peer high school GPA. Column (1) shows the estimated effect of tutorial peers. The positive coefficient has a p-value of 0.11 and shows that a one standard deviation increase in tutorial peer high school GPA increases a students' first year grade by 0.019 standard deviations. Columns (2) and (3) show the effect while separating the tutorial group by one's close- and distant peers. This reveals that the positive spillovers are entirely driven by close peers. The estimate for peer GPA when moving from tutorial to close peers in column (2) increases somewhat in magnitude and precision. It is statistically significant at the 5%-level. The estimate for distant peers in column (3) is economically and statistically indistinguishable from zero. Column (4) shows the estimates for close- and distant peer high school GPA are identical to (2) and (3) respectively when including both peer measures in one regression. These results imply that peer effects are entirely driven by social proximity.

In terms of the Dutch grading scale, columns (2) and (4) imply that increasing the close peers' high school GPA from 6.5 to 7 increases a student's grade from 7 to 7.14. This is economically small, but 2.1 times the size of Feld and Zölitz (2017), while Booij et al. (2017) find no peer spillovers in their linear-in-means specification. Both of these studies investigate spillovers in a similar context as ours; classroom peer effects at a public university in the Netherlands. This suggests that fostering social proximity has the capacity to generate larger spillovers than previously found in the literature.

Whereas students with good peers obtain higher grades, they are not necessarily better off if the only goal is to pass courses. We study the probability of passing a first-year course in panel B of Table 2.3, where the outcome variable is replaced with a pass-fail indicator. Column (1) shows that a one standard deviation increase in the high school GPA of one's tutorial peers increases the probability of obtaining a sufficient grade by 0.9 percentage points. This effect is significant at the 5%-level. Again, columns (2) to (4) show that these spillovers originate entirely from close peers.

Column (5) of both panel A and B includes an interaction effect between high school GPA of the close and distant peers. This interaction effect tests for possible complementarities between social proximity and classroom-level effects. For instance, having a superstar student in class posing insightful questions may only increase grades if one has high ability close peers to discuss the questions with. We find this interaction term is negative for grades and the probability of passing, but insignificant in both cases. We interpret this as showing that complementarities between both channels are unlikely to play a role.

2.5.2 Randomization Inference.

The results above use analytic standard errors. In this section we present p-values based on randomization inference for the baseline results on first-year grades, an alternative inference approach that does not rely on large sample approximations. This method involves re-drawing a large number (10,000) of randomly assigned hypothetical tutorial and close peer groups, respecting the size of the

Table 2.3: Peer Effects on First-Year Course Grades (Panel A) and Pass or Fail (Panel B)

	(1)	(2)	(3)	(4)	(5)
			Panel A: Grades (Standardized)		
Tutorial Peer GPA	0.0191				
	(0.0118)				
Close Peer GPA		0.0255**		0.0254**	0.0256**
		(0.0104)		(0.0106)	(0.0109)
Distant Peer GPA			0.0034	0.0008	0.0010
			(0.0131)	(0.0130)	(0.0131)
Close $\times$ Distant					-0.0150
Peer GPA					(0.0122)
Own GPA	0.3427***	0.3434***	0.3427***	0.3433***	0.3433***
	(0.0119)	(0.0119)	(0.0118)	(0.0119)	(0.0120)
Observations	18736	18736	18736	18736	18736
Adjusted R^2	0.323	0.323	0.322	0.323	0.323
			Panel B: Pass (1) or Fail (0)		
Tutorial Peer GPA	0.0090**				
	(0.0043)				
Close Peer GPA		0.0080*		0.0075*	0.0075*
		(0.0042)		(0.0043)	(0.0043)
Distant Peer GPA			0.0056	0.0048	0.0048
			(0.0049)	(0.0049)	(0.0049)
Close $\times$ Distant					-0.0005
Peer GPA					(0.0046)
Own GPA	0.1186***	0.1189***	0.1183***	0.1187***	0.1187***
	(0.0048)	(0.0048)	(0.0047)	(0.0048)	(0.0048)
Observations	18736	18736	18736	18736	18736
Pseudo R^2	0.187	0.187	0.187	0.187	0.187

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the tutorial- and close peers, and to the mean for distant peers. Own GPA refers to own high school GPA. All GPA measures are standardized.
3. Standard errors in parentheses, clustered on the tutorial level.
4. Panel A is estimated with OLS, Panel B uses Probit. Marginal effects are reported.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

original groups. For each of these hypothetical groups, we re-run the models presented in Panel A of Table 2.3 in order to assess the effect of the hypothetical peers' high school GPA on students' first-year grades. Comparing the actual estimate to the estimates from the simulated groups allows us to test the sharp null hypothesis that peer effects are equal to zero (Athey and Imbens, 2017). The results for the corresponding exact p-values are presented in Appendix Table A.2.5 and Figure A.2.3, which are nearly identical to those presented in Table 2.3. Given the similarity between the two approaches, the remainder of this paper uses analytic standard errors.

Additionally, these results address one of the concerns of Angrist (2014). He shows that the peer effects estimate is identical to the (scaled) difference between a 2SLS estimator using peer group dummies as instruments for individual high school GPA and an OLS estimator of individual GPA. In some settings this may lead to a spurious, mechanically driven finding of peer effects. In our setting, however, with random assignment of students to many small groups, there is little reason for this estimate to be different from zero in the absence of spillovers (Angrist, 2014). This is confirmed by the fact that the peer effect coefficients from the 10,000 hypothetical groups, containing unconnected students, are centred around zero.

2.5.3 Additional Outcomes.

In this section we turn our attention to five additional first-year outcomes: credit weighted GPA, number of credits, number of courses taken, student dropout, and tutorial attendance. We analyse the first four of these outcomes by estimating our baseline equations on the student level.

Table 2.4 shows the results, where columns (1) and (2) reveal that the positive effects on grades and passing rates have a cumulative effect on a student's GPA (p-value$<$0.01) and the number of credits she collects (p-value$=$0.13). The estimates indicate that a one standard deviation increase of close peer high school GPA increases a student's credit weighted GPA (total first-year credits) by roughly 0.04 standard deviations (0.52 credits). Column (3) and (4) reveal this increase in student performance is not due to the fact that peers impact the number of courses a student writes the final exam for.[22] Column (5) shows that peer GPA does not change the probability of student dropout, which is measured by an indicator variable that takes the value one if a student was no longer active in block five of their first year.

Appendix Table A.2.6 shows the results when analysing the impact of peer high school GPA on the percentage of tutorials attended per course in the first year. These estimates show that peers do not have an effect on average tutorial attendance. Recall, however, that students are required to attend 70 percent of the tutorials per course, so the scope for any improvement would be limited.

[22] Note that this also implies that course dropout is not influenced by peer GPA.

Table 2.4: Peer Effects on Additional Outcomes

	GPA Weighted by Credits	Number of Credits	Number of Courses	Followed the Course? Balanced Panel	Dropout
	(1)	(2)	(3)	(4)	(5)
Close Peer GPA	0.0450***	0.5230	0.0269	0.0034	0.0009
	(0.0146)	(0.3462)	(0.0497)	(0.0051)	(0.0083)
Own GPA	0.5073***	8.7081***	0.4747***	0.0539***	-0.0693***
	(0.0168)	(0.3560)	(0.0438)	(0.0054)	(0.0082)
Observations	2218	2218	2218	22180	2218
R^2	0.300	0.241	0.062	0.048	0.056
Binary Outcome	No	No	No	Yes	Yes

Notes:
1. All regressions include cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Column (1), (2), (3) and (5) are estimated on the student level. Column (4) creates a balanced panel on the student-course level, where the outcome variable takes the value one if a student wrote the final exam for that course and zero otherwise.
4. Column (1) has first-year credit weighted GPA as outcome variable and is based on the number of courses that the student took. Column (2), (3) and (4) refer to the number of credits obtained or the number of courses a student wrote the final exam for. Dropout in column (5) is one if a student did not write an exam in the last block of the first year and zero otherwise. Credit weighted GPA in column (1) is standardized, all other outcomes are unstandardized. Number of credits range from 1 to 60. Number of courses range from 1 to 10.
5. Across the six cohorts there are 78 students (3.4%) who confirmed their registration on the first day but for whom we observe no valid grade. These students dropped out before the first exam week. As we cannot calculate a GPA for them, these students are dropped from this analysis. Results do not change when we include these students.
6. Column (1), (2) and (3) are estimated with OLS, column (4) and (5) with Probit. Marginal effects are reported. The R^2 refers to the Adjusted and Pseudo R^2 respectively.
7. Standard errors in parentheses, clustered on the tutorial level.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.5.4 Robustness.

The results above show that peer GPA does not affect dropout, which implies our results are not contaminated by selection bias. However, the estimate for own GPA in Table 2.4 reveals that low GPA students take fewer courses and have a higher probability of dropping out by the end of their first year. This means that a student randomized into a tutorial group with many low ability rather than high ability students will experience a larger amount of course dropout among her peers, and thus have a smaller actual class size. This results in a positive correlation between peer GPA and class size, which could partly explain our baseline results if class size also impacts grades. Appendix Figure A.2.4 plots the number of students writing the final exam as a fraction of the initial students per block and separately for high, average, and low GPA close peer groups. This reveals that dropout increases during the year, being 15 to 20 percent at the end of the first year. It also reveals that dropout is somewhat larger for low GPA close peer groups.

We investigate whether our results are robust to class size and course dropout in Table 2.5, which presents the results of our baseline equation while including variables measuring class size and course dropout as explanatory variables. Column (1) includes a dummy for the assigned number of students to the close peer group at the start of the first year, column (3) for the actual number of students that wrote the exam for the course, and column (6) for the difference between the two. The latter is a measure for dropout per course. All three columns reveal a stable estimate for close peer GPA, suggesting that class size and course dropout are unlikely to explain our baseline results.

Table 2.5: Robustness of Baseline Peer Effects

	Grades (Standardized)						
	OLS	OLS	OLS	OLS	IV	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Close Peer GPA	0.0254**	0.0253**	0.0275**	0.0289***	0.0291***	0.0261**	0.0387***
	(0.0112)	(0.0107)	(0.0107)	(0.0108)	(0.0113)	(0.0103)	(0.0146)
Peer GPA × Assigned Class Size		-0.0059					
		(0.0093)					
Peer GPA × Actual Class Size				0.0042	-0.0043		
				(0.0076)	(0.0112)		
Peer GPA × (Assigned-Actual)							-0.0056
							(0.0052)
Own GPA	0.3435***	0.3436***	0.3436***	0.3438***	0.3438***	0.3433***	0.3433***
	(0.0120)	(0.0119)	(0.0119)	(0.0119)	(0.0119)	(0.0119)	(0.0119)
Observations	18736	18736	18736	18736	18736	18736	18736
Adjusted R^2	0.324	0.323	0.323	0.323	0.323	0.323	0.323
F-tests on				Assigned Class Size:	141.85		
Excl. Instruments				Its Interaction with Peer GPA:	475.24		
Class-Size Dummies	Yes	No	Yes	No	No	Yes	No
Robustness Check	Assigned Class Size			Actual Class Size		(Assigned-Actual)	

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Column (1) includes dummies for the number of students at the beginning of the year in the close peer group (assigned class size), column (3) for the number of students that wrote the exam for the course (actual class size on the course-cohort level), and column (6) for the difference between these two (assigned-actual). The latter is a measure for course dropout.
4. Assigned and actual class size are standardized in column (2), (4) and (5). The difference between the two in column (7) is not standardized. The coefficient on close peer GPA in column (7) measures spillovers in classes where there has been no course dropout (assigned-actual=0). Roughly 20 percent of the groups experience no course dropout and have a value of zero, where the average is 2.19.
5. Column (5) uses the assigned number of students and its interaction with close peer GPA as instruments for the actual number of students and for its interaction with close peer GPA.
6. Standard errors in parentheses, clustered on the tutorial level.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Columns (2), (4), and (7) of Table 2.5 include the assigned class size, actual class size, and course dropout as continuous variables, while also including their interaction with close peer GPA. The measures for original and actual class size in column (2) and (4) are standardized, while the difference between the two in column (7) is unstandardised. As such, the estimate for close peer GPA in (7) measures the peer effect for groups where there has been no course dropout. Again we find stable estimates for close peer GPA across all three columns. Moreover, we find the estimates for the interaction terms between peer GPA, class size, and course dropout to be unimportant. From this we conclude that the social proximity, and the corresponding nature of spillovers, is not different between classes of different size.

Whereas assigned class size is exogenous, one may have remaining concerns that actual class size is an outcome of close peer GPA. Therefore we report an additional specification in column (5) of Table 2.5, where we use assigned class size as an instrument for actual class size.[23] Using only the variation in actual class size that originates from the original assignment, we find our results to be virtually unchanged.

2.5.5 Heterogeneity.

Do the baseline estimates of Section 2.5.1 hide heterogeneity by own and peer ability? This question has important implications for policy. It is only when peer effects are non-linear that aggregate gains can be generated by reorganising peer groups.

Following Carrell et al. (2013) we test for heterogeneity using a two-way interaction model. We define low and high ability students to be in the bottom and top quartiles of high school GPA across the six cohorts. The remaining 50 percent of students are defined as being of average ability. For every student we calculate the (leave-out) proportion of low, middle, and high ability students separately for their close and distant peer groups. We estimate models with interactions of student's own ability type with the fraction of high and low ability peers. For each ability type, these interactions show the impact of increasing the proportion of high or low ability students by decreasing the proportion of average ability students. For example, Own Low $\times$ Peer High shows the estimated effect on student performance for low ability students of increasing the proportion of high ability students by decreasing the proportion of average ability students in the relevant peer group.[24]

[23] The variation in assigned class size comes partly from the original allocation and partly from the cases in which the administrator reallocates students across tutorial and close peer groups (see Section 2.2).

[24] Appendix Table A.2.7 repeats the first balancing test described in Section 2.4 while replacing average peer GPA as the outcome variable separately with the (leave-out) share of low, average, and high ability peers in the close and distant peer group. We find that student characteristics cannot explain the share of peers by ability type; only two out of the 35 estimated coefficients (γ_1 and γ_2) are significant, and the characteristics are always jointly insignificant. This result holds if we perform this balancing test separately for low, average, and high ability students.

Table 2.6: Heterogeneity by High School GPA of Peer Effects

	Grades (Standardized)				Pass (1) or Fail (0)	
	Close	Distant	Close	Distant	Close	Distant
	(1)	(2)	(3)	(4)	(5)	(6)
Share Peer High	0.1774*	0.0096				
	(0.0915)	(0.1104)				
Share Peer Low	-0.0483	0.0171				
	(0.1064)	(0.1108)				
Own High × Peer High			0.3659**	0.0007	0.2258***	0.1069
			(0.1456)	(0.1701)	(0.0730)	(0.0789)
Own High × Peer Low			-0.3036**	0.1024	-0.1141*	0.0329
			(0.1507)	(0.1506)	(0.0614)	(0.0666)
Own Avg × Peer High			-0.0257	-0.0687	-0.0081	-0.0153
			(0.1086)	(0.1239)	(0.0434)	(0.0508)
Own Avg × Peer Low			0.1063	0.0474	0.0698	0.0506
			(0.1196)	(0.1272)	(0.0463)	(0.0509)
Own Low × Peer High			0.3510**	0.1987	0.1146*	0.0884
			(0.1503)	(0.1624)	(0.0593)	(0.0655)
Own Low × Peer Low			-0.1492	-0.1654	-0.0400	-0.0689
			(0.2212)	(0.2002)	(0.0794)	(0.0692)
Observations	18736	18736	18736	18736	18736	18736
R^2	0.323	0.322	0.324	0.323	0.188	0.187
Binary Outcome	No	No	No	No	Yes	Yes

Notes:
1. All regressions include course-cohort fixed effects, controls; student number, gender, age, and distance to university, and own high school GPA.
2. Students are classified into dummies that refer to the bottom 25 percent (low), middle 25 to 75 percent (average), and top 25 percent (high) of high school GPA. The peer measures are the (leave-out) shares of students in the close (distant) peer group belonging to each category. The shares are unstandardized.
3. Odd columns include the shares for the close peer group and even columns for the distant peer group.
4. Column (1) to (4) are estimated with OLS, column (5) and (6) with Probit. Marginal effects are reported. The R^2 refers to the Adjusted and Pseudo R^2 respectively.
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6 presents our results. Column (1) and (2) first document that our baseline results from the linear-in-means specifications carry over to a model where we use the share of high and low ability students, rather than the mean of peer high school GPA, to measure peer ability. Next, column (3) and (4) show the heterogeneity results on first-year grades for the close and distant peer groups respectively, where column (5) and (6) do this for a pass-fail indicator. The results in column (3) and (5) reveal spillovers that are roughly linear in close peer ability, implying that the estimates of the linear-in-means model are insightful. Specifically, the columns show that the observed close peer spillovers are driven primarily by low and high ability students benefiting from social interactions with high ability students. Both high and low ability students are negatively affected by increasing the share of low ability students, insignificantly so for low ability students. Increasing the share of either high or low ability students appears to have no impact on average ability students. Conversely, column (4) and (6) show that the proportion of high and low ability types in one's distant peer group has no significant effect on grades or passing rates for any ability type, further supporting the lack of classroom-level effects.

The coefficient for Own High $\times$ Peer High in column (3) reveals that increasing the share of high ability students by 25 percent, the equivalent of replacing 3 out of 12 average ability students with 3 high ability students, increases the grade of a high GPA student by almost 0.1 standard deviation. To get a sense of the size of this effect, we follow Marie and Zölitz (2017) and compare it to other treatments known to have an impact on student performance in higher education. An estimate of 0.1 standard deviation is roughly twice the size of having a same-sex instructor (Hoffmann and Oreopoulos, 2009), resembles the effect of increasing professor quality by one standard deviation (Carrell and West, 2010), and is similar to the impact of a temporary restriction of legal cannabis access (Marie and Zölitz, 2017). It is perhaps useful to remark that 0.1 standard deviation corresponds to approximately half of the math gender gap in the fifth grade in the U.S. (Fryer Jr and Levitt, 2010).

In an additional analysis, we considered more restrictive definitions of high and low ability students to better reflect the concept of having superstar students or bad apples in the classroom. In particular, we defined superstar students as those having a GPA above 8.25 (cum laude) and bad apples as those having a GPA below 5.75; both categories form roughly one percent of our sample. Subsequently we constructed close and distant peer group dummies which are equal to one if the group contained such a student. Replicating the regression in column (1) of Table 2.6, while replacing the shares with a close-peer-group superstar and bad-apple dummy, we find the first is significantly positive and the latter to be insignificantly negative. Similar to column (2), both the superstar and bad-apples dummy for the distant peer group are smaller and statistically insignificant. Separating these effects by students' own ability, we find a similar pattern for low, average, and high ability students as documented in column (3) and (4) of Table 2.6. These results further support peer effects revolve around meaningful social interaction between peers, rather than classroom-level effects (results available upon request).

2.5.6 Group Assignment Policies.

The previous results imply that alternative assignment policies entail a transfer from one student group to the other. Therefore it is not possible to provide a Pareto-ranking of different policies. However, we can use the results in Table 2.6 to estimate the effects of alternative assignment policies. University administrators that want to maximize student grades can use such an exercise to weigh the grade benefits of one group against the costs of another.

Following Booij et al. (2017) we consider five alternative group assignment policies; low, average, high, three-way, and two-way ability tracking. Table 2.7 summarizes, for the average student as well as per ability type, the estimated change in a first-year course grade when switching from the current ability mixing regime to one of the five tracking policies. According to these estimates, the policy that will deliver the largest increase in student performance is the high tracking policy, whereby high ability students are grouped together and low and average students are mixed to form the remaining groups. Note however that this policy is predicted to decrease grades for low ability students compared to mixing.

A potential concern with using estimates based on ability mixing to inform alternative group assignment policies is that some peer configurations will not be covered by the data.[25] Such a problem was encountered by Carrell et al. (2013), who found that extrapolating from estimates based on ability mixing failed to predict the results of alternative group assignments. Given that social proximity is vital for the existence of peer effects, it may well be that such failures can be attributed to social proximity breaking down in more extreme group configurations. Our results are based on a setting in which social proximity has been fostered between close peers. If such fostering is achieved in more extreme group configurations then the results presented here may actually provide an accurate description of what will occur under alternative assignment policies. Given the lack of support they should still be treated with caution, however.

[25]This point has recently been made by Booij et al. (2017). In their study they manipulate the composition of groups to achieve a wider range of support in peer ability than observed under ability mixing.

Table 2.7: Estimated Effects of Alternative Group Assignments Compared to Mixing

			Effect For Student With		
		Average Effect	Low GPA [L]	Avg. GPA [A]	High GPA [H]
		(1)	(2)	(3)	(4)
Track Low	[L],[A,H]	-0.0400	-0.2030	-0.0290	0.1009^{**}
		(0.0512)	(0.1829)	(0.0300)	(0.0396)
Track Average	[A],[L,H]	0.0062	0.0476	-0.0202	0.0177
		(0.0115)	(0.0589)	(0.0432)	(0.0502)
Track High	[H],[L,A]	0.0686^{**}	-0.1036^{**}	0.0148	0.3484^{***}
		(0.0335)	(0.0486)	(0.0271)	(0.1167)
Three-way Tracking	[L],[A],[H]	0.0262	-0.2030	-0.0202	0.3484^{***}
		(0.0556)	(0.1829)	(0.0432)	(0.1167)
Two-way Tracking	[L,A],[A,H]	0.0123	-0.1247^{*}	0.0011	0.1717^{***}
		(0.0255)	(0.0730)	(0.0012)	(0.0551)

Notes:

1. For each alternative group assignment, we randomly allocate students depending on their ability type to groups of 14 to 15 students. The student types are low ability [L], average ability [A], and high ability [H] defined by the bottom quartile, two middle quartiles, and top quartile of high school GPA respectively.

2. Low (average or high) tracking involves grouping low (average or high) ability students together, while mixing the remaining students. Three-way tracking involves separate groups for each ability type. Two-way tracking involves defining students as either high or low ability, depending on whether their high school GPA is above or below the median. Groups are then composed of only high or low ability students.

3. For each student we subtract the actual leave-out ability shares (mixing) from the ability leave-out shares obtained via the alternative group assignments, denoted by $(x_{track} - x_{mixing})$. Then the average tracking effects are equal to $(\overline{x}_{track} - \overline{x}_{mixing})'\hat{\beta}$. Note that nearly identical estimates can be derived directly from column (3) of Table 2.6.

4. Standard errors are equal to $\sqrt{(\overline{x}_{track} - \overline{x}_{mixing})' V(\hat{\beta})(\overline{x}_{track} - \overline{x}_{mixing})}$, and shown in parentheses.

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

2.6 Nature of Social Interactions

Our results indicate better peers have small positive implications for a student's grades, passing rates, cumulative GPA, and credits in first year. These spillovers originate from peers with whom students are socially proximate and interact with. What is the nature of these social interactions? A possible answer to this question allows us to speak to the finer categorization of possible peer effect mechanisms listed by Sacerdote (2011), including peer-to-peer teaching or effects on student motivation or preferences.

We start with the use of course evaluations. Recall that the response rate, which is roughly 30 percent, is unrelated to a student's close peer GPA (see Appendix Table A.2.8). Hence, we worry little about sample selection when interpreting the following set of results. Table 2.8 uses data on self-reported lecture attendance and total study time to investigate whether the beneficial social interactions changed the inputs regarding the study process. Column (1) reports the effect of close peer high school GPA on an indicator for whether the student attended lectures. Column (2) does this for total study time (tutorials + lectures + self study). The estimates reveal that a student with better close peers is less likely to attend lectures, while reported total study time is not impacted. The estimate in column (1) suggests that a one standard deviation increase in close peer high school GPA decreases the probability to attend lectures by 1.8 percentage points. Due to the rough (binary) nature of the question, however, we are inclined to interpret only its sign and significance (p-value=0.019).[26] Recall that Appendix Table A.2.6 showed that tutorial attendance is unaffected by close peer high school GPA. Taken together, the estimates in column (1) and (2) suggest that students with better close peers substituted lecture attendance for additional self study.

Next we investigate the impact of close peer high school GPA on perceived lecturer and TA quality, and the perceived usefulness of lectures and tutorials. Column (3) and (4) indicate that having better close peers significantly decreases the perceived quality of the lecturer and usefulness of the lectures. This is consistent with, and further reinforces that, students substitute lecture attendance for additional self study. It seems most likely that this increase in self study involves close peers studying together. However, an alternative explanation might be that the beneficial student-to-student interactions only take place during the tutorials, after which individual self study takes place. If this is the case, we would expect students' perception of the quality of their TA and the usefulness of tutorials to increase when having better close peers. Column (5) shows that close peer high school GPA is unrelated to students' perceptions of the quality of the TA. Column (6) shows that there are

[26]For this question students are asked only about the extensive margin of their lecture attendance: "Have you attended lectures?". Even students who attended a few lectures may answer this question with yes (1) instead of no (0). As such, it may well be that these results understate the true reduction in lecture attendance.

Table 2.8: Peer Effects on Time Use and Additional Outcomes using Course Evaluations

	Attended Lectures	Total Study Time	Lecturer Quality	Usefulness Lectures	TA Quality	Usefulness Tutorials
	(1)	(2)	(3)	(4)	(5)	(6)
Close Peer GPA	-0.0180**	-0.1935	-0.0574***	-0.0585**	-0.0332	-0.0064
	(0.0077)	(0.1877)	(0.0195)	(0.0285)	(0.0241)	(0.0281)
Own GPA	-0.0139	-0.5414***	0.0348*	0.0268	0.0029	0.0024
	(0.0089)	(0.1484)	(0.0204)	(0.0201)	(0.0191)	(0.0308)
Observations	4361	4361	3560	2178	3560	2178
R^2	0.147	0.268	0.245	0.251	0.079	0.124
Binary Outcome	Yes	No	No	No	No	No

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. The dependent variable in column (1) is the answer to the question "Have you attended lectures?". The dependent variable in column (2) is the answer to the question "Average study time (hours) for this course per week (lectures+tutorials+self study)?" where we used the maximum for the interval to convert the categories into hours. The dependent variables in column (3) and (5) are the mean of the answers to the questions that evaluate the Lecturer/TA. The dependent variables in column (4) and (6) are the answers to the questions "Were the lectures/tutorials useful?". The dependent variables in column (3) until (6) are standardized.
4. Column (1) is estimated with Probit, the other columns with OLS. Marginal effects are reported. The R^2 refers to the Pseudo and Adjusted R^2 respectively.
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

no effects on the perception of the usefulness of the tutorials. Combining these results suggest that students exposed to better close peers substitute lecture attendance for collaborative self study.[27]

Next we turn to our data on student gender, high school location, and residence to shed additional light on the nature of the social interactions. We calculate for each student the leave-out proportion of females in their close peer group, the number of peers in their close peer group that attended the same high school as the student, the distance of the student's residence to the residence of every student in their close peer group, and the leave-out proportion of a student's close peers that live in the city in which the university is located.

Column (1) of Table 2.9 adds the leave-out share of females in the close peer group while interacting it with the female dummy in our baseline equation. The results replicate the finding of Oosterbeek and Van Ewijk (2014), who also find that the gender composition does not have an effect on student performance. Column (1) also documents an unchanged estimate for close peer high school GPA. This implies that the meaningful social interactions do not only take place in certain high ability groups with a high share of males or females. Column (2) shows that being assigned peers from one's former high school in the close peer group does not have implications for spillovers. Given that high school peers are most likely acquainted before university, this points to spillovers also being generated between formerly unknown peers.

If collaborative study meetings would take place outside university, we would expect to observe larger peer effects for students who live closer to their high ability peers. In column (3) we include the median distance of a student's residence to her close peers and interact this with close peer high school GPA. We do not find that a student who lives closer to her peers enjoys larger spillovers. This suggests that the study meetings take place on the university campus.

The notion that students benefit from collaborative self-study outside class implies that students would fail to benefit from having better close peers if these peers have other commitments that prevent such studying. We attempt at investigating this in column (4), which includes a dummy for whether the student lives in the city of the university and the leave-out share of students within their close peer group living in the city. First, notice that column (4) documents that city students perform significantly worse in their first year, scoring on average 0.11 standard deviations lower. Moreover, with our administrative tutorial attendance data we can show that the percentage of first-year tutorials attended per course is 0.07 standard deviations lower for city students (p-value=0.001). We conjecture that these findings partly reflect the large range of extra-curricular activities available to these students,

[27] Appendix Table A.2.8 reveals no effect of students' close peers on the remaining questions regarding their perceptions of the course.

Table 2.9: Peer Effects by Gender, Prior Bonds and Location

	Grades (Standardized)			
	(1)	(2)	(3)	(4)
Close Peer GPA	0.0250**	0.0286**	0.0254**	0.0190*
	(0.0104)	(0.0110)	(0.0104)	(0.0099)
Share of Female Peers	-0.0087			
	(0.0120)			
Female × Share of Female Peers	0.0290			
	(0.0232)			
Peer Same High School × Peer GPA		-0.0155		
		(0.0229)		
Distance of Peers to Your Residence × Peer GPA			-0.0091	
			(0.0089)	
Live in City				-0.1113***
				(0.0294)
Share of Peers that Live in City × Peer GPA				-0.0246**
				(0.0110)
Own GPA	0.3430***	0.3426***	0.3435***	0.3408***
	(0.0119)	(0.0116)	(0.0120)	(0.0122)
Observations	18736	18229	18736	18736
Adjusted R^2	0.323	0.324	0.324	0.325

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Column (1) includes the leave-out share of females in the close peer group (standardized) and its interaction with the gender dummy. Column (2) includes the number of students that attended the same high school (unstandardized) and its interaction with close peer GPA. For some students we do not observe their high school address, explaining the somewhat fewer number of observations. Column (3) includes the median distance of a students' peers to his or her residence (standardized) and its interaction with close peer GPA. Column (4) includes the leave-out share of peers that live in the city where the university is located (standardized) and its interaction with close peer GPA.
4. Standard errors in parentheses, clustered on the tutorial level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

most of whom are living outside of their parent's home for the first time.[28] The coefficient for the interaction of close peer ability with the proportion of close peers living in the city implies that peer effects vanish if all of one's close peers live in the city. Although we cannot definitively rule out other competing explanations, we believe this result is most consistent with the existence of spillovers depending upon peers not having busy social lives or other distractions outside of class.[29]

To summarize, these results suggest meaningful social interaction between close peers takes place on campus, where in place of attending lectures, students study together with their close peers. It seems that social interaction with high ability peers increases grades by increasing the productivity of (collaborative) self study. While only suggestive, these findings are consistent with laboratory evidence examining peer effects mechanisms. Kimbrough et al. (2017) find that low ability participants were able to solve more logic puzzles when allowed to socially interact with high ability participants, and audio recording revealed that these social interactions generated spillovers via peer-to-peer teaching.

2.7 Voluntary Sorting and Potential Implications for Group Assignment Policies

Our results indicate that peer effects in the classroom work through social proximity. The extent to which students sort out of their close peer groups over time, and become socially proximate with other, self-chosen peers, is therefore crucial to the evolution of peer effects from assigned close peers. For example, interventions aiming to help low ability students by matching them with high ability students may not be sustainable if the social proximity between these students wanes over time. Had classroom-level effects driven peer effects, any changes in social proximity would be of no concern.

In this section we analyze voluntary sorting and discuss its potential implications for group assignment policies. First, we track the effect of close peers on grades during the first year. We find that peer effects from close peers diminish over time; they are strongest in the first block and vanish by the fourth block of the first year. Second, we use detailed tutorial attendance data and present some evidence that the social proximity between close peers diminishes in a similar fashion during the first year. Third, we use second-year tutorial registration and confirm that students largely sort out of their assigned close peer group. Concurrently, students sort into new peer groups based on prior

[28]For example, these activities could include a fraternity membership, which is common among our student population living in the city. From the Dutch student survey "Studenten Monitor" we observe that students living outside of their parent's home spend in total roughly twice as much money on fraternity memberships and roughly 1.5 as much money on leisure activities than students living with their parents (`http://www.studentenmonitor.nl/`).

[29]Notice that this result is unlikely to be explained by non-city peers studying together in public transport. Column (3) in Table 2.9 has shown that spillovers are unrelated to having high ability peers closer to ones' residence.

Table 2.10: Peer Effects per Block

	Grades (Standardized)				
	Block 1	Block 2	Block 3	Block 4	Block 5
	(1)	(2)	(3)	(4)	(5)
Close Peer GPA	0.0404**	0.0361**	0.0318**	0.0080	0.0062
	(0.0177)	(0.0145)	(0.0145)	(0.0135)	(0.0178)
Own GPA	0.4139***	0.3451***	0.3849***	0.2537***	0.3026***
	(0.0159)	(0.0145)	(0.0176)	(0.0151)	(0.0160)
Observations	4271	4024	3650	3462	3329
Adjusted R^2	0.280	0.474	0.264	0.191	0.301

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Standard errors in parentheses, clustered on the tutorial level.
4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

bonds, ethnicity, and gender, but not on ability. While we cannot know with certainty the reason that academic spillovers from close peers vanishes during the first year, we believe the degree and type of voluntary sorting behavior provides an intuitive explanation.

2.7.1 Diminishing Peer Effects.

To study how peer effects evolve over time we repeat the analysis of close peer GPA on grades per block of the first year. The results are presented in Table 2.10 where the column number refers to the block being analyzed. Columns (1) and (2) reveal that during the first two blocks the estimates for close peer GPA are comparatively large (p-value<0.05). The magnitude slightly drops in block 3, while still being significant at the 5%-level. In blocks 4 and 5 spillovers become statistically indistinguishable from zero. Appendix Table A.2.9 shows that distant peers are unimportant throughout all five blocks. We investigate several potential explanations for the diminishing peer effects in the Appendix, such as differences in course types across blocks, direct effects of dropout, and measurement error in peer ability due to dropout. The results imply these explanations are unimportant.

2.7.2 First-Year Tutorial Attendance.

To study whether a reduction in social proximity could be a potential explanation, we analyze whether a student's tutorial attendance is associated with the attendance of their close and distant peers. Given a preference to attend tutorials with ones' friends, we interpret coordination of tutorial attendance among students as being indicative of social proximity. Let $Attendance_{isgct}$ be a binary variable taking the value one if student i attended tutorial session s, in group g, for course c, of cohort t. We run the following regression:

$$Attendance_{isgct} = \delta_0 + \delta_1 \overline{Att\,Close}_{(-i)sgc} + \delta_2 \overline{Att\,Distant}_{sgc} + G_{gct} + \delta_3 X_i + \epsilon_{isgct} \qquad (2.3)$$

Where $\overline{Att\,Close}_{(-i)sgc}$ and $\overline{Att\,Distant}_{sgc}$ are the proportions of individual i's close and distant peers who attend session s of course c. By running this regression per block, δ_1 and δ_2 detect any changes in attendance coordination as the first year progresses. Recall that across the seven weeks there are fourteen and seven tutorial sessions for large and small courses respectively.

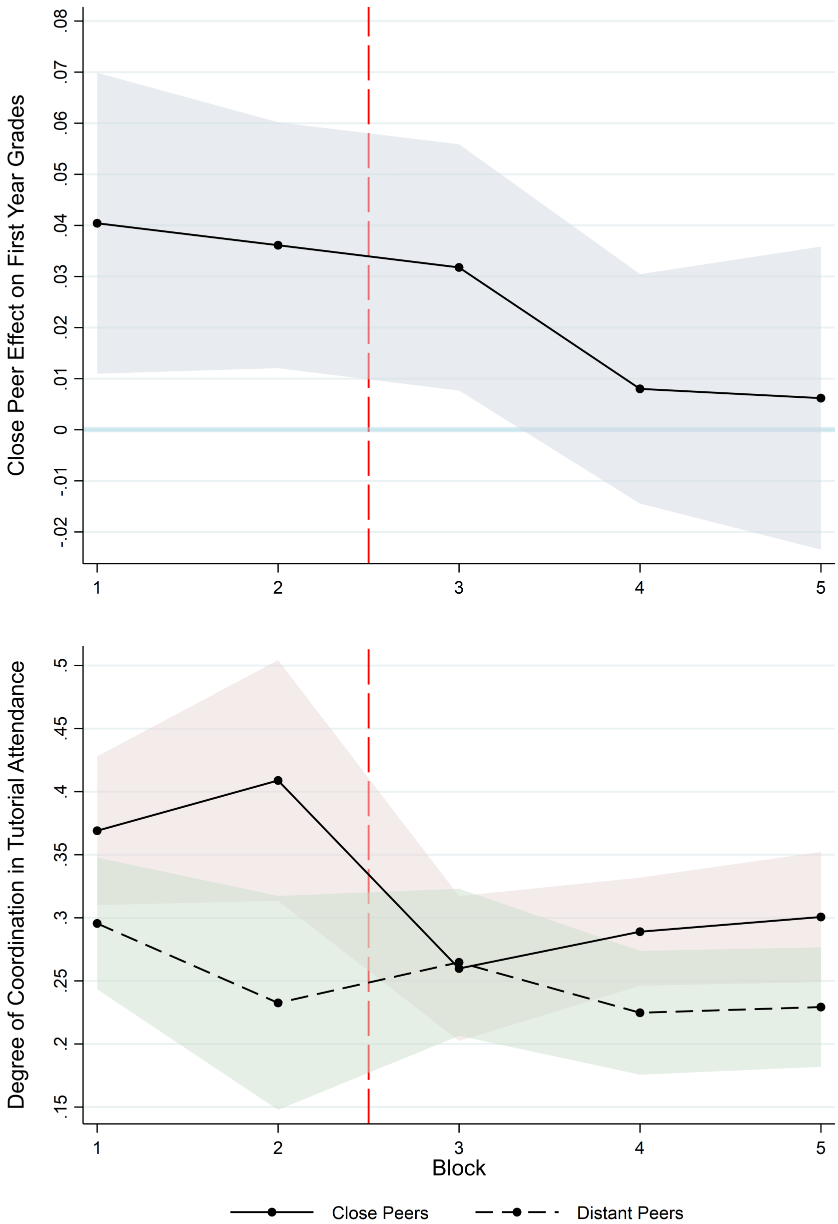
Equation (2.3) regresses attendance on its own group leave-out mean. If one is trying to detect causal peer effects this model would suffer from the reflection problem. Our goal, however, is to detect the degree of attendance coordination. Is a student more likely to go to tutorials with her close than distant peers, and does this change over time? The reflection problem poses no threat to answering this question. Another concern with such models is that group-level attendance shocks, such as bad weather, can result in coefficients that suggest peer coordination even if peers do not deliberately coordinate. Given that such shocks will take place at the tutorial level, both δ_1 and δ_2 are affected by these shocks. We will only compare their relative sizes and changes across blocks. Moreover, note that Equation (2.3) includes course-tutorial fixed effects (G_{gct}) to capture common shocks. The remaining control variables (X_i) are identical to the baseline regressions.

The results of the δ_1 and δ_2 coordination coefficients from these regressions per block are presented visually in Figure 2.3.[30] To highlight their potential relevance for the diminishing peer effects, the figure also contains a similar representation of the close group peer effect on grades during the first year.

We identify three main patterns. First, the degree of coordination in attendance between a student and her close peers is higher than between a student and her distant peers. This supports the notion that the close peer group meetings induced social proximity and further reinforces our results in Section 2.5. Second, the attendance coordination with close peers falls over time. Notably, the largest drop occurs after the second block, at which point there is a Christmas break. The timing of the break is indicated by the dashed vertical line in the figure. This drop is relatively large, significant (p-value=0.028), and stands in stark contrast to all other changes in coordination across blocks, which are relatively small and insignificant. Third, this drop in coordination after the second block is not

[30]The full regression results are presented in Appendix Table A.2.10. The table also presents p-values of a test for the equality of coefficients between adjacent blocks for close and distant peers separately and p-values of a test for the equality of the coefficients between close and distant peers within a block.

Figure 2.3: Diminishing Peer Effects in Grades (top graph) and Tutorial Attendance Coordination (bottom graph)



Notes:

1. Top graph shows the point estimates of close-peer effects on first year grades per block and the corresponding 90% confidence interval. The precise estimates can be found in Table 2.10.
2. Bottom graph shows the point estimates of first year tutorial attendance coordination for both close and distant peers and the corresponding 90% confidence interval. The precise estimates can be found in Appendix Table A.2.10.
3. The vertical dashed line indicates the timing of the first two-week break that occurs during the students' first year.

visible between students and their distant peers. We take this as evidence that, while there was initially a difference in the degree of social proximity between a student and her close and distant peers, this diminished over time. The Christmas break might have resulted in a severing of the bonds between close peers.

The results above provide some evidence that the social proximity between assigned close peers diminishes as time progresses. Are students sorting out of their close peer group into other groups? Second-year tutorial registration, which by then is under the purview of the students, provides us with an opportunity to analyse exactly this.

2.7.3 Second-Year Tutorial Choice.

All students in second year have to register for the tutorials a few weeks before the start of the course. If we assume that students prefer to be in a tutorial group with one's friends, then observing joint tutorial registration allows us to analyze peer group formation. We look for evidence of students co-registering based on shared characteristics; a phenomenon referred to as homophily. In particular, we use the following six characteristics: close and distant peer groups, ethnicity, gender, former bonds based on a student's high school, and ability (measured by high school GPA). Recall the program only admits Dutch students, and so students with a different ethnicity than Dutch are either first- or second generation immigrants. In the Dutch context, the categories European (81%, including Dutch), Arabic and Turkish (5%, referred to simply as Arabic from hereon in for simplicity), and Asian (14%) are ex-ante most relevant.[31]

Similar to the strategy of Marmaros and Sacerdote (2006), we first form all possible pairs of students who are observed to take course c in the second year of cohort t. Given N_{ct} students this procedure generates $(N_{ct} \times N_{ct} - 1)/2$ pairings of students.[32] Let $SecondYearTutorial(i,j)_{ct}$ be an indicator variable taking the value of one if both student i and j registered to the same second-year tutorial group and zero otherwise. We define a similar set of indicator variables for each of the characteristics listed above, taking the value of one if students i and j share that particular characteristic and zero otherwise. We then run the following regression per block:

$$SecondYearTutorial(i,j)_{ct} = \pi_0 + \pi_1 SharedCharacteristic(i,j) + C_{ct} + \epsilon(i,j)_{ct} \qquad (2.4)$$

[31] We determine ethnicity using the surname-based classification algorithm NamePrism (Ye et al., 2017). In the 1960s and 1970s many "guest workers" arrived to the Netherlands from North Africa and Turkey. Since these immigrants and their children are often referred to collectively as "allochtonen", we group students with these backgrounds together under the Arabic category.

[32] This is done by crossing the relevant list of student numbers with itself, removing all duplicate pairs (i, i), and keeping only one instance of the same pairing ((i, j) and (j, i)).

π_1 captures the change in the probability of two students sharing the same tutorial group in second year if they *e.g.* share the same gender. Equation (2.4) includes course-cohort fixed effects (C_{ct}), but, as the unit of observation is a student pair, it does not include other control variables. We cluster standard errors based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair (i, j).

Table 2.11 reports the results per block of the second year.[33] The last row reports the unconditional mean for the outcome variable, which is approximately 7 percent. The estimates for the shared characteristics can be compared to this mean, which reflects the probability of any two students registering together, independent of shared characteristics.[34]

The results reveal four main patterns. First, the Close Peer Group coefficient indicates that only some bonds from the close peer groups have remained up until the second year. The coefficient in block 1 is 0.06, which indicates that sharing a close peer group increases the probability of co-registration by 6 percent. As any two students have a 7 percent probability to co-register, a student registers together with 1 out of every 14 students. This becomes 2 out of 14 when the students originate from the same close peer group. Though this coefficient is significantly different from 0, it is far away from 1. This confirms the attendance results above and shows that by second year students have sorted out of their close peer groups to a large extent.

Second, a comparison of the Close Peer Group and Distant Peer Group coefficients reveals that, across blocks, the former is roughly 1.5 to 2 times larger than the later. The differences are statistically significant (p-values$<$0.05) and provide further evidence that the close peer group meetings indeed manipulated social proximity. Notice, however, that the Distant Peer Group estimates are also positive and statistically significant. Thus, while distant peers remain less important than close peers, a student is more likely to form bonds with her distant peers than with someone in a different first-year tutorial group altogether. Based on the estimate from block 1 of 0.035, on average students co-register with approximately 1 out of every 10 students from their distant peer group.

Third, Table 2.11 reveals that students sort into more homogeneous peer groups. Especially minorities, such as Arabic and female students are significantly more likely to appear in the same tutorial groups. The marginal effect of *e.g.* the coefficient for Both Arabic in block 2 indicates that two ethnically Arabic students are roughly 2 percentage points more likely to register together than *e.g.* an ethnic Arab with any other student. The largest predictor of co-registration is having shared the same high school, which increases the probability of being in the same second-year tutorial by roughly

[33] Notice that the last year we observe is 2014-15, and we therefore do not observed the second-year tutorial registration for the 2014 cohort.

[34] This unconditional mean coincides with our student-level data, where we observe roughly 200 students and 14 tutorials of 14 students each per course-cohort combination. As such, there is roughly a probability of 1/14 of registering in the same tutorial with any other student.

Table 2.11: Voluntary Sorting in Second-Year Tutorials (All Blocks)

	Same Tutorial in Second Year? Yes (1) or No (0)				
	Block 1	Block 2	Block 3	Block 4	Block 5
	(1)	(2)	(3)	(4)	(5)
Close Peer Group	0.0591***	0.0570***	0.0487***	0.0512***	0.0486***
	(0.0063)	(0.0052)	(0.0050)	(0.0054)	(0.0067)
Both Asian	0.0031	0.0029	0.0007	-0.0044	0.0127**
	(0.0049)	(0.0050)	(0.0053)	(0.0058)	(0.0055)
Both Arabic	0.0128	0.0227*	0.0258*	0.0578***	0.0273*
	(0.0108)	(0.0119)	(0.0136)	(0.0137)	(0.0160)
Both Europe	0.0043**	0.0044**	0.0020	0.0033	0.0056**
	(0.0019)	(0.0021)	(0.0019)	(0.0021)	(0.0023)
Both Female	0.0376***	0.0311***	0.0270***	0.0216***	0.0284***
	(0.0044)	(0.0044)	(0.0039)	(0.0040)	(0.0052)
Both Male	0.0010	0.0018	-0.0007	-0.0011	-0.0004
	(0.0021)	(0.0023)	(0.0022)	(0.0022)	(0.0028)
Same High School	0.0844***	0.0867***	0.0884***	0.0912***	0.0886***
	(0.0064)	(0.0072)	(0.0065)	(0.0066)	(0.0077)
Distant Peer Group	0.0350***	0.0396***	0.0385***	0.0342***	0.0309***
	(0.0063)	(0.0059)	(0.0056)	(0.0065)	(0.0067)
Both High GPA	-0.0009	-0.0023	0.0034	0.0039*	0.0030
	(0.0019)	(0.0025)	(0.0025)	(0.0024)	(0.0029)
Both Average GPA	0.0009	0.0030*	0.0015	-0.0017	0.0040
	(0.0016)	(0.0018)	(0.0016)	(0.0019)	(0.0025)
Both Low GPA	-0.0015	0.0000	-0.0012	0.0003	0.0038
	(0.0039)	(0.0045)	(0.0045)	(0.0043)	(0.0061)
Unconditional Mean	0.0756***	0.0755***	0.0764***	0.0766***	0.0746***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0008)
Observations	229428	214691	218183	188896	106630
Pseudo R^2	0.010	0.010	0.008	0.009	0.008

Notes:
1. All regressions include course-cohort fixed effects, other controls are excluded.
2. Block 5 has half the number of observations as one course does not have tutorials.
3. The unit of analysis is a student-pair. The outcome variable is one if both students in the pair registered for the same tutorial and zero otherwise. The explanatory variables are one if both students in the pair share the given characteristic.
4. Models are estimated with Probit. Marginal effects are reported.
5. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
6. Unconditional mean refers to the mean of the outcome variable in that particular block. Standard error reported in parentheses.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

8 percentage points across blocks. This supports our assumption that students seek to co-register to tutorial groups with existing friends, and rejects an explanation where student clustering is observed only due to shared preferences on the exact time at which the second-year tutorials are held.[35] Comparing the coefficients of the shared characteristic with the baseline unconditional mean of the outcome variable reveals that these effects are large. Sharing an Arabic ethnicity, or having shared the same high school, increases the baseline probability to register for the same tutorial by 33 and 110 percent respectively.

Fourth, the estimates for high school GPA reveal little to no co-registration of students with similar ability. The small degree of ability clustering is in line with the findings of Marmaros and Sacerdote (2006).[36]

2.7.4 Long-Term First-Year Bonds.

Which characteristics determine the long-term bonds that persist from the first-year close peer group? To investigate this, Table 2.12 shows results for a specification which includes interaction terms for each shared characteristic and the indicator for shared first-year close peer group. It appears that long-term first-year bonds are especially prevalent among close peers of the same gender; across the second year a pair of female (male) close peers are roughly 5 (3.5) percentage points more likely to form a long-term bond than a mixed gender pair. The estimates also reveal that long-term first-year bonds do not seem to be based on ability, which is consistent with our evidence of little to no clustering by ability presented above.

Note that in this specification the Close Peer Group coefficient provides a rare insight into the degree to which friendship groups can be institutionally manipulated against the formation of homogeneous subgroups based on gender, ethnicity, and prior bonds. More specifically, the coefficient measures the probability of co-registration among first-year close peers who share no observable characteristics. While relatively large and statistically significant in the first two blocks, as the second year progresses the coefficient diminishes in size and significance. This suggests that the manipulated social proximity further decreases in the long-term among students who differ on a wide range of characteristics.

[35]To this end, it is useful to note that across courses there are approximately two to three tutorial groups (of in total fourteen) taught at identical times. Thus, students with similar preferences regarding tutorial times could still register in different tutorial groups. We do not, however, observe the time of the second-year tutorial groups.

[36]We have performed a similar analysis for third-year course choice. The results of the regressions, per characteristic and for the ten most popular courses, are presented visually in Appendix Figure A.2.5. The conclusions are threefold. First, we find no evidence that close or distant peers choose the same courses in third year. Second, we find strong evidence of third-year course clustering based on shared high school, gender, and ethnicity. Third, in contrast to our results with the second-year tutorial registration, we find strong clustering based on ability. Taken together, this suggests that course choice also captures that students with some characteristics have preferences for certain topics, rather than reflecting bonding. For instance, high ability students sort into difficult courses.

Table 2.12: Characteristics of Long-Term First-Year Bonds (All Blocks)

	Same Tutorial in Second Year? Yes (1) or No (0)				
	Block 1	Block 2	Block 3	Block 4	Block 5
	(1)	(2)	(3)	(4)	(5)
Close Peer Group	0.0409***	0.0332***	0.0090	0.0190**	0.0129
	(0.0107)	(0.0093)	(0.0095)	(0.0091)	(0.0115)
Both Asian ×	-0.0299	-0.0711*	0.0122	-0.0069	-0.0808**
Close Peer Group	(0.0242)	(0.0387)	(0.0236)	(0.0262)	(0.0399)
Both Arabic ×	0.0034	-0.0264	0.0453	-0.0531	-0.0013
Close Peer Group	(0.0407)	(0.0530)	(0.0563)	(0.0693)	(0.0688)
Both European ×	0.0175*	0.0039	0.0116	0.0117	0.0028
Close Peer Group	(0.0097)	(0.0093)	(0.0079)	(0.0095)	(0.0106)
Both Female ×	0.0480***	0.0615***	0.0562***	0.0534***	0.0536***
Close Peer Group	(0.0149)	(0.0172)	(0.0150)	(0.0170)	(0.0180)
Both Male ×	0.0122	0.0283***	0.0391***	0.0351***	0.0416***
Close Peer Group	(0.0109)	(0.0109)	(0.0104)	(0.0099)	(0.0128)
Same High School ×	0.0221	0.0377	0.0143	0.0239	0.0537
Close Peer Group	(0.0254)	(0.0289)	(0.0324)	(0.0316)	(0.0343)
Both High GPA ×	-0.0175*	-0.0005	0.0006	-0.0155	-0.0209*
Close Peer Group	(0.0089)	(0.0107)	(0.0106)	(0.0103)	(0.0116)
Both Average GPA ×	-0.0078	-0.0023	0.0061	0.0025	0.0254**
Close Peer Group	(0.0086)	(0.0084)	(0.0089)	(0.0090)	(0.0102)
Both Low GPA ×	-0.0538**	-0.0308	0.0089	-0.0209	-0.0046
Close Peer Group	(0.0273)	(0.0195)	(0.0252)	(0.0237)	(0.0237)
Unconditional Mean	0.0756***	0.0755***	0.0764***	0.0766***	0.0746***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0008)
Observations	229428	214691	218183	188896	106630
Pseudo R^2	0.010	0.010	0.008	0.009	0.008

Notes:
1. Table shows results of a regression including all observable shared characteristics as predictors of shared second year tutorial, and with interactions between the shared characteristics and an indicator for shared first year close-peer group. Only results of shared first year close-peer group and the interaction terms shown.
2. All regressions include course-cohort fixed effects, other controls are excluded.
3. Block 5 has half the number of observations as one course does not have tutorials.
4. The unit of analysis is a student-pair. The outcome variable is one if both students in the pair registered for the same tutorial and zero otherwise. The explanatory variables are one if both students in the pair share the given characteristic.
5. Models are estimated with Probit. Marginal effects are reported.
6. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
7. Unconditional mean refers to the mean of the outcome variable in that particular block. Standard error reported in parentheses.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Policy makers and university administrators in both the U.S. and Europe have recently emphasized the importance of diversity in higher education. Table 2.12 implies that the group intervention mainly formed long-term bonds among students with similar characteristics and did little to promote long-lasting diversity on campus. We cannot rule out, however, that a more sustained or focused intervention would be more successful.

2.7.5 Implications of Voluntary Sorting for Peer Effects.

We have provided some evidence that the social proximity between close peers decreases during the first year. In turn, we have shown that by second year students' chosen peer groups hardly resemble their first-year assigned groups; they prefer to become socially proximate with others based on shared characteristics, such as gender and ethnicity. While we cannot know with certainty the reason that academic spillovers from close peers vanished during the first year, the voluntary sorting behavior provides an intuitive explanation. The social proximity between assigned close peers waned over time, which might have led to the corresponding decline in spillovers.

Another result is that students do not appear to choose their peers based on whether they are beneficial to their performance at university. We find no evidence of sizeable sorting by ability and close peers that stay together do not appear to base this choice on high school GPA. For instance, we do not find evidence that high (low) ability students sort into (out of) study partnerships with other high (low) ability students, though according to our peer effect estimates this would be academically beneficial. This stands in stark contrast to sorting based on other characteristics. Students are seemingly willing to trade off potentially higher grades in order to satisfy other preferences when choosing peers. Consistent with this, column (6) of Appendix Table A.2.13 shows that second-year chosen tutorial peers do not generate spillovers on student performance in second year.[37]

This might have further implications for group assignment policies. Policy makers may hope that targeted students would form new friendships with academically beneficial peers, thereby enjoying persistent peer effects despite sorting away from their assigned peers. Apparently this is not the case.

2.8 Conclusion

The promise of the peer effects literature is that simply reorganizing students among classes could increase overall student performance. Despite an abundance of papers aiming to properly identify

[37]Column (6) of Appendix Table A.2.13 shows that high school GPA of second-year chosen tutorial peers, while being instrumented with first-year assigned close peer GPA, has an insignificant effect on second-year grades.

spillovers, the literature has not yet delivered on this promise. A primary reason for this is our inability to understand the channels at work behind the various reduced-form estimates.

Our first set of results address this shortcoming. We focus on first-year student performance across six cohorts of economics undergraduate students at a large public university in the Netherlands. Students are randomly assigned to a tutorial group and one of two subgroups within their tutorial group. We take advantage of a university policy that stimulates social proximity within, and not between, these subgroups via a series of informal meetings at the start of the first year. We find the existence of spillovers on student performance that originate from students' socially proximate peers only. This implies that social proximity between peers, and the corresponding meaningful social interactions, are the driving force behind peer effects. Supplementary data suggests that these social interactions involve collaborative studying outside of class that occurs at university. Our non-linear estimates imply that alternative group assignment policies may result in aggregate, but not Pareto, improvements in performance.

The second part of this paper investigates the implications of voluntary sorting for group assignment policies. Given that peer effects arise due to social proximity, who students choose to become socially proximity with and how this evolves over time is crucially important. We first document that peer effects from assigned close peers diminish over time, and are completely absent by the end of first year. Using administrative data on daily tutorial attendance in first year and tutorial choice in second year we find that students increasingly sort out of their assigned close peer group into more homogeneous groups. This voluntary sorting behaviour foreshadows, and we argue provides an intuitive explanation for, the short-lived spillovers on student performance.

Similar to our analysis in Section 2.5, some researchers have used their reduced-form estimates on student performance to predict the effects of alternative group assignment policies (Booij et al., 2017) or to estimate the effects of optimal group assignment policies (Carrell et al., 2013). Such a practice usually assumes that there are no costs accompanying these effects. As our results imply that spillovers work solely through improving the productivity of (collaborative) self study, rather than through increasing teacher effort or decreasing leisure time, this assumption may be justifiable.

Our findings carry both good and bad news for those wishing to improve student performance using spillovers in similar settings. Encouragingly, it appears that a relatively uncomplicated and low cost intervention - in which social proximity between students was induced via several small meetings in the first weeks of university - could be used to generate larger spillovers than those previously observed. On a less encouraging note, we also find that any gains are likely to be short-lived, given that over time students increasingly sort out of their assigned peer groups. Social interactions appear to be too powerful to be constrained by a one-time manipulation of peer groups. A more sustained or intensive intervention may be necessary to ensure longer lasting benefits of group assignment policies.

2.A Appendix

Potential Explanations for the Decline in Peer Effects

Why do we observe that spillovers gradually diminish during, and become absent at the end of, the first year? Two overarching factors that vary during the first year, and could potentially drive the diminishing peer effect, are changes in the type of courses and dropout. Below we explore the evidence for each of these competing explanations.

Changes in the content, structure, and other characteristics of the courses during the first year could potentially drive the diminishing peer effect estimates. To explore this possibility, we look at heterogeneity in peer effects by course type. Following the classification of the university we group the ten first year courses into three categories: economics, business economics, and econometrics courses (see Appendix Table A.2.1). Appendix Table A.2.11 replicates our baseline specification while including an interaction between close peer ability and an indicator for course type, where economics courses are the baseline. The estimates reveal spillovers are statistically indistinguishable between the different types of courses. Feld and Zölitz (2017) reach similar conclusions, also at a public university in the Netherlands. In Lavy et al. (2012b) identification of peer effects is obtained via individual fixed effects together with the assumption that spillovers are the same across English, mathematics, and science courses. Note that Appendix Table A.2.11 does reveal the estimate of own high school GPA differs for the different types of courses. It appears that the returns to peer ability are disconnected from the returns to students' own ability.

It may be that the nature of the tutorial sessions changes from course to course, and that this has consequences for the existence of peer effects. Appendix Table A.2.1 reveals that the nature of the tutorial sessions is unrelated to the presence of peer effects. For example, tutorial descriptions are identical in Accounting and Microeconomics situated in block 1 and 2 and Marketing and Organisation & Strategy situated in block 4 and 5, while spillovers are only found in the former courses.

If courses get progressively easier during the first year, then the potential for peers to improve students grades could also diminish. To investigate this possibility Appendix Figure A.2.6 displays the coefficients on own and peer high school GPA per block, separately for small and big courses. Apart from a drop for the estimate on own high school GPA in block 4 for the big course (Marketing), the estimate for own GPA does not show a diminishing pattern across blocks. For instance, from block 2 to 3 this coefficient slightly increases, whereas the estimate for peer high school GPA decreases. Note that this evidence coincides with the results in Appendix Table A.2.11, where the returns to peer

GPA were detached from the returns to own GPA. Based on the three pieces of evidence presented above, we conclude changes in course type is an unlikely explanation for the diminishing peer effects.

A second potential explanation of diminishing peer effects is dropout. Indeed, Appendix Figure A.2.4 shows that dropout gradually increases as the blocks progress. Dropout could potentially reduce our peer effects estimates for at least two reasons; dropouts might be more responsive to peer high school GPA, and dropouts change the composition of the actual peer group for the remaining students. To investigate whether dropout interacts with the decline in peer spillovers, we repeat our robustness analysis of Section 2.5.4 and interact peer GPA with the number of course dropouts per close peer group for blocks 1 to 3 and block 4 to 5 separately. The results are presented in Appendix Table A.2.12. The estimates for close peer GPA imply that the decline in spillovers is present even in groups that did not experience any course dropout.

A consequence of dropout is that high school GPA of the initial close peer group becomes a worse measure of the actual ability of close peers. We overcome this potential problem by using an instrumental variable approach. For each student, per course, we calculate the average close peer GPA of only those close peers who are also observed to write the final exam for the course. We then repeat the regression of close peer GPA on first-year grades per block, while instrumenting the actual close peer GPA with the initially assigned close peer GPA. The first and second stage results of these regressions are presented in Appendix Table A.2.13. Panel A shows that assigned peer GPA is a strong instrument for actual peer GPA throughout the first year. Panel B shows that the decline in spillovers remains when using the variation in actual close peer GPA that originates from the assigned close peer GPA. From these results, we conclude that dropout is unlikely to be responsible for the diminishing peer effects during the first year.

Finally, Appendix Table A.2.14 repeats the analysis on lecture attendance and total study time for blocks 1 to 3 and blocks 4 to 5 separately. The negative estimate for close peer GPA on lecture attendance is only present in blocks 1 to 3. This suggests that the channel put forward in Section 2.6 - collaborative self study - is only present in the period for which we find significant peer effects.

Table A.2.1: Overview of the First Year

Course	Block	ECTS	Course Type	Tutorial Description	Exam Qs.
Accounting	1	8	Business Eco.	TA explains the assignments that students are supposed to make in preparation of the tutorial.	MC
Math I	1	4	Econometrics	TA briefly explains material of the week, after which the students make assignments in (unobserved) groups that count for a small percentage to the final grade.	Open
Microeconomics	2	8	Economics	TA explains the assignments that students are supposed to make in preparation of the tutorial.	MC (2009-11) and Open (2012-14)
ICT	2	4	Econometrics	TA briefly explains material of the week through predefined assignments, after which the students make assignments that count for a small percentage to the final grade.	Both MC and Open Qs.
Macroeconomics	3	8	Economics	TA explains the assignments that students are supposed to make in preparation of the tutorial.	MC
Math II	3	4	Econometrics	TA briefly explains material of the week, after which the students make assignments in (unobserved) groups that count for a small percentage to the final grade.	Open
Marketing	4	8	Business Eco.	TA explains the assignments that students are supposed to make in preparation of the tutorial.	Both MC and Open Qs.
Applied Statistics	4	4	Econometrics	TA briefly explains material of the week, after which the students make assignments in (unobserved) groups that count for a small percentage to the final grade.	MC
Organisation and Strategy	5	8	Economics	TA explains the assignments that students are supposed to make in preparation of the tutorial.	Both MC and Open Qs.
Financial Accounting	5	4	Business Eco.	TA explains the assignments that students are supposed to make in preparation of the tutorial.	MC

Notes: This is an overview of the year 2013-14. The courses are identical in the other years, with the only exception being that in 2014-15 the course “Skills” replaced “Financial Accounting”. Course type refers to the categorization made by the university. For a given course the use of tutorials only varies minor between the years. For all years and courses, the percentage of weekly tutorial assignments that count for the final grade is smaller than 20 percent. Mostly they count for 10 percent or they allow a student to make a bonus question on the final exam.

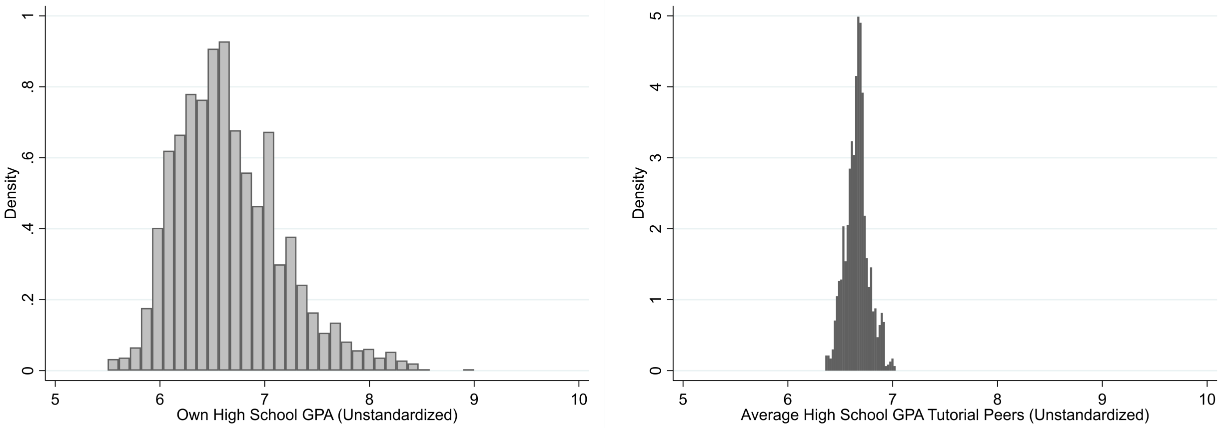
Table A.2.2: Overview of Categories and Questions in Course Evaluations

Question	Measurement scale	Category	Std.?
Objectives of course are clear	1-5	General	
Course is relevant for my studies	1-5	General	Yes
Course is interesting	1-5	General	
Course is well organized	1-5	Structure	Yes
Course material is understandable	1-5	Structure	
Can be completed within allocated study points	1-5	Fairness	
Time needed to complete exam is enough	1-5	Fairness	Yes
Exam reflects course content	1-5	Fairness	
Exam questions are clearly defined	1-5	Fairness	
Total study time (lectures+tutorials+self study)	1-10	Total study time	No
Have you attended lectures?	0-1	Lecture attendance	No
Lectures are useful	1-5	Lectures useful	Yes
Tutorials are useful	1-5	Tutorials useful	Yes
Lecturer is competent	1-5	Quality lecturer(s)	
Lecturer makes you enthusiastic	1-5	Quality lecturer(s)	Yes
Lecturer can be easily contacted	1-5	Quality lecturer(s)	
Lecturer provides sufficient assistance	1-5	Quality lecturer(s)	
TA gives good tutorials	1-5	Quality TA	
TA can be easily contacted	1-5	Quality TA	Yes
TA provides sufficient assistance	1-5	Quality TA	

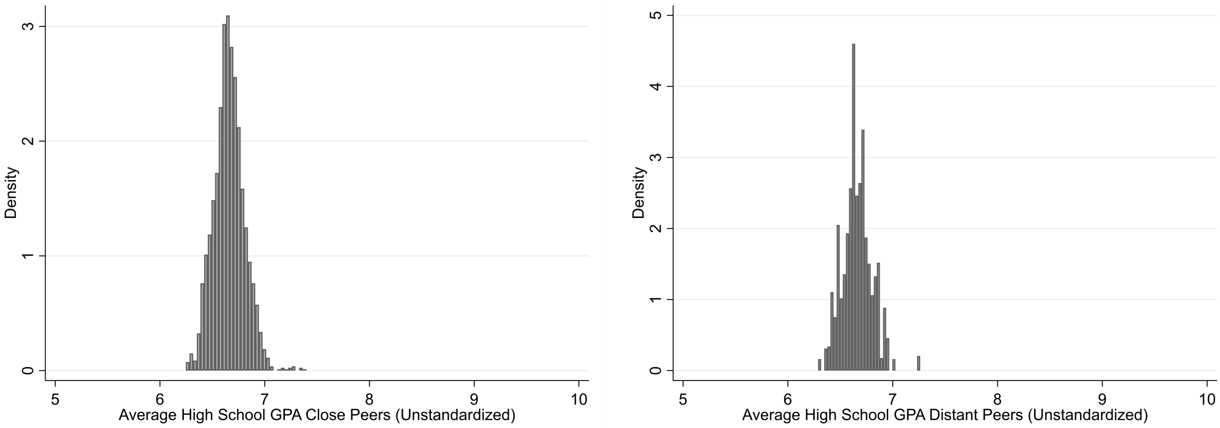
Notes: Questions are measured on a Likert scale, where 1 equals strongly disagree and 5 equals strongly agree, with the two exceptions being total study time (1 being 0 hours, 2 being $[1-5]$ hours, 3 being $[6-10]$ hours and 10 being ≥ 40 hours) and lecture attendance (1 being yes and 0 being no). We take the mean for questions within a category, ignoring potential missing values within a category. Std. refers to whether the (mean of a) category was standardized before the analysis.

Figure A.2.1: Histograms of High School GPA (Unstandardized)

(a) Own High School GPA (Left) and Tutorial Peer High School GPA (Right)



(b) Close Peer High School GPA (Left) and Distant Peer High School GPA (Right)



Notes:

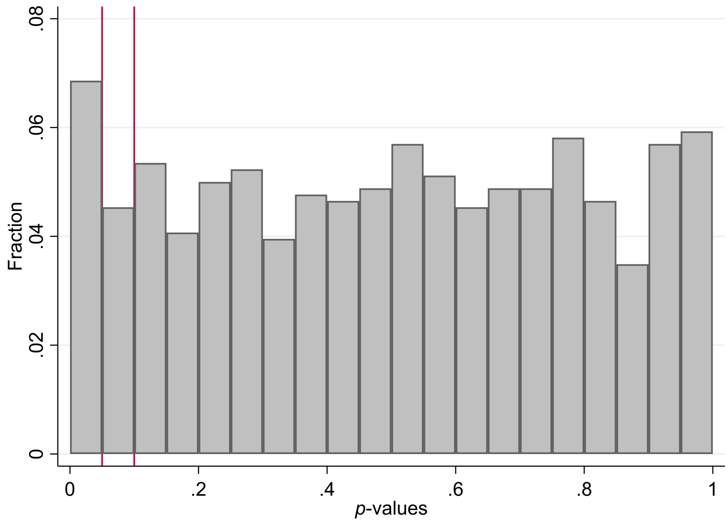
1. Figure shows histograms of student's own high school GPA, the leave-out mean for the tutorial- and close peer group, and the mean for the distant peer group.
2. In contrast to the leave-out mean for the close peer group, the mean for the distant peer group takes upon identical values for everybody in the same subgroup. This explains the somewhat more discrete nature of this figure. A histogram of the leave-in mean for the close peer group is similar to the mean for the distant peer group, where it would only change the peer-effects estimate on close peer high school GPA by a factor of $N_g/(N_g-1)$, where N_g is the size of close peer group g (Angrist, 2014).

Table A.2.3: Balancing Tests for Background Characteristics

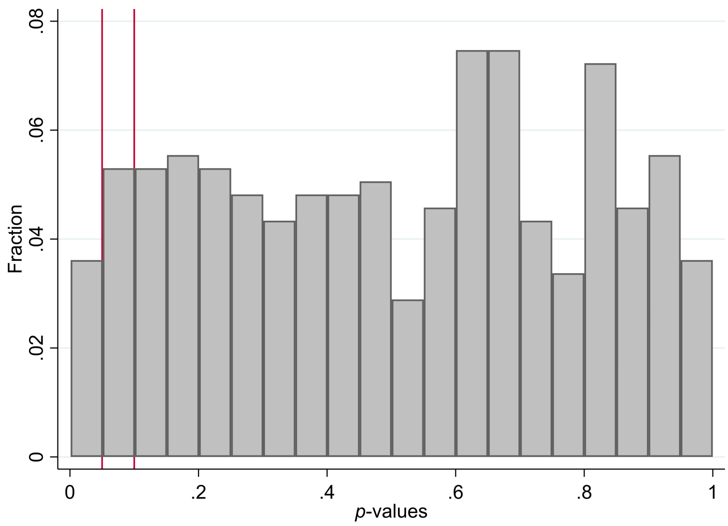
	Student Number	Gender	Age	Distance to University	High School GPA
	(1)	(2)	(3)	(4)	(5)
Close Peer Group 1	855.7164	0.0117	-0.7190	-5.8792	0.1693
	(3904.3123)	(0.1363)	(0.4737)	(9.0781)	(0.1668)
Close Peer Group 2	1578.1608	0.0117	-0.1796	-2.5438	0.0331
	(3904.3123)	(0.1363)	(0.4737)	(9.0781)	(0.1668)
Close Peer Group 3	-2206.2697	-0.0855	-0.3141	-7.1964	0.0785
	(4027.6709)	(0.1406)	(0.4887)	(9.3649)	(0.1721)
Close Peer Group 4	2209.5719	-0.0772	-0.5452	-3.3456	0.0247
	(4099.9048)	(0.1431)	(0.4975)	(9.5329)	(0.1752)
Close Peer Group 5	683.0497	0.0117	0.4360	13.8007	-0.0804
	(3904.3123)	(0.1363)	(0.4737)	(9.0781)	(0.1668)
Close Peer Group 6	257.0553	-0.1105	1.0621**	-1.9330	0.2936*
	(3802.7454)	(0.1327)	(0.4614)	(8.8419)	(0.1625)
Close Peer Group 7	-598.4830	0.0248	-0.3997	-1.4188	0.2320
	(3962.8418)	(0.1383)	(0.4808)	(9.2142)	(0.1693)
Close Peer Group 8	2902.1579	-0.0000	-0.5621	1.2335	0.3371**
	(3851.1900)	(0.1344)	(0.4673)	(8.9546)	(0.1646)
Close Peer Group 9	1121.8830	0.0117	-0.6279	0.0155	0.2005
	(3904.3123)	(0.1363)	(0.4737)	(9.0781)	(0.1668)
	⋮	⋮	⋮	⋮	⋮
Observations	2296	2296	2296	2296	2296
Adjusted R^2	0.832	-0.013	-0.005	-0.003	0.001
F-test	0.75	0.85	0.93	0.87	0.94
p-value	0.993	0.921	0.728	0.878	0.687

Notes:
1. Regressions include cohort fixed effects and dummies for the close peer group. No further controls are included.
2. The dependent variable is shown at the top of each column.
3. The F-test, and corresponding p-value, refer to a test for the joint significance of the close peer group dummies. It tests whether a large model with both cohort dummies and close peer group dummies can explain the background characteristics better than a small model with only cohort dummies.
4. Standard errors in parentheses.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2.2: Histograms of p-values of Balancing Tests



(a) p-values of Close Peer Group Dummies



(b) p-values of Tutorial Group Dummies

Notes:

1. Figures display histograms of the p-values of group dummies that originate from regressions where student characteristics are explained by group dummies.
2. The regressions were estimated for all student characteristics (student number, gender, age, distance to university, and high school GPA) separately for each cohort. The histograms include the p-values of all years and student characteristics combined.

Table A.2.4: Balancing Tests for TA Characteristics

	Is TA a PhD?	Is TA Female?
	Yes (1) or No (0)	
	(1)	(2)
Tutorial Peer GPA	-0.0041	-0.0148
	(0.0120)	(0.0199)
Own GPA	0.0005	-0.0042
	(0.0024)	(0.0036)
Observations	17535	6921
Adjusted R^2	0.254	0.345

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the tutorial peers. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Despite the two binary outcomes, we estimate the models with OLS. In some cases the course-cohort dummies predict the outcome variable perfectly, which means the Probit estimates for these dummies must be (minus) infinity.
4. Standard errors in parentheses, clustered on the tutorial level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

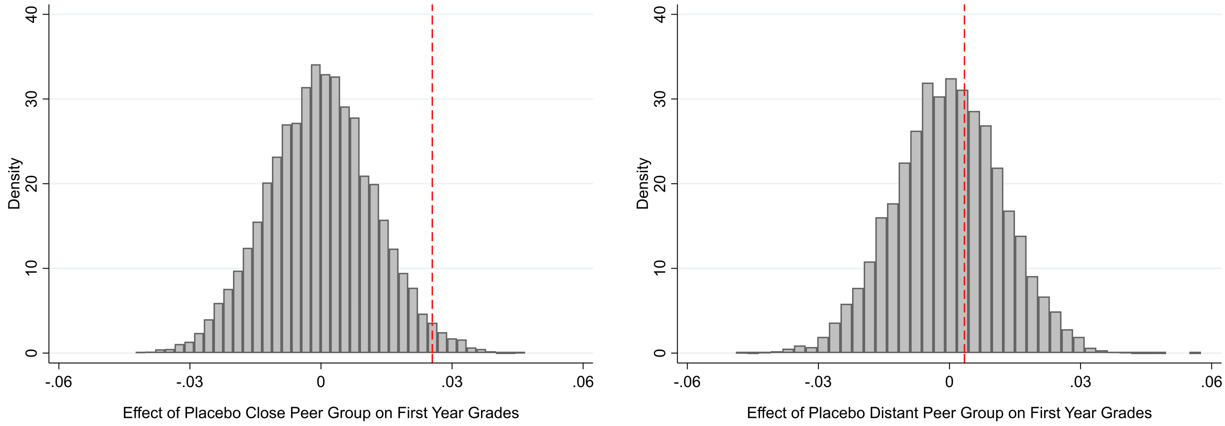
Table A.2.5: Randomization Inference and Exact p-values

	Simulated Mean (SD)	Estimated Value	Exact p-value
	Panel A: Separate Models		
Tutorial Peer GPA	-0.0002 (0.0128)	0.0191	0.1387
Close Peer GPA	-0.0001 (0.0123)	0.0255	0.0392
Distant Peer GPA	-0.0002 (0.0123)	0.0034	0.7852
	Panel B: Simultaneous Model		
Close Peer GPA	-0.0001 (0.0123)	0.0254	0.0408
Distant Peer GPA	-0.0002 (0.0123)	0.0008	0.9489
	Panel C: Simultaneous Model with Interaction		
Close Peer GPA	-0.00003 (0.0124)	0.0256	0.0412
Distant Peer GPA	-0.0002 (0.0124)	0.0010	0.9341
Close × Distant Peer GPA	-0.00003 (0.0126)	0.0150	0.2326

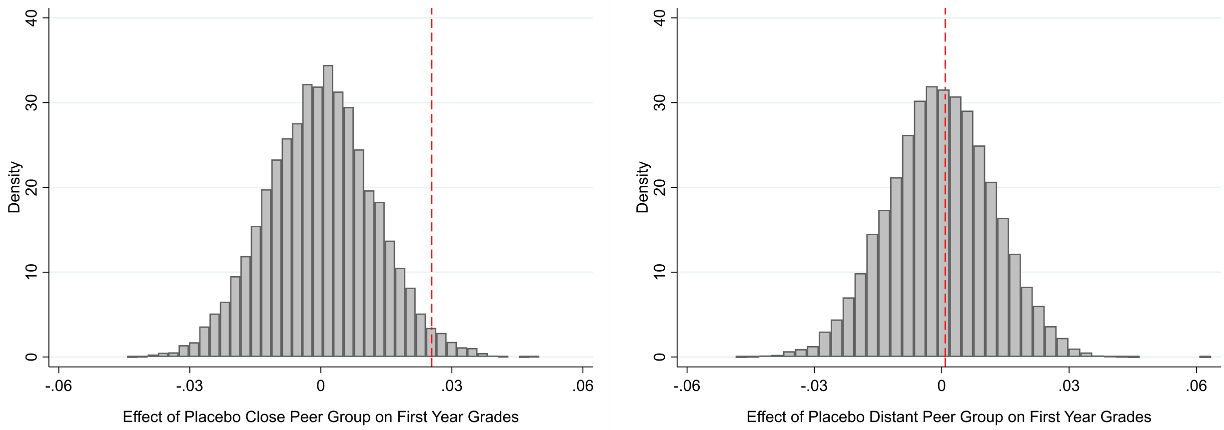
Notes:
1. Table summarizes the results of a randomization inference analysis of our baseline results presented in Table 2.3, in which we re-draw 10,000 alternative close- and tutorial peer group assignments. The table presents the mean and standard deviation of the coefficients under the 10,000 re-draws, the coefficient values under the actual assignment, and the exact p-value based on the randomization inference.
2. Panel A displays the results for models in which the peer GPA measures have been included separately. Panel B displays the results for a model in which the close and distant peer GPA measures have been included simultaneously. Panel C shows the results for a model in which the close and distant peer GPA measures, as well as their interaction, have been included simultaneously. Panels A, B and C correspond to columns (1) to (3), (4) and (5) of Panel A of Table 2.3, respectively.
3. The exact p-value shows the proportion of coefficients under the 10,000 re-draws for which a value at least as extreme as the actual value is observed.

Figure A.2.3: Histograms of Estimates Under 10,000 Group Assignment Re-draws

(a) Close Peer GPA (Left) and Distant Peer GPA (Right) on First Year Grades, Separate Models



(b) Close Peer GPA (Left) and Distant Peer GPA (Right) on First Year Grades, Simultaneous Model



Notes:

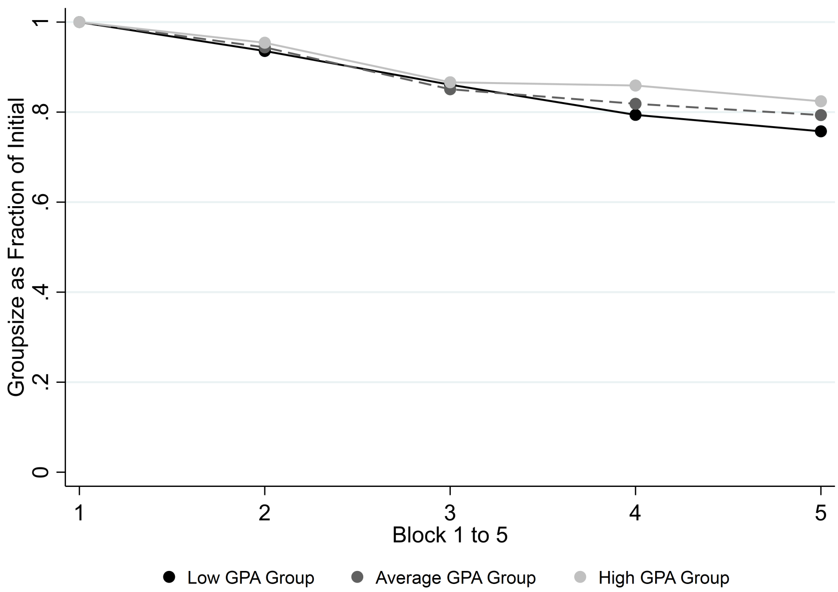
1. Figures show histograms of the estimates of close and distant peer GPA on first-year grades under 10,000 alternative group assignments.
2. Top figures (a) show results for models in which the peer GPA measures have been included separately. Bottom figures (b) show results for a model in which peer GPA measures have been included simultaneously.
3. Red dashed lines indicate the observed estimate under the actual assignment.

Table A.2.6: Peer Effects on First-Year Tutorial Attendance

	Attendance (% Tutorials Attended; Standardized)				
	(1)	(2)	(3)	(4)	(5)
Tutorial Peer GPA	-0.0122				
	(0.0176)				
Close Peer GPA		-0.0030		-0.0017	-0.0016
		(0.0131)		(0.0123)	(0.0121)
Distant Peer GPA			-0.0128	-0.0126	-0.0124
			(0.0149)	(0.0143)	(0.0143)
Close $\times$ Distant Peer GPA					-0.0146
					(0.0189)
Own GPA	0.0378***	0.0377***	0.0381***	0.0381***	0.0380***
	(0.0090)	(0.0092)	(0.0090)	(0.0091)	(0.0092)
Observations	18445	18445	18445	18445	18445
Adjusted R^2	0.122	0.122	0.122	0.122	0.123

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the tutorial- and close peers, and to the mean for distant peers. Own GPA refers to own high school GPA. All GPA measures are standardized. The outcome variable is the standardized percentage of tutorials attended per course.
3. Standard errors in parentheses, clustered on the tutorial level.
4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2.4: Course Dropout per Block



Notes:

1. This figure plots the number of students writing the final exam as a fraction of the initial students per block, separately for high, average, and low GPA close peer groups.
2. Low and high ability groups are in the bottom and top quartiles of close peer high school GPA. The average group refers to the middle 50 percent.

Table A.2.7: Balancing Tests for Non-linear Peer Ability

	Close Peer Group			Distant Peer Group		
	Share Low	Share Avg	Share High	Share Low	Share Avg	Share High
	(1)	(2)	(3)	(4)	(5)	(6)
Student Number	0.0017	-0.0007	-0.0010	0.0031	-0.0076	0.0045
	(0.0050)	(0.0071)	(0.0064)	(0.0047)	(0.0060)	(0.0052)
Female	0.0027	-0.0056	0.0029	0.0105**	-0.0085	-0.0021
	(0.0059)	(0.0074)	(0.0059)	(0.0048)	(0.0067)	(0.0063)
Age	-0.0008	-0.0001	0.0008	0.0007	-0.0025	0.0018
	(0.0027)	(0.0034)	(0.0029)	(0.0026)	(0.0030)	(0.0022)
Distance to	0.0025	-0.0027	0.0002	0.0029	-0.0019	-0.0010
University	(0.0021)	(0.0027)	(0.0023)	(0.0019)	(0.0024)	(0.0021)
Own GPA	0.0058**	-0.0050	-0.0008	-0.0032	0.0037	-0.0006
	(0.0024)	(0.0032)	(0.0036)	(0.0026)	(0.0025)	(0.0029)
Observations	2296	2296	2296	2296	2296	2296
Adjusted R^2	0.076	0.071	0.070	0.073	0.086	0.093
F-test	1.61	0.87	0.10	1.44	0.92	0.18
p-value	0.154	0.502	0.992	0.207	0.468	0.972

Notes:
1. All regressions also include cohort fixed effects.
2. The outcome variables are the (leave-out) proportion of low, middle, and high ability students separately for close and distant peer groups. Low and high ability students are defined as students in the bottom and top quartiles of high school GPA across the six cohorts, the remaining 50 percent is referred to as average ability. The dependent variables are unstandardized, where the independent variables are standardized except for the female dummy.
3. The F-test, and corresponding p-value, refer to a test for the joint significance of all the independent variables shown in the table.
4. Standard errors in parentheses, clustered on the tutorial level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.8: Peer Effects on Perceptions of Course using Course Evaluations

	Completed the Evaluation?	General	Structure	Fairness
	(1)	(2)	(3)	(4)
Close Peer GPA	0.0021	-0.0198	-0.0327	-0.0244
	(0.0084)	(0.0183)	(0.0254)	(0.0179)
Own GPA	0.0494***	0.0620***	0.0657***	0.1087***
	(0.0059)	(0.0151)	(0.0171)	(0.0165)
Observations	18736	3352	3352	3352
R^2	0.058	0.156	0.147	0.272
Binary Outcome	Yes	No	No	No

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. The dependent variable in column (1) equals one if the student completed the course evaluation and zero otherwise. The dependent variables in column (2) until (4) are the means of the answers to questions that embody the course characteristic showed in the top of the column. The dependent variables in column (2) until (4) are standardized.
4. Column (1) is estimated with Probit, the other columns with OLS. Marginal effects are reported. The R^2 refers to the Pseudo and Adjusted R^2 respectively.
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.9: Peer Effects for Distant Peers per Block

	Grades (Standardized)				
	Block 1	Block 2	Block 3	Block 4	Block 5
	(1)	(2)	(3)	(4)	(5)
Distant Peer GPA	-0.0159	-0.0014	0.0167	0.0064	0.0180
	(0.0199)	(0.0179)	(0.0168)	(0.0147)	(0.0184)
Own GPA	0.4133***	0.3442***	0.3836***	0.2534***	0.3021***
	(0.0159)	(0.0143)	(0.0176)	(0.0151)	(0.0158)
Observations	4271	4024	3650	3462	3329
Adjusted R^2	0.279	0.473	0.263	0.191	0.301

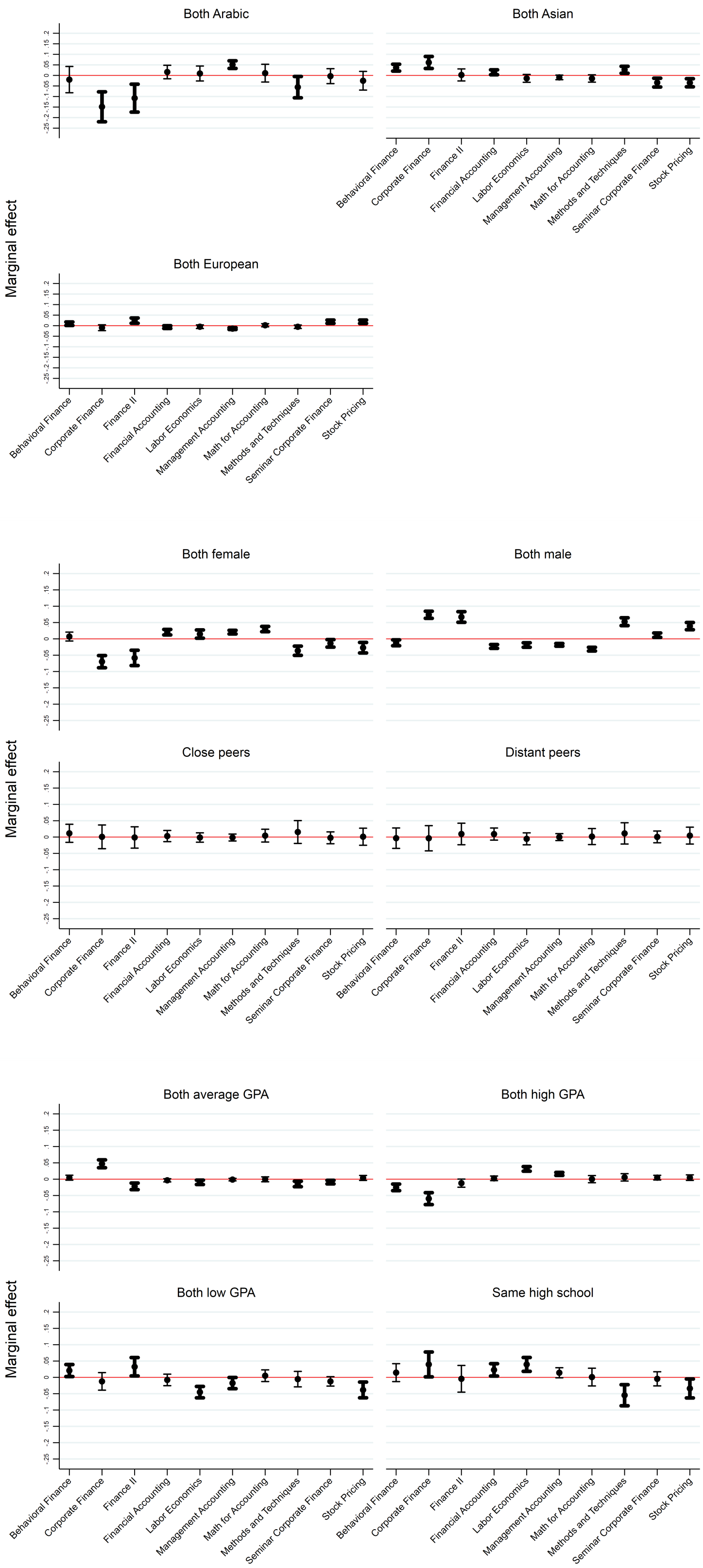
Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the mean of high school GPA for the distant peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. Standard errors in parentheses, clustered on the tutorial level.
4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.10: Coordination of First-Year Tutorial Attendance

	Attended Tutorial? Yes (1) or No (0)				
	Block 1	Block 2	Block 3	Block 4	Block 5
	(1)	(2)	(3)	(4)	(5)
Mean Attendance Close Peers	0.3690***	0.4089***	0.2598***	0.2890***	0.3007***
	(0.0355)	(0.0573)	(0.0345)	(0.0258)	(0.0310)
Mean Attendance Distant Peers	0.2956***	0.2326***	0.2647***	0.2247***	0.2293***
	(0.0314)	(0.0510)	(0.0350)	(0.0295)	(0.0285)
Observations	40321	40045	32920	33882	19654
Adjusted R^2	0.079	0.086	0.136	0.059	0.060
p-value Block $t = (t-1)$ Close		0.556	0.028	0.486	0.776
p-value Block $t = (t-1)$ Distant		0.287	0.578	0.335	0.908
p-value Close = Distant	0.173	0.078	0.889	0.107	0.133

Notes:
1. All regressions include course-tutorial fixed effects and controls; student number, gender, age, and distance to university.
2. Mean attendance refers to leave-out mean attendance per tutorial session for close peers and to the mean attendance per tutorial session for distant peers. The unit of analysis is on the student-tutorial-course level.
3. Block 5 contains somewhat less observations because the big course has 6 tutorials (one every week) instead of 13 to 14 tutorials (two every week).
4. The p-value "Block $t = (t-1)$" refers to a test for the equality of coefficients between adjacent blocks for close and distant peers separately. The p-value "Close = Distant" tests the equality of the coefficients between close and distant peers within a block.
5. The outcome is a binary variable, where the regressions are estimated with OLS. Our goal is to detect coordination in first-year attendance by relating the attendance of a student to her peers, we do not aim to estimate a causal peer effects regression. Probit estimates, and corresponding marginal effects, show qualitatively similar results.
6. Standard errors for the coefficients in parentheses, clustered on the tutorial level.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2.5: Voluntary Sorting in Third-Year Courses



Notes:

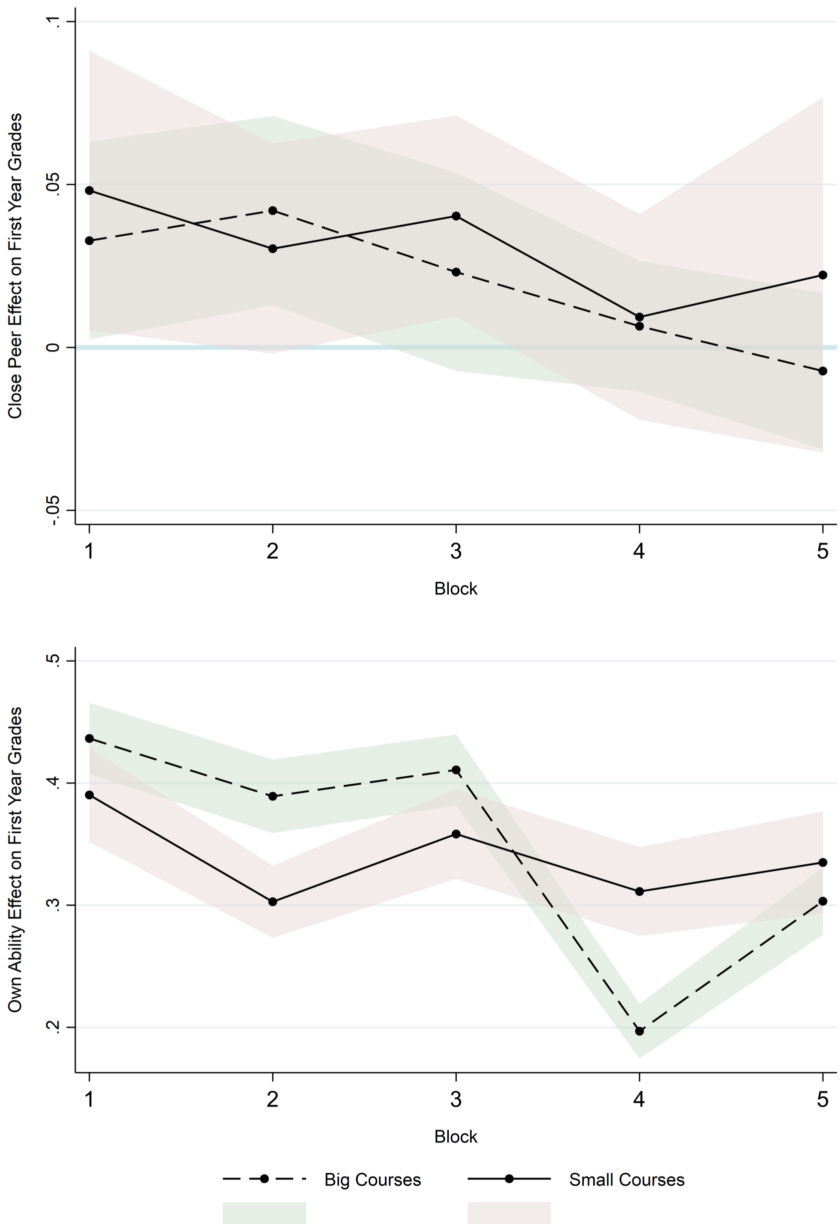
1. Figures display marginal effects and 90% confidence intervals of a Probit model that explains whether a student pair enrolled in the same course with their shared characteristics (*e.g.* both students in the pair are female).
2. The models are identical to the ones displayed in Table 2.11, only the binary outcome variable in this model is equal to one if a student pair enrolled in the same course and zero otherwise.
3. Significant marginal effects are made bold.

Table A.2.11: Peer Effects by Course Type

	Grades (Standardized)
	(1)
Close Peer GPA	0.0205*
	(0.0115)
Business Economics $\times$ Peer GPA	0.0002
	(0.0116)
Econometrics $\times$ Peer GPA	0.0120
	(0.0147)
Own GPA	0.3712***
	(0.0136)
Business Economics $\times$ Own GPA	-0.0535***
	(0.0115)
Econometrics $\times$ Own GPA	-0.0292**
	(0.0123)
Observations	18736
Adjusted R^2	0.323

Notes:
1. The regression includes course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. The dummy business economics is one for business-economics courses and the dummy econometrics is one for econometrics courses. The baseline consists of economics courses. Appendix Table A.2.1 shows which courses belong to which category.
4. The course dummies are not included as separate variables as they are a linear combination of the course-cohort dummies.
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.2.6: Effect of Peer GPA and Own GPA per Block



Notes:

1. Top graph shows the point estimates of close peer high school GPA on first year grades per block for big (8 ECTS) and small (4 ECTS) courses separately, and the corresponding 90% confidence intervals.
2. Bottom graph shows the point estimates of own high school GPA on first year grades per block for big and small courses separately, and the corresponding 90% confidence intervals.

Table A.2.12: Robustness of Peer Effects to Dropout Per Blocks

	Grades (Standardized)	
	Block 1-3	Block 4-5
	(1)	(2)
Close Peer GPA	0.0514***	0.0007
	(0.0162)	(0.0203)
Peer GPA $\times$ (Assigned-Actual)	-0.0092	0.0027
	(0.0061)	(0.0068)
Own GPA	0.3819***	0.2779***
	(0.0130)	(0.0138)
Observations	11945	6791
Adjusted R^2	0.352	0.257

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. The regressions include a measure for the difference between the number of students at the beginning of the year in the close peer group (assigned class size) and the number of students that wrote the exam for the course per close peer group (actual class size). This is a measure for course dropout and is not standardized.
4. The coefficient on close peer GPA measures spillovers in classes where there has been no course dropout (assigned-actual=0).
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.13: Instrumental Variable Analysis. Assigned Peer GPA is used as an Instrument for Actual Peer GPA

	First Year					Second Year
	Block 1	Block 2	Block 3	Block 4	Block 5	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Actual Peer High School GPA (First Stage)					
Assigned Close Peer GPA	0.8520***	0.8647***	0.8996***	0.8985***	0.9032***	0.0660***
	(0.0229)	(0.0223)	(0.0321)	(0.0360)	(0.0395)	(0.0214)
Own GPA	-0.0028	-0.0013	0.0006	0.0046	0.0043	0.0251*
	(0.0067)	(0.0063)	(0.0110)	(0.0129)	(0.0143)	(0.0139)
Adjusted R^2	0.918	0.865	0.801	0.745	0.723	0.166
F-test on Excl. Instrument	1386.07	1498.46	784.00	622.00	521.67	9.49
	Panel B: Grades (Standardized; Second Stage)					
Actual Peer GPA	0.0475**	0.0418**	0.0353**	0.0089	0.0069	0.1850
	(0.0205)	(0.0165)	(0.0162)	(0.0149)	(0.0196)	(0.3411)
Own GPA	0.4140***	0.3451***	0.3849***	0.2537***	0.3026***	0.3091***
	(0.0158)	(0.0144)	(0.0176)	(0.0150)	(0.0158)	(0.0203)
Observations	4271	4024	3650	3462	3329	10470
Adjusted R^2	0.280	0.474	0.263	0.190	0.301	0.196

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Own GPA refers to own high school GPA. All GPA measures are standardized.
3. In Panel A, the independent variable (Assigned Peer GPA) refers to the leave-out mean of high school GPA for the close peer group at the start of the first block in first year. This variable is used as an instrument for Actual Peer GPA, which is calculated on the course-cohort level and is equal to the leave-out mean of high school GPA for the close peer group in column (1) to (5) or for the tutorial peer group in column (6) while only taking into account the students who wrote the final exam of that course.
4. Panel B shows the results for the second stage, where Actual Peer GPA is the independent variable and has been instrumented with Assigned Peer GPA. The outcome variables are the standardized course grades for the first year per block in column (1) to (5) and for the second year pooled in column (6).
5. The number of observations for the second year in column (6) is lower than the baseline results for the first year. This is for three reasons; we do not observe the second-year grades of the 2014 cohort, students do not take all second-year courses in their second year, and for a small percentage we do not observe second-year tutorial choice.
6. Standard errors in parentheses, clustered on the tutorial level.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2.14: Peer Effects on Time Use per Blocks using Course Evaluations

	Attended Lectures		Total Study Time	
	Block 1-3	Block 4-5	Block 1-3	Block 4-5
	(1)	(2)	(3)	(4)
Close Peer GPA	-0.0223***	-0.0072	-0.1239	-0.3454
	(0.0083)	(0.0112)	(0.1922)	(0.2989)
Own GPA	-0.0118	-0.0176	-0.4963***	-0.6343***
	(0.0098)	(0.0124)	(0.1730)	(0.2050)
Observations	2995	1366	2995	1366
R^2	0.192	0.048	0.297	0.204
Binary Outcome	Yes	Yes	No	No

Notes:
1. All regressions include course-cohort fixed effects and controls; student number, gender, age, and distance to university.
2. Peer GPA refers to the leave-out mean of high school GPA for the close peer group. Own GPA refers to own high school GPA. Both GPA measures are standardized.
3. The dependent variable in column (1) and (2) is the answer to the question "Have you attended lectures?". The dependent variable in column (3) and (4) is the answer to the question "Average study time (hours) for this course per week (lectures+tutorials+self study)?" where we used the maximum for the interval to convert the categories into hours.
4. Column (1) and (2) are estimated with Probit, column (3) and (4) with OLS. Marginal effects are reported. The R^2 refers to the Pseudo and Adjusted R^2 respectively.
5. Standard errors in parentheses, clustered on the tutorial level.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 3

Going Dutch? Friendship Between Natives and Foreigners at University

Joint work with Matthijs Oosterveen

3.1 Introduction

Modern tertiary education can be characterized by an increased focus on internationalization. In addition to the development of international curriculum and an adoption of English as the language of instruction, the most notable consequence of this trend is a large increase in the admission of international and exchange students. In 2016, for instance, there were more than 1.6 million international students enrolled in universities across the EU.

However, the increased admittance of international students has not been without controversy. Critics hold that the vast increase in foreign students has led to less university places for native students, has put a strain on resources such as student housing, and that a switch to instruction in English in countries where it is not the native language puts local students at a disadvantage.

Conversely, supporters cite the large number of pro-social and educational benefits that foreign students transfer to local students. It is said that exposure to foreign students, and the accompanying cultural diversity, helps to reduce in-group bias, promotes international networks, and aids the development of intercultural skills. In the Netherlands for instance — one of the few EU countries where international student now make up more than 10% of the student population — the education minister wrote in a recent letter to parliament that the admission of international students helps native students deal with diversity, gain intercultural skills, and develop a global outlook (van Engelshoven, 2018). In a call to further promote internationalization in the sector, the Dutch education union claimed that through exposure to international students, local students "*learn to deal with other cultures, become familiar with traditions for other countries and are better off on the international labor market*".[1]

The scientific evidence for these arguments comes from the so-called contact hypothesis, a concept from psychology positing that interpersonal contact between individuals from different groups can reduce prejudice, promote friendships across a social divide, and aid inter-group relations (Allport et al., 1954). However, as the name suggests, any existence of these effects will depend crucially on the degree of contact between native and foreign students. If students self-segregate into homogeneous groups and have little meaningful contact with others from different backgrounds, then many of the purported benefits of an international campus will not be realized. Despite the lively debate surrounding the merits of the internationalization of education, there exists little evidence on the degree to which international and local students actually interact and form meaningful bonds, nor the degree to which universities themselves may be able to forge bonds between native and foreign students.

This paper aims to address this gap in the literature. Using administrative university data we seek to document the degree of interaction between native and foreign students. Rather than rely on self-reported measures of friendship, we elicit information about student friendships through their observed choices at university. Using actual student choices to infer dyadic ties avoids issues related to commonly used self-reported measures, such as miss-reporting and low response rate.

In addition, our setting allows us to study the extent to which universities can encourage native-foreign friendships through forced exposure. Opportunities to enforce exposure between students, such as in lectures halls and classrooms, are common at universities. These opportunities may therefore act as a natural remedy against self-segregation on campus. Similar interventions have been shown to aid inter-group student friendships in other contexts. We investigate the likely efficacy of such an intervention in encouraging native-foreign friendships.

Our administrative data comes from an economic bachelor program at a large university in the Netherlands, a country where the issue of the internationalization of higher education is being hotly debated. This program is taught in English and admits both students with a Dutch nationality as well as a sizable number of international students from a range of countries worldwide.[2]

The program has a unique structure that we exploit in order to elicit whether two students share a friendship. During the second year of this program, for each course, students must chose study groups; small groups in which course material is discussed. We observe both the study group choices of students for each course, as well as the exact time and date that these choices were made. Under the assumption that students who are friends will coordinate to register in multiple study groups together,

[1] `https://www.aob.nl/nieuws/kom-maar-op-met-die-buitenlandse-studenten-zegt-hoger-onderwijs`

[2] Our sample includes international students from 56 different countries. The three most frequent foreign countries are China (10.37%), Germany (9.86%), and Bulgaria (4.10%).

and that these coordinated choices will likely be made within a short period of time from one another, this data allows us to elicit proxies for the actual dyadic ties between students. In contrast to the second year, the first year study groups are allocated at random by the university, and do not differ per course. We use this random assignment to investigate the degree to which a year-long forced exposure within the close geographic proximity of the study groups can encourage subsequent native-foreign friendships that would otherwise not have occurred.

Our findings suggest that universities may be missing out on the full extent of the proposed benefits of the internationalization of education; interaction between local and international students should not be assumed to occur by default. When studying what appear as students' best, and likely most influential, friends we find that 40% (50%) of native (foreign) students have only native (foreign) best friends, far more than we would expect given the proportions of both student types on campus. Studying all dyadic ties, we find that native students are on average 20% less likely to form a friendship with a foreign student than a fellow native student. This preference remains even after controlling for potential differences in ability, gender, and other characteristics of the students.

Our study of the effects of the year-long first-year study group exposure shows that while being forced to share a small geographic proximity does increase the probability that a native-foreign friendship will occur, this only holds for the most "compatible" pairs of students. The more culturally distinct from the Netherlands the country of origin of the foreign student, the less likely it is that a friendship forms through exposure. Native students from municipalities with a high share of votes for a far-right anti-immigrant political party sort away from foreign students they are exposed to in their first year. Finally, we find no evidence that native (foreign) students forced into first year study groups with a large number foreign (native) students go on to make friends with foreign (native) students outside of this study group.

While relevant for the debate surrounding the internationalization of education, we believe these results also hold importance for broader topics. Firstly, a large body of research has shown the importance of friends in determining various outcomes, academic and otherwise. Insights into the mechanisms governing interaction between individuals are therefore crucial to policies aiming to induce peer effects. Indeed, attempts at harnessing peer effects have failed due to a lack of understanding about the degree to which policies can encourage friendships between diverse peers (Carrell et al., 2013).

Second, these results relate to the literature concerned with the effects of exposure to diversity at university, one aspect of which is diversity based on nationality. Studies have shown that friendships between students of different backgrounds can result in pro-social changes in attitudes and behavior (Boisjoly et al., 2006; Carrell et al., 2019). However, little is known about how to encourage bonds between such students, especially students of different nationalities. This paper adds important evidence on the formation of native-foreign ties.

Finally, we believe these results speak to the recent discussions surrounding migration trends and the rise in xenophobic sentiments. For countries within the Europe Union, for instance, the free movement of people and an influx of asylum seekers has meant that these societies have become increasingly international and multicultural. As evidenced by the rise of far-right anti-immigration parties across Europe, there exists a sizable amount of animosity to foreigners in some countries.[3] The negative economic and social impacts of such animosity can be widespread. Insights which can inform policies aiming to reduce xenophobia - such as this paper's investigation into the degree to friendships between diverse individuals can be encouraged - are therefore important.

The remainder of this paper is structured as follows. The subsequent section briefly summarizes the existing evidence on student sorting behaviours at university and other closely related literature. Section 3.3 describes the institution context. Section 3.4 outlines the data used in this paper, and our method for eliciting friendship information from student choices. Section 3.5 presents results on the sorting behaviour of natives and foreign students. Section 3.6 investigates the effect that a year-long forced exposure has on the occurrence of native-foreign friendships. Section 3.7 concludes.

3.2 Existing Research

To the best of our knowledge, there is currently no quantitative research documenting the general friendship patterns of students in European countries, and no evidence on the sorting patterns of native and foreign students. Existing evidence in the economics literature has so far come only from students in US universities, and has focused on interactions between "white" a "black" students.

In a discussion that mirrors the disputes around international students highlighted in this paper, the benefits of so-called "affirmative action" admission policies in the US have also been debated. Proponents of such policies claim that increased racial diversity is beneficial to all students, through a reduction in segregation, increased understanding between groups and reduced discrimination.[4] Motivated by the fact that such benefits will not be realized if black and white students have little actual interaction, researchers have set out to document the level of friendship between students belonging

[3]The average vote share for such parties in Europe has risen from 5% in 1997 to 16% in 2017 (`https://www.bloomberg.com/graphics/2017-europe-populist-right/`).

[4]In a recent United States Supreme Court decision that ultimately ruled in favor of affirmative action policies (*Fisher vs. University of Texas*), the majority judges wrote that the affirmative action admission requirements were justified based on their "*destruction of stereotypes*", their promotion of "*cross-racial understanding*" and because they prepare students "*for an increasingly diverse work force and society*".

to these groups. Given the similarity to the current paper in both motivation and methodology, a selection of important papers in this literature are summarized below.

A seminal paper examining student sorting patterns is Marmaros and Sacerdote (2006). Using the volume of email communication between students to elicit friendship ties, the authors aim to document the determinants of student friendship at a US university. Their findings are consistent with homophily; students tend to form bonds with those who are similar to them. In particular their results suggest a large degree of racial segregation. On average, two white students will interact three times as much as a black-white pair. Students at this university are also allocated to college dorms in a random fashion. Exploiting this, the authors highlight the importance of forced exposure to friendship formation. Two students randomized into the same dorm, regardless of race, become three times more likely to form a friendship.

Foster (2005) and Mayer and Puller (2008) also document student sorting patterns at university. The former uses students' roommate choice to study the determinants of friendship. The latter use friendship data from Facebook to simulate the likely effectiveness of policies aimed at reducing social segmentation. Both papers' results largely confirm those of Marmaros and Sacerdote (2006); they document large racial sorting patterns, and highlight the role of exposure for subsequent friendship.

Others have sought to provide some explanations for the high levels of sorting between diverse students. Using data from selective universities in the US, Arcidiacono et al. (2013) suggest that the academic miss-match between black and white students drives a wedge between the groups and limits interaction. Camargo et al. (2010), however, show that segregation persists even when controlling for potential differences in ability between students. Using roommate allocation, their analysis shows that black and white students are in fact highly compatible as friends once introduced. They tentatively suggest that misperceptions about friendship compatibility between students of different groups may drive segregation.

There is also a highly related strand of literature that goes one step further by investigating if the purported benefits of exposure to diversity at university exist at all. This emerging literature examines if exposure to minorities or marginalized groups changes students' attitudes, behaviors or preferences in a manner consistent with the contact hypothesis. While we do not explicitly examine changes in attitudes or preferences, the current paper also contributes to this larger literature by examining the effects of diversity on friendship formation at university.

For instance, both Boisjoly et al. (2006) and Burns et al. (2013) use roommate assignment to study how being matched with a minority black student changes the beliefs of white students. Carrell et al. (2019) study how exposure to black squadron members during their first year affects the second-year roommate choice of white squadron members at the U.S Air Force Academy. In general, these studies consistently find evidence in line with the predictions of the contact hypothesis.[5]

3.3 Institutional Context

We study the sorting patters of native and foreign students within a three year economics undergraduate program at a large Dutch public university. This program is taught in English, and is open to both Dutch and foreign students. The academic years are divided into five blocks. One block lasts eight weeks; seven weeks of teaching and one week of exams.[6]

3.3.1 First Year.

During the first year, students have to follow ten compulsory courses covering the basics of economics, business, and econometrics. Students have to follow one light and one heavy course per block, which make up four and eight credits respectively. Sixty credits account for a full year of study (5 blocks $\times$ 12 credits).[7] Instruction takes place via lectures and study groups, which both last for 1 hour and 45 minutes. Lectures are of large scale, where the course material is explained with little to no interaction between students and lecturer or between students themselves. The study groups contain roughly 24 students, are guided by a teaching assistant (TA), and mostly consist of active discussion about course related assignments. The TA's are either senior students or PhD students.

Students are assigned to a first-year study group in a random fashion, and must follow all study groups for all first-year courses with this group. Details of this allocation process are presented in Appendix Section 3.A.[8] Also presented are balancing tests that cannot reject that the final allocation of students to first year study group was random.

There are two study group sessions per week and one study group session per week for the heavy courses and light course, respectively. This amounts to 5 hours and 15 minutes of forced exposure time to study group peers per week. Students must attend 70 percent of the study group sessions per course. If they do not meet this requirement, they are prohibited from taking the exam and must wait a full year before they can take the course again. Our data shows that compliance with the 70 percent

[5]Recent economic investigations of the contact hypothesis have also taken place in contexts outside the university. See for instance Dahlberg et al. (2012); Nekby and Pettersson-Lidbom (2017); Steinmayr (2016); Schindler and Westcott (2018).

[6]There is also a resit period at the end of the academic year. This takes place in the few weeks after block five, and is the only opportunity for students to resit exams.

[7]At this university, credits are referred to as European Credit Transfer System (ECTS). This is the common measure of student performance in Europe to accommodate credits across institutions. One ECTS is equivalent to 28 hours of study.

[8]Briefly, students are randomly sorted on a list upon arrival at the university on the first day of the program. Study group membership is then determined from this list on a rotating basis.

rule is almost perfect. This confirms that students have a significant amount of exposure to and interaction with their first-year study group peers. These interactions are not only academically oriented. Students' transition from high school to university involves disruptions of old friendships and the creation of new bonds (Thiemann, 2017). The tutorial groups provide one of the first opportunities for students to create a network and interact with others in their program.

3.3.2 Second Year.

The structure of the program's second year is identical to the first year. Students follow ten compulsory courses, one light and one heavy course per block, which can be considered as followups of the first-year courses. The key difference from the first year, however, is that study group choice is now under the purview of the student. A few weeks before the start of each block, students must electronically register for a study group for both courses in that block. The lone exception to this is the final block of the second year, in which one of the two courses has no study groups. We observe these choices, as well as the exact time and date (down to the second) at which students make their study group choice for each course. Registering for the same study group for both courses in a block amounts to more than 5 hours of forced exposure peer week.

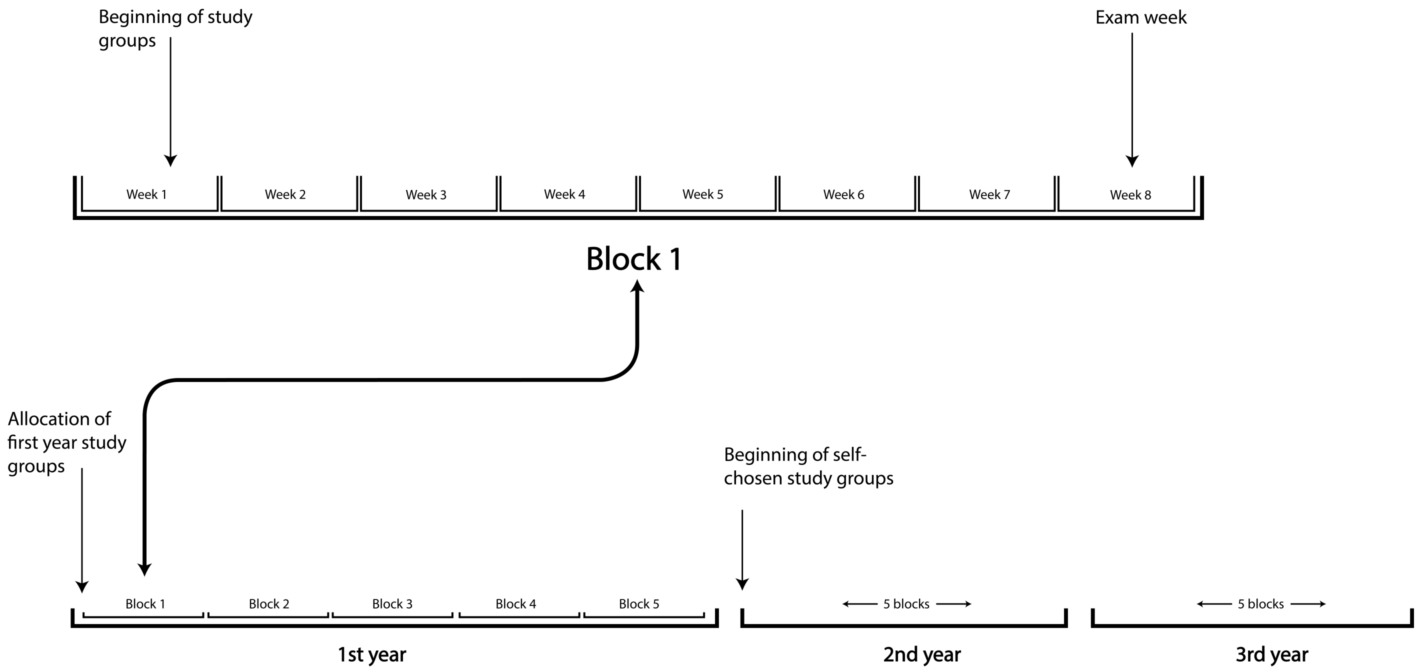
For two second-year courses students must also form working groups of 3 to 4 members, within which they complete important group assignments. These working groups entail a significant amount of interaction and collaboration. For one of these courses, for instance, students must write a report together and spent a minimum of 10 hours on the project per week.

Since both empirical and anecdotal evidence suggest that students strategically coordinate their study and working group registrations in order to register in the same group as their friends, we use both the self-chosen second-year study groups and working groups to elicit information about student friendships. Exactly how we use this information is discussed in more detail in Section 3.4.2. A visual overview of the first block and the entire three year program is presented in Figure 3.1.

3.4 Data

Our main source of data is the administrative database of the university between the academic years 2008-9 and 2014-15. We observe all incoming students for these 7 cohorts, their courses taken, and study group chosen across these years. We have also extended this data with the working group choice for two courses in the second year. Additionally we observe a rich set of student characteristics; gender, age, residential address, high school location for the majority of Dutch students, and nationality as stated on their passport.

Figure 3.1: An overview of the characteristics of the undergraduate Economics program relevant to our study



3.4.1 Descriptive Statistics.

Our main sample is the 781 students across 7 cohorts who are observed to take at least a full block in the second year. Summary statistics of this sample are given in Table 3.1. Panels A and B present the characteristics of the native and foreign students, respectively, per cohort. While the majority of students are male, there are a higher proportion of women in the foreign student sample, and foreign students are marginally older. Panel C summarizes the origin of the student sample per cohort. The three most represented countries are The Netherlands (*i.e.* native students), China, and Germany. A small fraction of students originate from outside both Europe and Asia.

3.4.2 Eliciting Friendships.

Recall that for each course in the second year students must chose their own study group. Students are able to begin registering for these groups by computer approximately two weeks before the start of a block. Depending on the number of students taking a course and the size of the study groups, there are between three and five groups to chose from per course. Importantly, students are unable to see which other students are already registered in each study group before registering themselves. In addition, we also observe students' working groups for two courses in their second year. While, due to their larger number, our main source of friendship data will be based on the study group choices, we will also present results using the working group data to provide additional evidence.

Table 3.1: Descriptive Statistics per Cohort

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15	
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Total
Panel A: Native Sample								
Female	0.35 (0.49)	0.47 (0.51)	0.33 (0.47)	0.35 (0.48)	0.35 (0.48)	0.27 (0.45)	0.42 (0.5)	
Age	19.23 (1.04)	18.94 (0.78)	19.31 (1.59)	19.08 (0.6)	19.06 (0.89)	19.14 (1.01)	18.95 (0.62)	
Distance (km)	23.58 (23.29)	32.38 (39.59)	31.72 (38.49)	35.56 (45.52)	33.81 (33.44)	23.22 (27.19)	29.1 (37.16)	
No. of Second Year Courses	9.13 (1.57)	9.07 (1.41)	8.04 (1.24)	8.92 (1.59)	8.65 (0.94)	9.64 (0.8)	9.77 (0.9)	
No. of Students	31	30	49	51	68	56	65	350
Panel B: Foreign Sample								
Female	0.4 (0.49)	0.53 (0.51)	0.38 (0.49)	0.39 (0.49)	0.35 (0.48)	0.35 (0.48)	0.48 (0.5)	
Age	20 (1.56)	20.13 (1.49)	20.44 (1.89)	20.47 (1.91)	19.91 (1.86)	20.13 (1.81)	19.35 (1.13)	
No. of Second Year Courses	9.11 (1.26)	9.21 (1.25)	8.66 (0.86)	9.22 (1.26)	8.76 (0.64)	9.81 (0.93)	9.92 (0.39)	
No. of Students	55	34	56	59	71	79	77	431
Panel C: Place of Origin (Proportion)								
Dutch	0.36	0.47	0.47	0.46	0.49	0.41	0.46	0.45
Chinese	0.15	0.16	0.16	0.10	0.08	0.07	0.07	0.10
German	0.13	0.13	0.08	0.07	0.13	0.11	0.06	0.10
Europe East	0.17	0.14	0.14	0.21	0.13	0.19	0.18	0.17
Europe North (Non-Dutch)	0.16	0.16	0.11	0.15	0.16	0.17	0.15	0.15
Europe South	0.03	0.02	0.05	0.01	0.02	0.07	0.05	0.04
Asia	0.22	0.20	0.19	0.15	0.15	0.12	0.12	0.16
Other	0.05	0.02	0.04	0.03	0.05	0.04	0.04	0.04

Notes:
1. Table shows the mean and standard deviation per cohort of native (Panel A) and foreign (Panel B) student characteristics. Panel C shows the proportions of the country of origin of students per cohort.
2. Age is evaluated on January 1^{st} in the academic year that the cohort started. Distance to University refers to the number of kilometers from a student's registered address to the university.

Using the study group data, we take all students who are observed to have taken both courses in block b of the second year of cohort t. Given N_{bt} students, each student then has $N_{bt} - 1$ potential friends in block b. Under the assumption that friends will coordinate their tutorial registrations so that they end up in the same classes, we classify two students as being friends in a particular block b if both students choose the same study groups for all courses in that block, *and* if, for each course in the block, the student's registration decisions are made within 12 hours of one another. The latter condition is designed to ensure as far as possible that two students who choose the same study groups are indeed coordinating without being too restrictive, given that under coordination registration decisions will likely occur within a small time period from one another.[9] We repeat this process for each of the 5 blocks of the second year. As stated earlier, block 5 only has one course in which study groups are used. We therefore only use coordination decisions within one study group to elicit friendships in this final block.

There are two courses in the second year for which we have information on the working groups chosen by students. Using this data, we take all students in cohort t who are observed to have taken the second year course c. Under the assumption that students prefer to work on group assignments with their university friends, we classify two students as friends if they appear in the same working group. We repeat this process for both of the courses that contain working groups.

Table 3.2 displays a range of summary statistics for the friendship data elicited using the methods above. The study group data reveals that students coordinate with on average approximately 7 fellow students per block. This amounts to on average approximately 25 unique friendships across the whole second year.[10] Using the working group data, we identity approximately 3 friends per student per course, which amounts to a total of roughly 5 unique friends.

[9] For instance, it is common for students to share their registration decisions with their friends over text message, or to register together simultaneously on adjacent computers.

[10] Note that the larger number of friends in block 5 may be at least partly explained by the fact that we observe only one study group in this block.

Table 3.2: Descriptive Statistics of Friends from Study and Working Group Data

	Block 1	Block 2	Block 3	Block 4	Block 5	Total
	Panel A: Mean Number of Friends from Study Group Data					
Native Students	5.21 (7.40)	6.59 (6.68)	5.84 (5.20)	5.83 (6.13)	12.84 (10.20)	25.51 (17.15)
Foreign Students	4.88 (6.81)	5.57 (6.35)	4.73 (5.01)	6.58 (7.03)	11.44 (9.84)	24.62 (17.07)
	Panel B: Mean Number of Friends from Working Group Data					
	Course 1	Course 2				Total
Native Students	2.83 (1.04)	2.64 (0.66)				4.60 (1.47)
Foreign Students	2.92 (0.95)	2.65 (0.65)				4.63 (1.48)

Notes:

1. Tables shows the mean and standard deviation of the number of friends per block, separately for native and foreign students.

2. Panel A shows the results from the study group data. Panel B shows the results from the working group data

3.4.3 Do We Capture Friendship?

A potential concern with eliciting friendship data using students' choices regarding their study and working groups, rather than self-reported data, is that coordination behavior does not amount to friendship. For the reasons stated below we believe that our measures of friendship are valid.

Firstly, the fact that the friendship patterns we observe using our data are in line with what one would expect to see provide reassurance of the accuracy of these measures. Appendix Table A.3.1 shows how the probability of two students being classified as friends using the study group pair data (column (1)) and the working group data (column (2)) depends on various characteristics.[11] These results reveal for both data sources that sharing a gender, country, or academic ability significantly increases the probability that two students are classified as friends. The absolute difference in age is also a significant predictor of friendship in the study group pair data. These patterns are in line with generally observed patterns of homophily in networks. A further test of the accuracy of these measures is examining if our friendship measures are sensitive to students who may have an existing relationship before entering university. It seems sensible to assume that students from the same high school are more likely to be friends. The results using our measures are in line with this; as Appendix Table A.3.1 shows, by far the strongest predictor of being classified as friends is having attended the same high school.

[11] Section 3.5.2 describes these specifications in further detail.

Secondly, if two students are indeed coordinating, it is less important if they actually consider themselves as friends due to the consequences of this coordination. Students who end up in one another's study and working groups are forced to spend a significant amount of time in each others presence working and studying together. Our interest is in capturing students that socially interact. Thus, regardless of their perceived friendship status, we are ultimately interested in coordinating students. Note that self-reported data, on the other hand, may incur its own measurement problems. The perceived strength of a friendship may differ from individual to individual, and students who report themselves as friends are not guaranteed to be exposed to one another or work together regularly. Our measures do not suffer from these issues.

Finally, although it may be the case that our friendship measures include a degree of noise, it is not clear how this noise would induce sorting patterns between natives and foreigners in the data. These friendship measures are primarily used as dependent variables in our regression analyses. As such, although the standard errors for our regression estimates may increase, they will remain unbiased.

3.5 Sorting Results

3.5.1 Unconditional Native-Foreigner Sorting Patterns.

Using the friendship we now document some basic evidence on the sorting patterns of native and foreign students. Firstly, for each student, we calculate their five "top" friends. These are defined as the five fellow students who appear the most number of times as a student's friends across all blocks of the study group data. Table 3.3 shows, for both native and foreign students, the observed percentage of students with differing amounts of native students in their top five friends. Additionally, we calculate the percentage of students in each category we would expect to see if it was the case that students would choose their friends independent of nationality. We present p-values from a test of equality between the actual observed percentages and the expected percentages.

The patterns in Table 3.3 are suggestive of a sizeable degree of segregation between native and foreign students. Approximately 10% of native students have 5 native top friends. Under the assumption of nationality-independent friendship formation, we would expect this to apply to only roughly 2% of native students. Similarly, we observe that 10% of foreign students have only fellow foreign as top friends, whereas under nationality-independent friendship formation we would expect this to apply to only 5% of foreign students. In both cases we strongly reject the null hypothesis that the observed percentages are equal to the expected percentages under a scenario of no segregation. In general, consistent with segregation, we observe a surplus (shortage) of native top friends for native (foreign) students.

Table 3.3: Top 5 Friends from Study Group Data

	Number of Native Friends in Top 5					
	0	1	2	3	4	5
Native Students						
Actual	5.34%	15.43%	24.33%	25.52%	20.18%	9.20%
Expected	5.00%	19.72%	33.12%	27.61%	12.31%	2.24%
p-value Actual=Expected	0.78	0.03	0.00	0.38	0.00	0.00
Foreign Students						
Actual	10.10%	25.42%	28.78%	26.38%	7.19%	1.92%
Expected	5.24%	20.76%	34.17%	26.87%	11.03%	1.91%
p-value Actual=Expected	0.00	0.03	0.82	0.38	0.00	0.99

Notes:
1. Table shows the percentage of students with various numbers of natives in students in their top 5 friends using the study group data, separately for native and foreign students.
2. Top 5 friends are defined as the 5 students that appear as a students friend most frequently across the second year.
3. The Expected rows show the percentages that would be expected if students do not sort based on nationality. These are calculated as the average values of 100 simulations in which students choose study groups randomly.
4. The p-value row presents the results of a T-test of the null hypothesis that the observed percentage is equal to the expected percentage.

One potential issue with our top friends measure is that students may coordinate equally with a group of friends larger than five, and therefore there may exist a large number of ties in the ranking of the top five friends. In cases where these ties are larger than five they must be broken arbitrarily. To avoid these issues, we also consider an alternative measure of close friendship. We define a student's "best" friend or friends as those students who appear the most number of times as a friend across all blocks in the study group data. The number of best friends varies from student to student, depending on the number of friends they coordinate equally with. Table 3.4 summarizes, for both native and foreign students, a number of statistics associated with these best friends. Namely, we calculate the mean proportion of natives in each student's best friends, the percentage of students who have all native best friends, and the percentage of students who have no native best friends. Also presented in Table 3.4 are the statistics we would expect to observe under nationality-independent friendship formation, and the p-values from a test of equality between the observed values and these expected values.

Table 3.4 reveals that the mean proportion of native students' best friends who are fellow natives is 0.56, and 10 percentage points more than would we expect to see under nationality-independent friendship formation. Roughly 40% of native students have only native best friends, more than double what would be expected under the alternative scenario. Conversely, the average proportion of native

Table 3.4: Best Friends from Study Group Data

	Origin of Best Friend(s)		
	Mean Prop. Native	All Native	None Native
Native Students			
Actual	0.56	42.30%	29.13%
Expected	0.46	17.33%	21.32%
p-value Actual=Expected	0.00	0.00	0.00
Foreign Students			
Actual	0.34	22.86%	50.81%
Expected	0.44	15.98%	22.77%
p-value Actual=Expected	0.00	0.00	0.00

Notes:
1. Table shows the mean proportion of native students, and the percentage of students who have all native and no native best friends, separately for native and foreign students.
2. Best friends are defined as the students that are tied for first place as the most frequent occurring friends across the second year.
3. The Expected rows show the percentages that would be expected if students do not sort based on nationality. These are calculated as the average values of 100 simulations in which students choose study groups randomly.
4. The p-value row presents the results of a T-test of the null hypothesis that the observed percentage is equal to the expected percentage.

Table 3.5: Friendship Data from Working Groups

	Number of Native Friends in Working Groups				
	0	1	2	3	All Native
Native Students					
Actual	16.03%	27.17 %	25.27%	31.52%	39.95%
Expected	20.06%	41.79%	30.46%	7.68%	12.39%
p-value Actual=Expected	0.03	0.00	0.01	0.00	0.00
Foreign Students					
Actual	42.08%	33.48%	18.55%	5.88%	9.73%
Expected	19.72%	41.27%	30.86%	8.05%	13.02%
p-value Actual=Expected	0.00	0.00	0.00	0.05	0.00

Notes:
1. Table shows the percentage of students with various numbers of natives in students in their working group friends, separately for native and foreign students.
2. Working group friends are defined as those students that shared a working group.
3. All Native columns shows the proportion of students that have only native working group friends.
4. Expected rows show the percentages that would be expected if students do not sort based on nationality. These are calculated as the average values of 100 simulations in which students choose working groups randomly.
5. The p-value row presents the results of a T-test of the null hypothesis that the observed percentage is equal to the expected percentage.

best friends among foreigners is 0.34, 10 percentage points less than the expected proportion. 50% of foreign students have only foreign students as best friends, compared to an expected value of 22%.

Finally, Table 3.5 presents sorting evidence using friendship ties elicited from the working group data. We examine the number of native friends, *i.e* the number of native students among the two to three individuals in a student's working group, for both foreign students and native students. Consistent with the patterns presented above, Table 3.5 reveals a large degree of sorting between native and foreign students. The number of native students with three fellow native students in their working group is roughly 30%, in comparison with an expected value of 8%. Roughly 40% of foreign students, on the other hand, have no native friends according to our working group data, in comparison with the expected 20%. Given that the total number of students varies somewhat between working groups, the last column calculates the percentage of students for whom all their working group friends are native. Again, we observe an surplus (shortage) of native friends for native (foreign) students. In all cases we strongly reject the hypothesis that the observed patterns are consistent with what we would expect under nationality-independent friendship formation.

Table 3.6: Descriptive Statistics of Dyads from Study and Working Group Data

	Panel A: Student Pairs from Study Group Data					
	Block 1	Block 2	Block 3	Block 4	Block 5	Total
Native-Native Pairs	8,839	6,998	8,001	6,641	8,124	38,603
Foreign-Foreign Pairs	13,123	11,134	12,422	11,361	12,761	60,801
Native-Foreign Pairs	21,668	17,830	20,171	17,498	20,532	97,699
Total	43,630	35,962	40,594	35,500	41,417	197,103
Mean of Friendship Variable	0.04	0.06	0.05	0.06	0.11	0.06
	Panel B: Student Pairs from Working Group Data					
	Course 1	Course 2				Total
Native-Native Pairs	8,796	8,695				17,491
Foreign-Foreign Pairs	12,017	13,308				25,325
Native-Foreign Pairs	20,740	21,632				42,372
Total	41,553	43,635				85,188
Mean of Friendship Variable	0.02	0.02				0.02

Notes:
1. Table shows the number of native-native, foreign-foreign, and native-foreign student pairs per block using the study group pair data (Panel A), and the working group pair data (Panel B).
2. Mean of Friendship Variable gives the average value of the friendship variable, equal to one if the student pair are categorized as friends, in a particular block.

3.5.2 Regression Approach.

The above results strongly suggest a sizeable degree of segregation between native and foreign students. However, one difficulty with interpreting these unconditional patterns is that differences in other determinants of friendship may be driving the sorting. For instance, differences in the gender composition of foreign and native students may result in what appears to be sorting based on nationality, but is actually explained by sorting based on gender.

To address this we follow Marmaros and Sacerdote (2006), Mayer and Puller (2008) and Foster (2005) in examining the probability that a pair of students are friends within a regression framework. Using our study group data, we generate all possible pairs of students for each block. Given N_{bt} students in a particular block b, this produces $(N_{bt} \times N_{bt} - 1)/2$ pairings of students.[12] Using our study group friendship data, we define a variable $Friends(i, j)_{bt}$ that is equal to one if the student pair (i, j) are indeed classified as friends in block b (*i.e.* registered for the same study group in both courses within 12 hours of each other), and zero otherwise. An identical process is applied to the working group data.

Table 3.6 gives a summary of the student pair data, as well as the mean value of the friendship variable in the different periods. Across all blocks the study group data produces 155,686 student

[12]This is done by crossing the relevant list of student numbers with itself, removing all duplicate pairs (i, i), and keeping only one instance of the same pairing ((i, j) and (j, i)).

pairs. An overall mean of the friendship variable of 0.06 indicates that two randomly chosen students in a block have a 6% chance of being friends. The working group data provides 85,188 pairs, 2% of which are classified as being friends.

In order to analyze native-foreign sorting patterns, we run linear probability models of the following form:

$$Friends(i,j)_{bt} = \alpha_0 + \alpha_1 BothNative + \alpha_2 BothForeign + \alpha_3 X + C_{bt} + \epsilon(i,j)_{bt} \tag{3.1}$$

Where $BothNative$ and $BothForeign$ are dummy variables equal to one if students i and j are both native or both foreign, respectively. Vector X optionally includes variables capturing if students i and j share the same gender or ability, as well as the absolute difference in their age in years.[13] Finally, C_{bt} captures block-cohort fixed effects. The coefficient α_1 then gives the difference in the probability of friendship of a native-native pair compared to the reference category of a native-foreign pair, controlling for potential differences in the gender, ability and age of the student pair. Similarly, α_2 gives the conditional difference in the probability of friendship of a foreign-foreign pair compared to a native-foreign pair.[14]

Although our focus will be on the study group pair data, our strategy for all regressions in the remainder of the paper will be to also present results using the working group pair data as a type of robustness check. By comparing results from friendship data derived from two distinct sources we hope to lessen the probability of spurious findings. We follow this convention for all of our remaining analyses.

Using the study group pair data, column (1) of Table 3.7 presents the results of Equation (3.1) excluding the control variables. Column (2) adds variables adjusting for potential differences in the gender, ability, and age of the student pair. Columns (3) and (4) repeat the same specifications for the working group pair data. These patterns are consistent with a substantial degree of sorting between native and foreign students. Not accounting for differences in age, gender and ability, column (1) reveals that two native students are 1.47 percentage points more likely to be friends than a native-foreign pair, according to the study group pair data. When differences in basic characteristics are taken into account in column (2), this difference in probability reduces only marginally to 1.31 percentage points, and remains highly significant. These effects are large when one considers that the unconditional mean of the outcome variable is 0.063. The reduction in the probability of friendship

[13]We define students in the top quartile of the average of first year grades at high ability, students in the bottom quartile as low ability, and the remaining students as average ability.

[14]We cluster our standard errors on a variable that takes on different values for each unique combination of the student pair's first year study groups. All continuous variables are standardized, and thus their coefficients should be interpreted in terms of standard deviations. We follow these conventions for the remainder of the paper.

Table 3.7: Sorting Results From Regression Approach

	Friends Yes (1) or No (0)			
	Study Group Pair Data		Working Group Pair Data	
	(1)	(2)	(3)	(4)
Both Native	0.0147***	0.0131***	0.0140***	0.0135***
	(0.0027)	(0.0026)	(0.0019)	(0.0020)
Both Foreign	0.0031	0.0037*	0.0122***	0.0126***
	(0.0021)	(0.0021)	(0.0015)	(0.0015)
Same Gender		0.0035*		0.0149***
		(0.0018)		(0.0015)
Same Ability		0.0035**		0.0059***
		(0.0016)		(0.0013)
Age Diff.		-0.0029***		-0.0007*
		(0.0007)		(0.0004)
Unconditional Mean	0.0630	0.0630	0.0235	0.0235
	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Observations	197,103	197,103	85,188	85,188
R^2	0.02	0.02	0.00	0.01

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

for native-foreign pair then amounts to a reduction of roughly 20% of the baseline. Similar results are found in columns (3) and (4) using the working group pair data. In this case, since the unconditional mean of the friendship variable is lower, the difference of 1.35 percentage points between the probability of a friendship between a native-native and native-foreign pair is almost 60% of the baseline.

The patterns in this section point to significant and substantial segregation between native and foreign students on campus. It is outside the scope of this paper to make any normative statement about the degree to which these patterns are desirable or not.[15] However, given that non-trivial interactions are necessary for diversity to yield benefits (Camargo et al., 2010), whatever benefits do arise from an internationally diverse campus could be further increased if the segregation between native and foreign students was reduced. Depending on the institutional aims and the expected gains from further interaction, universities may wish to implement policies aimed increasing the levels of friendship and interaction between natives and foreigners, thereby ensuring that any benefits are fully realized. The following section explores the efficacy of one such policy.

3.6 Encouraging Native-Foreign Friendships

Considering the structure of a typical university program, where students must attend lectures and study groups, a natural approach to encouraging international friendships may be to use such opportunities as a means to force native and foreign students to be exposed to one another within a small geographic proximity for an extended time. Evidence from existing studies suggests that such policies are effective in producing friendships and interaction between diverse students. For instance, multiple studies have documented that interracial student pairs randomly allocated to share a dorm or a close equivalent of a study group were significantly more likely to subsequently become friends (Marmaros and Sacerdote, 2006; Camargo et al., 2010; Foster, 2005).

The setting of the bachelor program allows us to investigate the effect of prolonged forced exposure on the probability of subsequent native-foreign friendships. We take advantage of the fact that students in their first year of the program are randomly allocated to a study group of approximately 24 students, with whom they follow multiple weekly tutorials with for the entire first academic year. As outlined in Section 3.3.1, during their first year, students must meet with their allocated study group three times a week, and must attend at least 70% of these weekly sessions per course in order to pass. Thus, students have regular and forced exposure to their first-year study group peers; a back-of-the

[15] It should be noted that the apparent preference for segregation has at least one glaring educational consequence; because study and working groups are self-chosen, as is the case in many universities, students' classrooms and working groups in turn become less internationally diverse.

envelope calculation reveals that even students who barely comply with the 70% threshold rule must be present in their first-year study group for a total of just over 131 hours across the first academic year.[16]

Studying how forced and prolonged exposure influences subsequent friendship is often difficult using observational data; students who are geographically proximate will likely have many unobservable characteristics that also determine the probability of friendship. Individuals with similar interests and characteristics tend cluster together. It may then be these shared interests and characteristics, rather than exposure, that determines friendship. Moreover, students who are already friends are likely to choose to be exposed to one another. Disentangling the actual effect of exposure on the probability of friendship is therefore difficult. Taking advantage of the randomized allocation to first year study groups ensures that we avoid these issues. Appendix Section 3.A presents various balancing tests to provide evidence that allocation to first year study groups was indeed random.

3.6.1 Exposure Through First Year Study-Group.

We examine the effect of this first-year exposure on the subsequent probability of friendship between a native and a foreign student. We run regressions of the following form:

$$Friends(i,j)_{bt} = \alpha_0 + \alpha_1 FirstYear + \alpha_2 X + C_{bt} + \epsilon(i,j)_{bt} \tag{3.2}$$

Where $FirstYear$ is an indicator variable taking the value of one if student i and student j were randomized into the same first-year study group, and zero otherwise, vector X_i contains additional variables to control for the characteristics of the student pair, and C_{bt} captures block-cohort fixed effects.

These results are presented in Table 3.8. Using the study group pair data, column (1) gives the results of Table 3.8 without the addition of control variables. The coefficient for $FirstYear$ implies that a student pair who are exposed to one another through being randomized into the same first year study group are 1.84 percentage points more likely to be classified as friends than a pair not in the same first year study group. Column (5) repeats this specification for the working group pair data, in which an exposed first-year pair are 2.59 percentage points more likely to be friends. In column (2) and (6) we break down this effect the type of student pair by adding interactions of $FirstYear$ and indicators for native-native and foreign-foreign pairs. The $FirstYear$ coefficient then gives the exposure effect for native-foreign pairs only. While the exact results differ somewhat

[16]Complying with the 70% attendance rule for both the heavy and light course amounts to attending 15 study-group sessions per block, and 75 across the whole year. Each session lasts for 1 hour and 45 minutes. The minimum total attendance time at these sessions to enter the second year is then equal to $(1.75 \times 75 =)$ 131.25 hours.

Table 3.8: Exposure Effect on Student Pairs

	Friends Yes (1) or No (0)							
		Study Group Pair Data				Working Group Pair Data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
First Year	0.0184***	0.0136**	0.0136**	0.0010	0.0259***	0.0196***	0.0195***	0.0032
	(0.0057)	(0.0063)	(0.0064)	(0.0070)	(0.0026)	(0.0034)	(0.0035)	(0.0037)
Both Native		0.0131***	0.0115***	0.0115***		0.0106***	0.0101***	0.0101***
		(0.0027)	(0.0027)	(0.0027)		(0.0017)	(0.0018)	(0.0018)
Both Foreign		0.0018	0.0024	0.0023		0.0112***	0.0115***	0.0115***
		(0.0021)	(0.0021)	(0.0021)		(0.0015)	(0.0015)	(0.0015)
First Year × Both Native		0.0109	0.0110	0.0111		0.0220***	0.0221***	0.0221***
		(0.0084)	(0.0084)	(0.0085)		(0.0063)	(0.0063)	(0.0063)
First Year × Both Foreign		0.0084	0.0085	0.0090		0.0056	0.0058	0.0065
		(0.0070)	(0.0071)	(0.0071)		(0.0047)	(0.0047)	(0.0047)
First Year × Same Gender				0.0235***				0.0305***
				(0.0061)				(0.0052)
Pair Controls			✓	✓			✓	✓
Unconditional Mean	0.0630	0.0630	0.0630	0.0630	0.0235	0.0235	0.0235	0.0235
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Observations	197,103	197,103	197,103	197,103	85,188	85,188	85,188	85,188
R^2	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. Pair controls include the absolute difference in age, and indicator variables capturing if the pair share the same gender and ability.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

between the study group and working group data, they suggest that the exposure effect may be largest for same-type pairs (*i.e.* native-native pairs and foreign-foreign pairs). The addition of controls for differences in age, gender and ability in column (3) and (7) leaves these estimate unchanged for both data sources. Finally, columns (4) and (8) add an interaction term between $FirstYear$ and the same gender indicator. The fact that the main $FirstYear$ effect is no longer significantly different from zero with the addition of this interaction implies that the exposure effect predominately works through native-foreign pairs of the same gender.

Although the results in Table 3.8 suggest that the exposure effect may be largest for native-native and foreign-foreign pairs, they do imply that exposure between native-foreign pairs has some potential to reduce their propensity to segregate. These effects are large when viewed in the context of the relatively rare event of student friendship. The increase in friendship probability due to exposure for native-foreign pairs amounts to 22% and 83% of the unconditional friendship probability for the two data sources respectively. Also notable is the fact that native-foreign $FirstYear$ coefficients in Table 3.8 closely match or exceed in absolute size the sorting coefficients in Table 3.7. This implies forced exposure could virtually entirely offset the difference in the probability of a native-native and a native-foreign friendship, at least for those of the same gender.[17] The following sections explore this native-foreign exposure effect in more detail.

3.6.2 Heterogeneity in Exposure Effect.

The results above are in line with the findings of a number of previous studies (Marmaros and Sacerdote, 2006; Camargo et al., 2010; Foster, 2005); forcing students to share a close geographic proximity for an extended time, on average, promotes friendships. However, what has so far been unaddressed by the literature is the degree to which individual characteristics of either student may alter this friendship promoting effect.

Exploring heterogeneity in the exposure effect by individual characteristics is interesting for at least two reasons. Firstly, given that one of the goals of an intentional campus is for native students to form beneficial friendships with students from backgrounds that are appropriately distinct from their own, it is important to see if exposure can encourage such friendships. For instance, do the results above hold for students coming from radically different cultures, or are they restricted to native-foreign students who are culturally similar? Secondly, there are natural limits on the effectiveness of policies based on shared close geographic proximity. Namely, only a limited amount of students can

[17] A proportion of the native-foreign friendships under analysis in Table 3.7 are those occurring through forced exposure in the first year tutorial group. A comparison of the coefficients in Table 3.7 and Table 3.8 may therefore not be entirely appropriate. Appendix Table A.3.2 repeats the sorting regressions in Table 3.7 only using student pairs who were not in the same first year study group, where we find similar results.

physically share the same geographical space. Therefore, it may be sensible in some circumstances to "ration" forced exposure to only those students who could potentially benefit the most.

We identify three factors that may alter the potential benefits of a native-foreign friendship. Firstly the diversity-related benefits may differ depending on the cultural distances between the pair. It seems likely, for instance, that a native Dutch student will gain more intercultural skills from interaction with a foreign Chinese student than a Belgian student, given the large degree of cultural similarity and familiarity between large parts of Belgium and the Netherlands.

Second, the potential benefit of a native-foreign partnership may differ depending on the existing attitudes of the native student. Natives who arrive at university with no existing prejudices or in-group biases will have little room to improve their intercultural skills, whereas native students with xenophobic attitudes will have a large potential to benefit from a partnership with a foreigner.

Finally, the benefits of friendship with a foreign student may depend on a native student's previous exposure to foreigners. Natives growing up in highly international neighbourhoods are more likely to have already benefited from diversity-related exposure, while those with little existing experiences interacting with foreigners could potentially derive larger benefits from partnerships with foreign students.

We use our detailed student administrative data to investigate the heterogeneity in the exposure effect by these three factors. We calculate the cultural distance between the country of origin of each international student and the Netherlands. To do so we use the widely used cultural indices developed by Hofstede (2001). The cultural distance between the country of origin of the foreigner and the Netherlands is calculated using the method of Kogut and Singh (1988).[18]

For use as control variables when examining the role of cultural distance, we also retrieve the geographical distance between the country of origin of the foreigner and the university campus, as well as a measure of common languages spoken between the country of origin of the foreign student and the Netherlands.[19] This latter variable is defined as the probability that an individual from the foreign student's country and an individual from the Netherlands will be able to understand one another in some common spoken language (Melitz and Toubal, 2014).

As a proxy for the attitudes a native student may have towards foreigners before entering university we use the voting behavior of their municipality. Municipalities act as small administrative regions and number 335 within the Netherlands. We retrieve the average voting share received by the

[18]The Hofstede measures consider four different cultural traits; power distance, uncertainty avoidance, masculinity/femininity, and individualism. Cultural distance from the Netherlands is then measured for each country j as $CD_j = \sum_{i=1}^{4}((I_{ij} - I_{in})^2/Var_i)/4$ where I_{ij} is the index of the ith cultural dimension for country j, Var_i is the variance of the index of the ith dimension, and n denotes the Netherlands.

[19]This variable is taken from the CEPII database: `http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele.asp`

Party for Freedom (PVV) in the 2010 Dutch general elections within the municipality in which each native student went to high school. The PVV is generally recognized as a far-right, anti-immigration party.[20] In the 2010 general election they received their highest vote share to date with 15.4% of the vote, becoming the third most popular party. We reason that students from municipalities with higher vote shares for PVV are more at risk of having xenophobic views. As a measure of the likely exposure a student has had with foreigners before entering university, we retrieve the proportion of their municipality population classified as being of a "Dutch background" (*i.e.* non-immigrant background).

For use as control variables when examining the characteristics of the native student, we retrieve numerous other municipality-level variables; the average vote share for the other major political parties in 2010, the distance of the municipality from the university campus, the percentage of the municipality population who have an undergraduate university degree or higher, and the percentage of the municipality collecting unemployment benefits.[21]

We assign individual students with municipality- and country-level characteristics. While country-level measures have successfully been used as a proxies for culture, beliefs and attitudes in other contexts (Fernandez and Fogli, 2009), the evidence we present in this section should therefore be viewed with the appropriate caution. We note that any noise introduced as a result of these imperfect measures will bias our results towards zero.

Given our interest in the potential heterogeneity of forced exposure only for native-foreign friendships we now restrict our focus to solely the native-foreign student pairs in our student pairs data. This amounts to 97,699 native-foreign pairs in the study group pair data, and 42,372 native-foreign pairs in the working group pair data (see Table 3.6). In order to examine heterogeneity of the exposure effect, our strategy will be to run specifications of the following form on this restricted data set:

$$Friends(i^N, j^F)_{bt} = \alpha_0 + \alpha_1 FirstYear + \alpha_2 X + \alpha_3 \big[FirstYear \times X\big] + C_{bt} + \epsilon(i^N, j^F)_{bt} \quad (3.3)$$

The variables are defined as in Equation (3.2), and the superscripts N and F refer to native and foreign students, respectively. Coefficient α_3 then indicates how $FirstYear$ varies with some characteristic X — say, PVV vote share of the native student's municipality — and thus how the year-long friendship inducing exposure effect depends on the characteristics of either one of the student pair. We separately examine the heterogeneity in the exposure effect for the characteristics of the foreign student (*i.e.* their cultural distance from the host country) and of the native student (*i.e.* their municipality's voting behaviour and proportion of non-immigrant Dutch).

[20]For instance, among their stated policy goals is the deportation asylum seekers, a banning of immigration from "Islamic" countries (much of the immigration to the Netherlands in recent years has come from Turkey and Morocco), and for the Netherlands to leave the EU. The PVV has also been critical of the number of international students in the Netherlands.

[21]For the latter two percentages we use the average value of these measures between 2007 and 2014.

Table 3.9: Heterogeneity in Exposure Effect by Foreign Characteristics

	Friends Yes (1) or No (0)			
	Study Group Pair Data			
	(1)	(2)	(3)	(4)
First Year	0.0136**	0.0138**	0.0138**	0.0173**
	(0.0062)	(0.0062)	(0.0062)	(0.0070)
Cultural Dist.		0.0023*	0.0043***	
		(0.0013)	(0.0017)	
First Year × Cultural Dist.		-0.0086**	-0.0086**	
		(0.0037)	(0.0037)	
Top 25% Cult. Dist.				0.0027
				(0.0036)
First Year × Top 25% Cult. Dist.				-0.0157**
				(0.0074)
Unconditional Mean	0.0592	0.0592	0.0592	0.0592
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Country Controls			✓	✓
Observations	97,699	97,699	97,699	97,699
R^2	0.02	0.02	0.02	0.02

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs are included in the regressions.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. Country controls include the geographical distance in kilometers between the country of origin of the foreigner and the university campus, and a measure of common languages spoken between the country of origin of the foreign student and the Netherlands.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We begin by examining the heterogeneity in the exposure effect by the foreigner's characteristics using the study group pair data in Table 3.9. Column (1) repeats the earlier specification showing only the exposure effect for native-foreign pairs. Column (2) adds the standardized cultural distance from the country of origin of the foreigner of the student pair to the Netherlands, as well as an interaction between this and the $FirstYear$ indicator. Column (3) adds controls for both the standardized distance in kilometres between the country of origin on the foreigner and the university campus, and the standardized common language measure. Finally, column (4) replaces the continuous cultural distance variable with an indicator that takes the value of one if the foreign student is in the top quartile of the cultural distance measure. Appendix Table A.3.3 presents the results using the working group pair data.

These results suggest that the impact of the exposure effect is heavily dependent on the cultural similarity between the native and foreign student. Column (2) reveals that a one standard deviation increase in the cultural distance measure reduces the $FirstYear$ coefficient by roughly 60%. Column (3) shows that the reduction in the exposure effect cannot be explained simply by potential language differences or geographical distance. Controlling for these factors leaves the $FirstYear$ interaction unchanged. When we consider foreign students in the top quartile of the cultural distance distribution in column (4), the increase in the probability of a native-foreign friendship brought about by exposure has been basically erased.[22] A test for the sum of the coefficients being different from zero is strongly insignificant (p-value=0.82). Qualitatively similar patterns are found in the working group pair data in Appendix Table A.3.3.

Table 3.10 examines the heterogeneity in the exposure effect by characteristics of the native student using the study group pair data. Note that the sample size has reduced as for only approximately 80% of native students are we able to match their high school data to a municipality. Column (1) again repeats the specification including only the exposure effect. Column (2) adds the standardized share of votes received by PVV in the municipality in the 2010 general elections, as well as an interaction of this variable with $FirstYear$. Column (3) adds the municipality level controls. Column (4) replaces the continuous vote share variable with an indicator variable taking the value of one if the native student originates from the municipalities in the top quartile of PVV votes. Columns (5) to (6) repeat the specifications in columns (2) to (3) replacing the PVV vote share with the share of the municipality population of the native student who are classified as non-immigrant Dutch. Appendix Table A.3.4 presents the results using the working group data.

[22]The countries in this category are Albania, Austria, China, Cyprus, Ecuador, Greece, Macedonia, Mexico and the Philippines.

Table 3.10: Heterogeneity in Exposure Effect by Native Characteristics

	Friends Yes (1) or No (0)						
	Study Group Pair Data						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Year	0.0103 (0.0078)	0.0103 (0.0079)	0.0104 (0.0078)	0.0146 (0.0091)	0.0104 (0.0078)	0.0105 (0.0077)	0.0060 (0.0080)
Vote % Far-Right		0.0042*** (0.0015)	0.0048 (0.0032)				
First Year × % Far-Right		-0.0071* (0.0040)	-0.0072* (0.0039)				
Top 25% Far-Right				-0.0079* (0.0045)			
First Year × Top 25% Far-Right				-0.0233** (0.0116)			
Pop. Dutch					-0.0014 (0.0013)	-0.0024 (0.0021)	
First Year × Pop. Dutch					0.0098** (0.0041)	0.0100** (0.0041)	
Top 25% Pop. Dutch							-0.0006 (0.0050)
First Year × Top 25% Pop. Dutch							0.0189* (0.0100)
Unconditional Mean	0.0588 (0.0009)	0.0588 (0.0009)	0.0588 (0.0009)	0.0588 (0.0009)	0.0588 (0.0009)	0.0588 (0.0009)	0.0588 (0.0009)
Municipality Controls			✓	✓		✓	✓
Observations	71,024	71,024	71,024	71,024	71,024	71,024	71,024
R^2	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Notes:
1. All regressions include block-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs are included in the regressions.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. Municipality controls include the average vote share for the other major political parties in 2010, the geographic distance in kilometers from the municipality to the university campus, the percentage of the municipality who have an undergraduate university degree or higher, and the percentage of the municipality who collect unemployment benefits.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notable in Table 3.10 is that the $FirstYear$ effect is no longer significant. Compared to column (1) of Table 3.8, the point estimates have reduced slightly and the standard errors have increased. Given that the point estimates of $FirstYear$ are comparable to earlier specifications, and that the exposure effect remains significant when using the working group data in Table A.3.4, this loss of significance may be due to the reduction in observations. Nevertheless, strictly speaking we cannot reject the hypothesis for this restricted sample that native-foreign pairs allocated to the same first year study group are equally as likely to become friends than pairs in different study groups. Despite this, column (2) reveals that the probability of a friendship developing between a native-foreign pair allocated to the same first year study group depends on the PVV vote share of the municipality of the native student; a one standard deviation in the vote share for this far-right party decreases the exposure effect by 0.07 percentage points. This effect remains even when adjusting for a set of comprehensive municipality-level controls in column (3). Column (4) implies that for native students in the top quartile of the PVV voting share, the probability of a friendship with a foreign student in their first year study group is significantly reduced by 2.33 percentage points. Despite the $FirstYear$ dummy not being significant, comparing the magnitude of the PVV interaction with the main $FirstYear$ point estimate suggests that any increase in the probability of native-foreign friendship brought about by the exposure effect are completely absent for native students in the top quartile of the PVV voting share. Similar patterns are found in columns (2), (3), and (4) of Appendix Table A.3.4 using the working group pair data.

Columns (5), (6), and (7) of Table 3.10 examine how the proportion of non-immigrant Dutch in the population of a native student's municipality may alter the exposure effect. Natives who may have had little previous interactions with foreigners appear to be most sensitive to exposure effect. The results in column (6), which presents the specification with municipality level controls added, show that a one standard deviation increase in the proportion of non-immigrant Dutch in a native's municipality significantly increases the exposure effect by 1 percentage point, which amounts to a doubling of the $FirstYear$ effect. Focusing only on the municipalities in the top quartile of the non-immigrant Dutch variable, the interaction effect indicates that the exposure effect is increased by 1.89 percentage points, significant at the 10% level. However, these patterns are not present in columns (5), (6) and (7) of Appendix Table A.3.4 where the working group pair data is used. Overall, we therefore view the evidence for a larger effect of exposure for natives with little prior interaction with foreigners as suggestive only.

Although tentative, these results add some important qualifiers to the average exposure effect found in Section 3.6.1, as well as earlier findings in the literature. Forced and prolonged exposure within a close geographic proximity does appear to promote friendship, but the effectiveness of this exposure varies with the characteristics of the students. The institutional manipulation of friendships between native-foreign student pairs who may otherwise no meet and thus for whom interaction may be most beneficial — those from different genders and cultures, and those including students at risk of having existing xenophobic attitudes — appears to be difficult through forced exposure.

3.6.3 Exposure Effect Multiplier.

The results above establish that native-foreign friendships can be encouraged through forced exposure in the first year study groups, though this effect does appear to have some important heterogeneity. It may also be the case that exposure to foreigners (natives) in the first year study group stimulates friendships with foreigners (natives) outside of these study groups. In other words, there may be some "multiplier" effect on foreign-native friendships from exposure. Such an effect could mean that exposure is a far more efficient method of promoting native-foreign friendships than suggested by the estimates above. Namely, not every foreign and native student would have to be exposed to one another in order for segregation to be eliminated.

To investigate this we run regressions of the following form:

$$Friends(i^N, j^F)_{bt} = \alpha_0 + \alpha_1 PropForeigner_{i^N} \\ + \alpha_2 PropNative_{j^F} + \alpha_3 X + C_{bt} + \epsilon(i^N, j^F)_{bt} \quad (3.4)$$

Where $PropForeigner_{i^N}$ is the proportion of foreign students in the native student's first year study group, $PropNative_{j^F}$ is the proportion of native students in the foreign student's first year study group, and the remaining variables are defined as in Equation (3.2). To study the effect of the conditions of the first year study group on the propensity of native-foreign friendships outside of these study groups, we only consider so-called non-exposed native-foreign student pairs; native-foreign pairs not allocated to the same first year study group. Therefore, α_1 (α_2) captures the change in the probability of a non-exposed native-foreign friendship brought about by a one standard deviation change in the proportion of foreigners (natives) in the native (foreign) student's first year study group.

Table 3.11 presents the results of these regressions using the study group pair data. Column (1) presents the results of Equation (3.4). Based on the fact that the exposure effect seemingly works primarily through same-gender pairs, as shown in Table 3.8, we calculate the proportion of foreigners (natives) in the native (foreign) student's first year study group who are of the same gender as the student themselves. Column (2) presents the results using these same gender proportions now considering only non-exposed native-foreign pairs of the same gender. Column (3) adds an interaction term between the two proportions. Column (4) checks for possibly heterogeneity in the multiplier effect

Table 3.11: Multiplier Effect of Exposure

	Friends Yes (1) or No (0)			
	Study Group Pair Data			
	(1)	(2)	(3)	(4)
Prop. Foreigner	-0.0019			
	(0.0013)			
Prop. Native	-0.0008			
	(0.0016)			
Prop. Foreigner (Same Gen.)		0.0041**	0.0040**	0.0039
		(0.0018)	(0.0017)	(0.0024)
Prop. Native (Same Gen.)		-0.0018	-0.0018	-0.0022
		(0.0020)	(0.0020)	(0.0024)
Prop. Foreigner (Same Gen.) × Prop. Native (Same Gen.)			-0.0018	
			(0.0020)	
Top 25% Far-Right				0.0021
				(0.0047)
Top 25% Cult. Dist.				0.0024
				(0.0049)
Prop. Foreigner (Same Gen.) × Top 25% Far-Right				-0.0080
				(0.0049)
Prop. Foreigner (Same Gen.) × Top 25% Cult. Dist.				0.0050
				(0.0035)
Prop. Native (Same Gen.) × Top 25% Far-Right				-0.0044
				(0.0054)
Prop. Native (Same Gen.) × Top 25% Cult. Dist.				-0.0015
				(0.0046)
Unconditional Mean	0.0574	0.0559	0.0559	0.0559
	(0.0008)	(0.0011)	(0.0011)	(0.0011)
Observations	83,169	44,237	44,237	32,702
R^2	0.02	0.02	0.02	0.02

Notes:
1. All regressions include block-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs not allocated to the same first year study group are included in the regressions. Columns (2), (3), and (4) consider only pairs of the same gender.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

by the cultural distance and PVV vote shares of the foreigner and native, respectively. The results for the identical specifications using the working group pair data are presented in Appendix Table A.3.5.

While column (1) of Table 3.11 suggests that neither of the total proportions appear to be significantly related to the probability of a native-foreign friendship, column (2) implies that native students exposed to more foreign students of the same gender are induced to make new foreign friends of the same gender with whom exposure is not forced. A one standard deviation in the proportion of same gender foreign students in a first year study group (an increase in the proportion of foreigners by 0.10) significantly increases the probability of a friendship with a non-exposed foreigner by 0.4 percentage points. We do not find an equivalent effect for foreign students. Columns (3) and (4) reveal no convincing evidence of an interaction between the proportions of the native-foreign pair, nor of any systematic differences in these effects by the characteristics of either the native or the foreign student. Overall, the results in column (2) and (3) provide some evidence of a small multiplier effect of exposure for native students. However, given that Appendix Table A.3.5 shows no significant effect across any specification when using the working group pair data, we view this evidence with appropriate caution.[23]

3.7 Conclusion

A commonly cited benefit of the increased admittance of foreign students is the facilitation of beneficial native-foreigner friendships. Despite this assertion, and the ongoing debate surrounding the internationalization of education, empirical evidence documenting the actual native-foreign sorting patterns remain scarce.

This paper addresses the gap in the literature. Using a novel friendship elicitation method based on students' choices at university, our first set of results show a significant and substantial degree of segregation between native and foreign students. Focusing on best friends — likely those friendships that have the largest influence on individuals — we find that 42% (50%) of native (foreign) students have only fellow native (foreign) students within this important group. Our regression framework reveals that even when controlling for potential differences in age, ability and gender, any native-foreign friendship is 20% less likely than a native-native friendship.

Such a degree of segregation between native and foreign students may be surprising for two reasons. First, the program we study is taught in English and designed with an international focus. Second, according to commonly observed empirical patterns in social networks, the tendency to segregate should be least pronounced where the group sizes are roughly even - as is the case in our

[23] It should be noted that the variation in the proportion of natives and foreigners is limited by the nature of the random assignment. This lack of support may reduce our ability to detect a multiplier effect, should it exist.

setting (Currarini et al., 2009). Despite both the nature of the program, and the roughly equal number of native and foreign students, we nevertheless observe a non-trivial degree of segregation.

We consider the effectiveness of a seemingly natural remedy for segregation on campus; the promotion of native-foreign friendships through forcing such students to be exposed to one another within a small geographic proximity for an extended period. While we find forced exposure can promote native-foreign friendships, this finding comes with some important caveats. Namely, encouraging native-foreign friendships between students who may derive the largest benefit is difficult; the exposure effect is apparently absent for student pairs between whom the cultural distance is large, and for native students at risk of having xenophobic views.

The most obvious obstacle to reducing segregation through forced exposure is the reality that only a small number of students can be forced to share a close geographic space. The fact that we find no convincing evidence of a large multiplier effect of exposure, and the fact that a native (foreign) student cannot be forced to be exposed to every foreign (native) student in their cohort, means that forced exposure cannot entirely eliminate segregation.

Our results imply that there may be a gap in the rhetoric of the purported benefits of the internationalization of education and what actually occurs on university campuses. Given the segregation patterns we observe, universities could be missing out on fully realizing the gains from an international student body. Moreover, it appears there is no quick fix to address segregation within the normal structure of university programs. A deeper understanding of the reasons for segregation of students from different backgrounds is required if universities wish to yield these benefits.

3.A Appendix

First Year Study Group

The main assumption underlying our identification of the exposure effect in the first year tutorials is that unobserved characteristics determining friendship are not systemically correlated with first year study group assignment. Random allocation of students to study groups, the details of which are described below, makes this identifying assumption likely to hold.

3.A.1 First Year Study Group Allocation.

On the first day of the academic year, every student who has preregistered to the program is invited to come to campus where they must confirm their registration.[24] This happens through approximately 10 to 15 administrative personnel, who add names and students' details to an electronic list.

Study group membership is then determined from this list. After being sorted on a randomly determined ID, the first student on the list is assigned to study group 1, the second student to study group 2, and so forth. This process continues until the maximum study group has been reached, at which point allocation begins again at study group 1. Students who are late to register and students who wish to be re-assigned due to irresolvable scheduling conflicts outside of university are allocated at the discretion of the university administrator. We do not observe this in our dataset. Though we were informed that these cases are rare, the final group size and composition may differ slightly from the initial assignment.

3.A.2 Balancing Test.

We provide evidence that the final allocation of students to study groups was indeed random using a series of balancing checks. Given that the allocation process takes place at the beginning of the program, the following specifications include all students observed to register for the first year of the program including those students that drop out before the start of their second year.

First, we analyse whether the proportion of foreign students in a student's first year study group can be explained by their background characteristics X_i.

$$PropForeigner_{ig} = \gamma_0 + \gamma_1 X_i + T_t + \epsilon_{igt}$$

[24]This confirmation process is done so that no-shows are not taken into account when forming the tutorial groups.

The background characteristics X_i include age, gender, and student number.[25] We include cohort fixed effects (T_t) as randomization into groups takes place per cohort. Estimates of γ_1 significantly different from zero would indicate that our identifying assumption may be violated. Appendix Table A.3.6 shows the results of this test, where column (1) to (3) examine the degree to which student's background characteristics determine the proportion of foreigners, the proportion of European foreigners, and the proportion of non-European foreigners respectively.

Our second balancing check tests for systematic clustering of certain characteristics within the first year study groups. We regress each background characteristic above - as well a dummy variable $Native$ indicating if a student is native or not - on first year study group dummies and cohort fixed effects. Next, in a separate model we regress the same student characteristics only upon cohort fixed effects and perform an F-test on the small versus big model. If the larger model explains the variation in the background characteristics significantly better than the smaller model, this may indicate that students with certain characteristics have been clustered in study groups. These results are presented in Appendix Table A.3.7.

The findings for both our tests are consistent with the final allocation of students to study groups being random. Across the three specifications in Appendix Table A.3.6 we find all student characteristics to be individually and jointly insignificant. Appendix Table A.3.7 shows that for no background characteristics does the F-test reject the null hypothesis that the parsimonious model fits the data equally as well as the extended model.

[25] When students register for the program they are automatically assigned a student number. This number is used for administrative purposes during their time at university. Given that a student is allocated a student number at the time of registration, it may correlate with their motivation to take the program.

Table A.3.1: Validation of Friendship Measures

	Friends Yes (1) or No (0)	
	Study Group Pair Data	Working Group Pair Data
	(1)	(2)
Same Country	0.0142***	0.0173***
	(0.0025)	(0.0020)
Same Gender	0.0033*	0.0145***
	(0.0018)	(0.0015)
Same Ability	0.0035**	0.0059***
	(0.0016)	(0.0013)
Age. Diff	-0.0027***	-0.0001
	(0.0008)	(0.0004)
Same School	0.1353**	0.1397***
	(0.0602)	(0.0510)
Unconditional Mean	0.0630	0.0235
	(0.0010)	(0.0010)
Observations	197,103	85,188
R^2	0.02	0.01

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.2: Sorting Results Excluding First Year Study Group Pairs

	Friends Yes (1) or No (0)			
	Study Group Pair Data		Working Group Pair Data	
	(1)	(2)	(3)	(4)
Native-Native	0.0131***	0.0112***	0.0106***	0.0099***
	(0.0027)	(0.0027)	(0.0017)	(0.0018)
Foreign-Foreign	0.0018	0.0025	0.0112***	0.0116***
	(0.0021)	(0.0021)	(0.0015)	(0.0015)
Same Gender		0.0001		0.0104***
		(0.0017)		(0.0012)
Same Ability		0.0026		0.0040***
		(0.0016)		(0.0012)
Age Diff.		-0.0034***		-0.0012***
		(0.0007)		(0.0003)
Unconditional Mean	0.0604	0.0604	0.0195	0.0195
	(0.0006)	(0.0006)	(0.0005)	(0.0005)
Observations	167,452	167,452	72,410	72,410
R^2	0.02	0.02	0.00	0.01

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. All student pairs who allocated to the same first year tutorial have been excluded.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
6. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.3: Heterogeneity in Exposure Effect by Foreign Characteristics (Working Groups)

	Friends Yes (1) or No (0)			
	Working Group Pair Data			
	(1)	(2)	(3)	(4)
First Year	0.0205***	0.0198***	0.0198***	0.0233***
	(0.0018)	(0.0033)	(0.0033)	(0.0040)
Cultural Dist.		0.0005	0.0007	
		(0.0008)	(0.0011)	
First Year × Cultural Dist.		-0.0093***	-0.0093***	
		(0.0023)	(0.0023)	
Top 25% Cult. Dist.				0.0015
				(0.0022)
First Year × Top 25% Cult. Dist.				-0.0159***
				(0.0055)
Unconditional Mean	0.0171	0.0171	0.0171	0.0171
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Country Controls			✓	✓
Observations	42,372	42,372	42,372	42,372
R^2	0.00	0.01	0.01	0.01

Notes:
1. All regressions include block/course-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs are included in the regressions.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. Country controls include the geographical distance in kilometers between the country of origin of the foreigner and the university campus, and a measure of common languages spoken between the country of origin of the foreign student and the Netherlands.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.4: Heterogeneity in Exposure Effect by Native Characteristics (Working Groups)

	Friends Yes (1) or No (0)						
	Working Group Pair Data						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First Year	0.0138***	0.0138***	0.0138***	0.0164***	0.0138***	0.0138***	0.0145***
	(0.0036)	(0.0036)	(0.0036)	(0.0040)	(0.0036)	(0.0036)	(0.0038)
Vote % Far-Right		0.0011	0.0021				
		(0.0009)	(0.0019)				
First Year ×		-0.0038	-0.0038				
		(0.0027)	(0.0027)				
Top 25% Far-Right				0.0062**			
				(0.0030)			
First Year × Top 25% Far-Right				-0.0151**			
				(0.0058)			
Pop. Dutch					-0.0002	-0.0010	
					(0.0007)	(0.0012)	
First Year × Pop. Dutch					0.0012	0.0013	
					(0.0025)	(0.0025)	
Top 25% Pop. Dutch							-0.0026
							(0.0025)
First Year × Top 25% Pop. Dutch							-0.0026
							(0.0069)
Unconditional Mean	0.0149	0.0149	0.0149	0.0149	0.0149	0.0149	0.0149
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Municipality Controls			✓	✓		✓	✓
Observations	30,822	30,822	30,822	30,822	30,822	30,822	30,822
R^2	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Notes:
1. All regressions include block-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs are included in the regressions.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. Municipality controls include the average vote share for the other major political parties in 2010, the geographic distance in kilometers from the municipality to the university campus, the percentage of the municipality who have an undergraduate university degree or higher, and the percentage of the municipality who collect unemployment benefits.
8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.5: Multiplier Effect of Exposure (Working Groups)

	Friends Yes (1) or No (0)			
	Working Group Pair Data			
	(1)	(2)	(3)	(4)
Prop. Foreigner	0.0012			
	(0.0009)			
Prop. Native	0.0001			
	(0.0009)			
Prop. Foreigner (Same Gen.)		0.0013	0.0012	0.0019
		(0.0014)	(0.0014)	(0.0019)
Prop. Native (Same Gen.)		0.0001	0.0001	-0.0011
		(0.0013)	(0.0013)	(0.0016)
Prop. Foreigner (Same Gen.) $\times$ Prop. Native (Same Gen.)			-0.0018	
			(0.0016)	
Top 25% Far-Right				0.0024
				(0.0030)
Top 25% Cult. Dist.				0.0003
				(0.0030)
Prop. Foreigner (Same Gen.) $\times$ Top 25% Far-Right				-0.0019
				(0.0030)
Prop. Foreigner (Same Gen.) $\times$ Top 25% Cult. Dist.				0.0017
				(0.0029)
Prop. Native (Same Gen.) $\times$ Top 25% Far-Right				-0.0035
				(0.0031)
Prop. Native (Same Gen.) $\times$ Top 25% Cult. Dist.				0.0017
				(0.0031)
Unconditional Mean	0.0140	0.0175	0.0175	0.0153
	(0.0006)	(0.0175)	(0.0175)	(0.0153)
Observations	42,372	42,372	42,372	30,822
R^2	0.00	0.01	0.01	0.01

Notes:
1. All regressions include block-cohort fixed effects.
2. The unit of analysis is a student-pair in a particular block. The outcome variable is one if the students are classified as friends in that particular block. Only native-foreign student pairs not allocated to the same first year study group are included in the regressions. Columns (2), (3), and (4) consider only pairs of the same gender.
3. Models are estimated with OLS.
4. Standard errors in parentheses, clustered based on a variable that takes upon unique values for every combination of first-year tutorial groups of each student pair.
5. All continuous variables are standardized.
6. Unconditional mean refers to the unconditional mean of the outcome variable. Standard error reported in parentheses.
7. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.6: Proportion of Foreigners in First Year Study Group Balancing Tests

	Prop. Foreigner	Prop. Foreigner (Euro.)	Prop. Foreigner (Non-Euro.)
	(1)	(2)	(3)
Student Number	-0.0857	-0.0137	-0.0872
	(0.1510)	(0.0977)	(0.1423)
Female	-0.0363	-0.0395	0.0059
	(0.0370)	(0.0324)	(0.0420)
Age	0.0235	-0.0040	0.0338
	(0.0330)	(0.0250)	(0.0274)
Observations	961	961	961
R^2	0.175	0.353	0.150
F-test	0.48	0.57	0.56
p-value	0.700	0.635	0.644

Notes:
1. Regressions show the relationship between the background characteristics and proportion of foreigners in a student's first year study group. All regressions include cohort fixed effects.
2. The dependent variable is shown at the top of each column. The dependent variables, $StudentNumber$ and Age are standardized.
3. The F-tests, and corresponding p-values, refer to a test for the joint significance of all background characteristics.
4. Standard errors in parentheses, clustered on the first year study group level.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3.7: Clustering of Background Characteristics Test

	(1) Native	(2) Student Number	(3) Age	(4) Female
Study Group 1	0.1333	-0.0404	-0.0587	0.0111
	(0.1486)	(0.0607)	(0.2962)	(0.1437)
Study Group 2	0.0000	0.0545	0.0278	-0.0556
	(0.1601)	(0.0654)	(0.3191)	(0.1548)
Study Group 3	0.0652	0.0260	0.0230	0.1486
	(0.1455)	(0.0594)	(0.2899)	(0.1406)
Study Group 4	-0.0652	-0.0253	0.1091	-0.0272
	(0.1455)	(0.0594)	(0.2899)	(0.1406)
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Observations	961	961	961	961
R^2	0.046	0.960	0.055	0.033
F-test	1.09	1.30	1.03	0.56
p-value	0.331	0.113	0.423	0.982

Notes:
1. Regressions include cohort fixed effects and dummies for the first year study group. No further controls are included.
2. The dependent variable is shown at the top of each column. The $StudentNumber$ and Age variables are standardized.
3. The F-test, and corresponding p-value, refer to a test for the joint insignificance of the study group dummies. The test is for whether a large model with both cohort dummies and study group dummies can explain the variance in the background characteristics better than a small model with only cohort dummies.
4. Standard errors in parentheses.
5. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 4

Health Disparities by Income in Spain Before and After the Economic Crisis

Joint work with Pilar García-Gómez, Eddy Van Doorslaer and Tom van Ourti

Published in Health Economics

4.1 Introduction

The Great Recession that started in 2008 was for most OECD countries the worst economic contraction since the 1930s (Jenkins et al., 2012). While falling incomes and rising unemployment have been the most visible consequence of the crisis, an additional concern is whether any effects have been unequally spread across the income distribution.

The importance of studying inequalities, both in income and other dimensions, is widely appreciated. While the European Union has targeted health inequality reduction as a key policy goal and warned of "*the negative consequences for health, social cohesion and economic development if health inequalities are not effectively tackled*" (European Commission, 2009, p. 5), the crisis has interfered with the execution of some of these policies (European Commission, 2013).

The aim of this paper is to examine what has happened to the social gradient in health before and after the crisis. We focus on Spain, one of the EU countries confronted with a severe economic recession, and employ a decomposition method that has been used to examine the evolution of income, health and inequality during a period of rapid growth in China (Baeten et al., 2013). We use new EU Statistics on Income and Living Conditions (SILC) panel data spanning the period 2004-2012, including both a period of substantial economic growth (2004-2007) as well as of recession (2009-2012). We examine the extent to which the evolution of health disparities by income was associated with changes in income growth, in income inequality and in differential income mobility of various socio-demographic groups.

The economic growth pattern in Spain since the mid-1990s can be summarized by a pre-crisis trend and a post-crisis trend. Figure 4.1 shows the evolution of unemployment (right axis), and GDP growth and real average annual wage growth (left axis) between 2002 and 2013. Prior to the crisis, the country experienced extended economic expansion with real GDP growing at approximately 3% per year and unemployment falling below 8% in 2007. Despite this extended period of labour demand, wage growth was minimal (Carrasco et al., 2011).

The effects of the global financial crisis become obvious in Spain beginning in 2008: when GDP growth fell from 3% to below -3% between the first quarter of 2008 and the first quarter of 2009, while unemployment roughly doubled in the same period. Youth unemployment became particularly high, with unemployment in the 15-24 age group doubling between 2007 and 2009 to stand at more than 35% (OECD, 2016b). Males, who were overrepresented in highly cyclical forms of employment, were hit disproportionally (De la Rica and Rebollo-Sanz, 2017). Jenkins et al. (2012) reaches similar conclusions, noting that the largest fall in employment in Spain during the crisis period was concentrated among young people under the age of 25, especially young men.[1]

Particularly important to understanding both the boom and bust years in Spain is the expansion and collapse of the housing market. The nominal house price per square meter in Spain tripled between 1997 and 2007 (Bonhomme and Hospido, 2012). Parallel to this housing boom was an expansion in the construction sector. Between 1998 and 2008 the share of construction in Spain's GDP increased by 4 percentage points to 10.7% (Gonzalez and Ortega, 2013). From 1997 to 2006 the share of construction in male employment rose from 14% to more than 20% (Bonhomme and Hospido, 2012). However, in 2008 as the effects of the subprime mortgage crisis in America spread and Europe began to enter recession, the Spanish housing sector crashed. Most of the sharp rise in unemployment in Spain was due to the collapse of this sector. From the first quarter of 2008 till the last quarter of 2009, construction experienced a 20% per annum drop in employment (Bentolila et al., 2012).[2]

Alongside the changes in average levels of income there have been changes in its distribution. In general, income inequality in Spain has followed a counter-cyclical pattern – during boom years income inequality decreased while during bust years it rose (Lacuesta and Izquierdo, 2012; Carrasco

[1]In the years immediately following the crisis GDP growth rebounded slightly before falling back to negative growth, whereas unemployment steadily increased to above 25% in 2012. Real wages actually increased between 2009 and 2010. However, this was due to compositional effects, as less experienced workers with lower paying temporary contracts were the first to lose their jobs (Puente and Galán, 2014). As the crisis progressed however, real wages contracted.

[2]The reasons for the extraordinary growth until 2006, and the collapse after 2008 in the construction sector are still up for debate. See Gonzalez and Ortega (2013) and Bentolila et al. (2012) for more details.

Figure 4.1: Unemployment and GDP growth in Spain

Notes:

1. GDP growth and unemployment data taken from the Instituto National de Estadística.
2. Wage growth data taken from OECD database.

et al., 2011; Pijoan-Mas and Sánchez-Marcos, 2010). Bonhomme and Hospido (2012) show that the housing and construction sector is once again key to understanding these trends as the construction sector is one of the main employers of young, uneducated, relatively disadvantaged groups of (usually) men Aparicio Fenoll (2010).

A separate question is the extent to which the Great Recession has affected health. There is a large literature documenting that worsening economic conditions are associated with reduced mortality (Ruhm, 2000; Stuckler et al., 2009), although recent evidence has been more mixed. Ruhm (2015) suggests that the relationship may be disappearing over time, while others have found that mortality trends, except for suicides, continued to improve during the recent European-wide recession (Regidor et al., 2014; OECD, 2016a).

There is little evidence on the evolution of health inequalities during the recent crisis.[3] Most observers appear to assume that it will widen existing gaps following reductions in welfare spending that increase the vulnerability of those with lower education levels, who are also more likely to be unemployed (European Commission, 2013).

One strand of literature has focused on the comparison of income-related health inequalities (IRHI) across countries and over time. Doorslaer and Koolman (2004), for instance, documented

[3]Bacigalupe and Escolar-Pujolar (2014) review four studies on Spain during the Great Recession, and conclude that the evidence points in the direction of increasing health inequalities.

the variation in degrees of IRHI for 13 EU countries in 1996 and showed that IRHI tends to be larger in countries with larger income inequality, but also that the relative income position of Europeans that are not working and not in good health, like the retired and the disabled, was critical. Van Ourti et al. (2009) decomposed the evolution of IRHI between 1994 and 2001 for the same 13 EU countries. They found that the income elasticity of health was crucial for understanding the evolution of IRHI, although the period considered was one of economic growth for most European countries. An extended version of this decomposition was used by Baeten et al. (2013) to decompose the evolution of IRHI in China into the contributions of various factors like income growth, income inequality, and income mobility, as well as a number of regional-demographic factors associated with health. They found that the substantial rise in IRHI over the period of double-digit income growth (1991-2006) was associated with rising income inequality, but especially with the adverse health and income experience of older women lacking pension or other social protection. It is this decomposition method that we use in this paper.

Our findings indicate the following: inequality in health by income was slightly rising before the crisis, but started falling sharply after 2009 when the recession hit Spain. The main reason for this reversal is the differential effect of the crisis on the incomes of young and old Spaniards: while pensioner incomes were relatively shielded against the erosion in the post-crisis years, this does not hold for the incomes of younger groups. Loss of employment and of earnings in employment meant that these relatively healthier groups moved downwards in the income ranking, thereby lowering the association between health and income rank. As a result, IRHI in 2012 was lower than in the years prior to 2009, a somewhat surprising by-product of an otherwise discomforting period in recent Spanish history.

4.2 Decomposing the Evolution of IRHI with a Balanced Cohort

We use the decomposition method of Baeten et al. (2013) to estimate the evolution of IRHI and to shed light on the relative importance of (a) income growth, (b) the evolution of income inequality, (c) income mobility, and (d) the evolution in non-income factors (such as demographics) that are associated with health. This section describes the decomposition approach for a balanced cohort of n individuals that we observe at the start (period 1) and end (period 2) of a time interval.

4.2.1 Choice of Health Inequality Index.

We use the corrected concentration index (CC) (Erreygers, 2009) because it satisfies the mirror condition and it is insensitive to equal health additions (cf. absolute inequality) (Erreygers and Van Ourti, 2011). When health is bounded between 0 and 1, it can be written as:

$$CC(h_t|y_t) = \frac{8}{n^2}\sum_{i=1}^{n} z_{it}h_{it} \tag{4.1}$$

where h_t and y_t are the health and income distribution in period $t = 1, 2$, h_{it} describes the health of individual i and z_{it} is a weight that depends on the income rank of individual i. This weight takes the value 0 for the individual with median income, and increases (decreases) linearly for individuals with higher (lower) than median income levels.[4]

4.2.2 Descriptive Model for Health.

We use a simple descriptive[5] model that links health linearly and additively to its associated factors:

$$h_{it} = \alpha + \theta(y_{it}) + x'_{it}\beta \tag{4.2}$$

where α is an intercept parameter; $\theta(y_{it})$ is a non-linear function of income; x_{it} represents a vector of K non-income variables (e.g. demographics), and β is its associated parameter vector.[6] It is important to allow for a very general functional form for $\theta(.)$ since the actual functional form will largely determine the relative importance of the contribution of (a) income growth and (b) income inequality in our decomposition approach.

4.2.3 Evolution of IRHI Over Time.

Our interest lies in decomposing changes in IRHI. Combining Equation (4.2) and Equation (4.1) leads to[7]:

[4] z_{it} takes $(1-n)/2$ for the poorest individual and $(n-1)/2$ for the richest individual.

[5] It should be stressed that our goal is not to estimate a causal model of health; our sole aim is to decompose changes in the (partial) association between health and income rank. As we have neglected to include potentially endogenous variables such as education or lifestyle the non-linear income function features as the sole potentially endogenous variable. We have deliberately not addressed its potential endogeneity since we are interested in documenting the association between changes in the distribution of income and the evolution of IRHI in Spain. Turning to the underlying mechanisms is only sensible after the magnitude of this association has been established, and after the relative importance of "income growth", "mean-preserving income changes", and "income mobility" has been understood.

[6] After results are presented we return to the assumption that Equation (4.2) is deterministic.

[7] The intercept parameters drop out in Equation (4.3) since $\sum_{i=1}^{n} z_{it} = 0$.

$$
\begin{aligned}
CC(h_2|y_2) - CC(h_1|y_1) &= \frac{8}{n^2}\Big[\sum_{i=1}^{n} z_{i2}h_{i2} - \sum_{i=1}^{n} z_{i1}h_{i1}\Big] \\
&= \frac{8}{n^2}\sum_{i=1}^{n}\Big\{\Big[z_{i2}\theta(y_{i2}) - z_{i1}\theta(y_{i1})\Big] + \beta\Big[z_{i2}x_{i2}^{'} - z_{i1}x_{i1}^{'}\Big]\Big\}
\end{aligned}
\tag{4.3}
$$

Equation (4.3) shows that we can disentangle the change in IRHI into a part due to changes in the association between the income rank and the income effect $(z_{i2}\theta(y_{i2}) - z_{i1}\theta(y_{i1}))$ and a part due to changes in the association between the income rank and the non-income factors $(z_{i2}x'_{i2} - z_{i1}x'_{i1})$. In order to isolate the contributions of (a) income growth, (b) the evolution of income inequality, (c) income mobility, and (d) the evolution in non-income factors, we construct two hypothetical health states in period 2 using Equation (4.2) – health under average income growth (h_{i2}^{ag}) and health under no income growth (h_{i2}^{ng}). For the former, we calculate an individual's health in period 2 in the scenario that everyone's income changed proportionally to the average income gain (or loss) between period 1 and period 2. We denote this income as y_{i2}^{ag}. For the latter, we estimate an individuals health in period 2 in the scenario that there was no income change between period 1 and period 2 y_{i2}^{ng}. In each scenario we allow all non-income variables to change as they actually did.

4.2.4 Decomposition Method.

Given that $CC(h_2^{ag}|y_2^{ag}) = CC(h_2^{ag}|y_1)$ and $y_{i2}^{ng} = y_{i1}$, the change in IRHI can be expressed as:

$$
\begin{aligned}
CC(h_2|y_2) - CC(h_1|y_1) = &\underbrace{CC(h_2|y_2) - CC(h_2^{ag}|y_1)}_{\text{income inequality \& mobility}} + \underbrace{CC(h_2^{ag}|y_1) - CC(h_2^{ng}|y_1)}_{\text{average income growth}} \\
&+ \underbrace{CC(h_2^{ng}|y_1) - CC(h_1|y_1)}_{\text{non-income factors}}
\end{aligned}
\tag{4.4}
$$

which can be further disentangled as the sum of 4 terms (note that $z_{i2}^{ag} = z_{i2}^{ng} = z_{i1}$):

$$
\begin{aligned}
&CC(h_2|y_2) - CC(h_1|y_1) \\
&= \frac{8}{n^2}\sum_{i=1}^{n}\Bigg\{\underbrace{z_{i1}\big[\theta(y_{i2}^{ag}) - \theta(y_{i1})\big]}_{\text{average income growth}} + \underbrace{\big[z_{i2}\theta(y_{i2}) - z_{i1}\theta(y_{i2}^{ag})\big]}_{\text{income inequality}} \\
&\quad \underbrace{+ (z_{i2} - z_{i1})\Big(\sum_{k=1}^{K}\beta^k x_{i2}^k\Big)}_{\text{income mobility}} + \underbrace{z_{i1}\Big[\sum_{k=1}^{K}\beta^k(x_{i2}^k - x_{i1}^k)\Big]}_{\text{non-income factors}}\Bigg\}
\end{aligned}
\tag{4.5}
$$

Equation (4.5) shows that the evolution of IRHI can be written as the sum of (a) average income growth, (b) the evolution of income inequality, (c) income mobility, and (d) the evolution in non-income factors.

The first term, average income growth, captures the change in IRHI when everyone's income changes in proportion to the average income change. As all incomes grow proportionally, there is no change in the rankings (z_{it}'s). Therefore this term captures whether the health responsiveness to proportional income changes ($\theta(y_{i2}^{ag}) - \theta(y_{i1})$) is, on average, larger or smaller for those with lower (negative z_{i1}) versus higher incomes (positive z_{i1}) in period 1. If the health responsiveness is larger for the initially richest part of the population, then this term will be positive. The sign (and magnitude) of this term depends on the functional form of $\theta(.)$, but also on whether incomes have increased or decreased on average.

The second term captures the evolution of income inequality – that is, the health difference attributed to the difference between the true income in the second period and the income under the scenario of average income growth ($\theta(y_{i2}) - \theta(y_{i2}^{ag})$). If the health returns from income growth are increasing with income ($\theta'(.) > 0$), if there is no income re-ranking ($z_{i2} = z_{i1}$) and if – relative to the average income growth scenario – the rich become richer while the poor loose, then the second term will be positive. In a scenario with income re-ranking, one cannot a priori assign a direction to term 2.

Term 3 – "income mobility" – captures the association between income re-ranking ($z_{i2} - z_{i1}$) and the non-income factors in the second period, weighted by the β^k coefficients. One can further decompose term 3 into separate contributions for each non-income variable since term 3 is additively separable. In our empirical application, the non-income variables are dummy variables. In this case, the contribution of each non-income dummy can be large (compared to the reference category) because (a) health is considerably higher or lower among the individuals belonging to the non-income dummy (β^k), (b) income re-ranking is substantial for these individuals ($z_{i2} - z_{i1}$), and/or (c) a substantial share of the sample belongs to this non-income dummy ($\sum_{i=1}^{n} x_{i2}^{k}$).

Term 4 measures the association between changes in non-income factors and initial income ranks. If the non-income factors include age and location, then term 4 isolates the effect of ageing and migration on the change of IRHI. For example, if many people with high initial income ranks migrate to a location which is associated with better health then this term will be positive. In what follows we refer to the terms 1, 2, 3 and 4 as income growth, income inequality, income mobility, and non-income factors.

Table 4.1: Rotation Groups Overview

	Rotation Group					
	1	2	3	4	5	6
2004	✓					
2005	✓	✓				
2006	✓	✓	✓			
2007	✓	✓	✓	✓		
2008		✓	✓	✓	✓	
2009			✓	✓	✓	✓
2010				✓	✓	✓
2011					✓	✓
2012						✓

Notes:
1. Table presents the structure of the EU-SILC rotation groups for Spain between 2004 and 2012. Ticks indicate the years covered by each rotation group.

4.3 Data and Empirical Implementation

We use 6 rounds of the Spanish EU-SILC dataset spanning 2004-2012. It includes the period before and during the financial and economic crisis that affected Spain from 2009 onwards. As the EU-SILC dataset is set up as a rotating panel every year between 2004 and 2009, a new random sample is drawn and followed for 4 years, after which it is dropped. We use the term "rotation group" for each of these random samples. For example, the first rotation group is drawn in 2004 and lasts till 2007; the second covers 2005-2008; and the sixth and last rotation group covers 2009-2012. Hence, for the full period of 2004-2012 we have 6 rotation groups in total, and these constitute different and independent samples.[8] In total, we have 122,592 observations (see Table 4.1 for more details).

4.3.1 Key Variables.

The two main variables of interest are self-assessed health (SAH) and household income. The SAH responses derive from the question: "*How is your health in general? Is it: (1) very good, (2) good, (3) fair, (4) bad, (5) very bad?*" As our income variable we use total disposable household income during the previous 12 months. We adjust for household size and inflation by dividing by the square

[8] In 2004 and 2012, we observe only one rotation group (group 1 & 6); in 2005 and 2011 we simultaneously observe 2 rotation groups (group 1/2 & 5/6), in 2006 & 2010 3 rotation groups (group 1/2/3 & 4/5/6), while for the years 2007, 2008 and 2009, we simultaneously observe 4 separate rotation groups (group 1/2/3/4; 2/3/4/5 & 3/4/5/6).

root of household size[9] and by applying the Spanish CPI index with base year 2012. We remove observations with negative incomes.[10]

4.3.2 Estimating the Health Model.

We estimate the model for health in Equation (4.2) using an interval regression (with threshold values imposed as in Van Doorslaer and Jones (2003)), as the CC computation requires a health indicator measured on a cardinal scale.[11] The predicted values are used as our main health indicator, and can be interpreted as predicted health utility index (HUI) scores (Van Doorslaer and Jones, 2003).[12] We use a second degree polynomial for the income function $\theta(y_{it})$ as it is a parsimonious functional form that is sufficiently flexible to avoid predetermining the effect of proportional income changes on health.[13] The remaining independent variables are dummy variables for each region in Spain and age category dummies for both males and females.[14] Age is categorised into the following groups: 16 to 26 years, 26 to 36 years, 36 to 46 years, 46 to 56 years, 56 to 66 years, 66 to 76 years, 66 to 76 years, and more than 76 years of age.

4.3.3 Empirical Implementation of Decomposition Method.

Because of the rotating design of EU-SILC we cannot directly compare IRHI measured for the same individuals in 2004 and 2012. This complicates both the implementation of the decomposition and the estimation of the empirical health model. The decomposition requires at least two observations of the same individual over time. We apply the decomposition to each of the 6 rotation groups separately

[9]Decompositions using alternative equivalence scales, such as the OECD-modified scale, did not significantly change the results.

[10]Negative incomes can occur in the EU-SILC data due to debt, but make up less than 1% of the observations. They are problematic as in the hypothetical average income movement scenario these individuals will see their incomes drop when on average incomes rise. However, decompositions that included these observations did not change the qualitative features of our results.

[11]This involves estimating an ordered probit model, with the thresholds imposed from the empirical distribution function of HUI in the *Canadian National Population Health Survey* 1994-1995 (HUI=1 equals maximum health and HUI=0 equals minimum health). Several studies using this approach (Van Doorslaer and Jones, 2003; Lauridsen et al., 2004; Lecluyse and Cleemput, 2006) have found the health inequality estimates to be rather insensitive to the threshold values imposed.

[12]While the predicted HUI scores only reflect health changes resulting from changes in the explanatory variables, Van Doorslaer and Jones (2003) show that the interval regression approach is the preferred approach when calculating health inequality indices. One might also calculate the conditional predictions from the interval regression model given the observed SAH levels, but then the predicted HUI scores would no longer be a linear combination of the explanatory variables, and therefore not be amenable to our decomposition approach.

[13]As explained before, the signs and size of term 1 and 2 (“income growth” and “evolution of income inequality”) largely depend on whether the health responsiveness to proportional income changes decreases or increases with rising incomes. This is left open with a second order income polynomial, but not with other popular choices in the empirical literature. For example, when one would favour the natural logarithm of income, one would impose that a proportional change in income has the same health effect for every individual (and hence one would force term 1 to be zero).

[14]The EU-SILC categorises Spain into 18 different regions: Galicia, Asturias, Cantabria, País Vasco, Navarra, La Rioja, Aragón, Madrid, Castilla y León, Castilla-La Mancha, Extremadura, Cataluña, Comunidad Valenciana, Baleares, Andalucía, Murcia, Ceuta y Melilla, and Canarias.

Table 4.2: Balanced Panel Observations per Rotation Group

Rotation Group	Individuals per Wave	Total Observations
1	4,193	16,772
2	4,996	19,984
3	5,099	20,396
4	5,575	22,300
5	5,617	22,468
6	5,168	20,672
Total		122,592

Notes:
1. Table presents the number of unique individuals, and the total number of observations per rotation group.

and within each rotation group to a balanced panel only.[15] While we calculate the decomposition for each of the 6 rotating panels, we only present three of these; a before crisis group: 2004-2007; a group covering both before and when the crisis occurs: 2007-2010; and a group that covers the crisis period: 2009-2012.[16]

We first pool the data from all 6 rotation groups and run the interval regression model described above.[17] We remove the individuals belonging to the top 1% of incomes (calculated on the full pooled sample) as these observations have a disproportionate effect on the functional form of income.[18] The estimated parameters of the pooled model are then used to decompose the 3 rotation groups which span the entire 2004-2012 period, leaving us with the observations per rotation group as shown in Table 4.2. Each of the rotating panels uses a different base year. In the 2004-2007 rotation group, we first compare the change in IRHI for 2004-2005, then 2004-2006, then 2004-2007. We next take the second rotation group (which spans 2007-2010) and compare the change in IRHI between 2007 and each following year. For the 2009-2012 rotation group, 2009 is the base year. In total there are then 9 changes of IRHI to be decomposed.

We use the sample weights of the first year of each rotating panel provided with the EU-SILC data. In the interval regression model, we also allow for robust standard errors and cluster at the

[15] Summary statistics of the full unbalanced panel sample are similar to those of the balanced panel. Nor did the evolution of IRHI using the unbalanced panels for each rotating group differ markedly, suggesting that attrition is not an important driver of our main findings, although we cannot entirely rule out that explicitly accounting for mortality as in Petrie et al. (2011) would have disproportionally hit the older and poorer age groups.

[16] Decomposition results for the 2005-2008, 2006-2009 and the 2008-2011 rotation groups are not presented for reasons of clarity and brevity. They are in line with the results presented and available upon request.

[17] The assumption of constant coefficients may be questionable in the case of pre- and post-crisis Spain, since the relationship between income and health may have changed. To test the robustness of our main findings we also decomposed the periods 2004-2007 using coefficients estimated on pre-crisis observations (before 2009) only, and 2009-2012 using only post crisis observations (after 2008). This did not change our results.

[18] While including the top 1% of incomes does change the function form of income due to some extreme outliers (in particular among very high incomes), it does not change the overall results of the decomposition. Nevertheless in order to achieve an income function that is not unduly influenced by outliers we remove the top 1% of incomes.

individual level. Statistical inference of the decomposition method is obtained after bootstrapping the entire procedure with 1,500 replications. The bootstrap sampling is bias-corrected, and clustered at the primary sampling unit of the EU-SILC.

4.4 Results

4.4.1 Summary Statistics.

Table 4.3 displays variable means for each wave of rotation group 1, 4 and 6. Panel (a) includes variables most important to the analysis, whereas panel (b) provides additional background information on the labour market. The health variable refers to the predicted HUI score.

Income is rising in each successive year for rotation group 1, as well as rotation group 4 until 2009. As income refers to the last 12 months, the drop observed in 2010 refers to an income fall in 2009, during which Spain was fully immersed in the economic crisis. In rotation group 6 income falls in each wave compared to the last. The effect of the crisis is also visible in the occupational category changes. The proportion unemployed in 2009 almost doubles from the previous year to approximately 11%. In subsequent years the proportion of employed individuals decreases every year. This does not appear to be due to ageing and retiring individuals; while the proportion of retirees does increase slightly, it is the unemployed category that shows the sharpest increase.

Income inequality, as measured by the Gini coefficient was rather stable, although opposite trends can be observed before and after 2009. Income inequality appears to have been slightly falling during the"boom" years, and began to increase once the crisis started. This is in line with the findings of others, such as Jenkins et al (2013).

Figure 4.2, with the CC per year for each rotation group, shows that IRHI has not been stationary over the sample period.[19][20] Until 2009, there is a slightly significant upward trend.[21] Since the beginning of the crisis, however, IRHI fell quite steeply. This is confirmed by comparatively large and significant falls in IRHI in the final two rotation groups.

Column 1 of Tables 4.4 and 4.5 shows the coefficients from the interval regression for the age-sex and region dummies.[22] Note that older age groups consistently report lower health than younger.

[19] One should only use the 95% confidence intervals to compare IRHI between rotation groups, since different waves within each rotation group are dependent samples.

[20] The fact that similar trends are observed using different rotation groups indicates that the trend is not simply driven by a particular rotation group.

[21] Appendix Table A.4.1 shows the numerical changes of IRHI between waves for each rotation group, and indicates the significance of such changes.

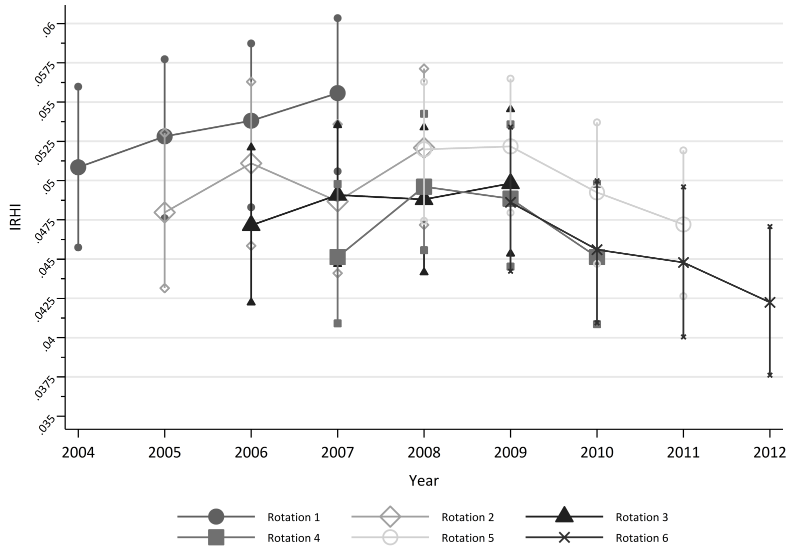
[22] The coefficients of the income polynomial are suppressed.

Table 4.3: Summary Statistics

		(a)				(b)					
Rotation Group	**Wave**	**Household Income (€)^a**	**Age^b**	**Female (%)**	**Health**	**Labour Market Status (%)^c**					**Gini**
						Employed	Self-Employed	Unemployed	Other	Disabled	
1	2004	23962.08	47.2	0.531	0.886	0.41	0.074	0.08	0.258	0.02	0.31
1	2005	25156.27	48.3	0.531	0.884	0.411	0.083	0.077	0.244	0.024	0.31
1	2006	25602.62	49.3	0.531	0.883	0.426	0.081	0.061	0.242	0.019	0.3
1	2007	26056.85	50.2	0.531	0.881	0.426	0.08	0.063	0.228	0.021	0.3
4	2007	27554.9	46.7	0.513	0.888	0.465	0.065	0.06	0.242	0.021	0.28
4	2008	28775.31	47.6	0.513	0.886	0.46	0.069	0.074	0.217	0.024	0.29
4	2009	29900.84	48.5	0.513	0.883	0.427	0.065	0.117	0.198	0.028	0.29
4	2010	28475.24	49.5	0.513	0.881	41.4	0.065	0.123	0.199	0.032	0.3
6	2009	29912.82	47.5	0.513	0.885	0.401	0.068	0.116	0.233	0.028	0.29
6	2010	28798.27	48.4	0.513	0.882	0.388	0.066	0.135	0.211	0.034	0.3
6	2011	27366.47	49.4	0.513	0.88	0.39	0.059	0.12	0.217	0.032	0.3
6	2012	26425.08	50.4	0.513	0.878	0.371	0.06	0.149	0.201	0.032	0.31

Notes:
1. Table presents the weighted means for key variables, as well as additional labour market and income inequality measures, calculated on the balanced sample of rotation group 1, 4 & 6.
2. ^{a}In 2010 Euros.
3. ^{b}The yearly ageing of the balanced sample does not always equal one since survey months differed slightly from wave to wave.
4. ^{c}Refers to current self-reported labour market status; "employed" includes full- and part-time workers, while "other" includes students, "domestic tasks", and other "inactive" persons.

Figure 4.2: IRHI per Wave per Rotation Group



Notes:

1. Bars indicate 95% confidence intervals.

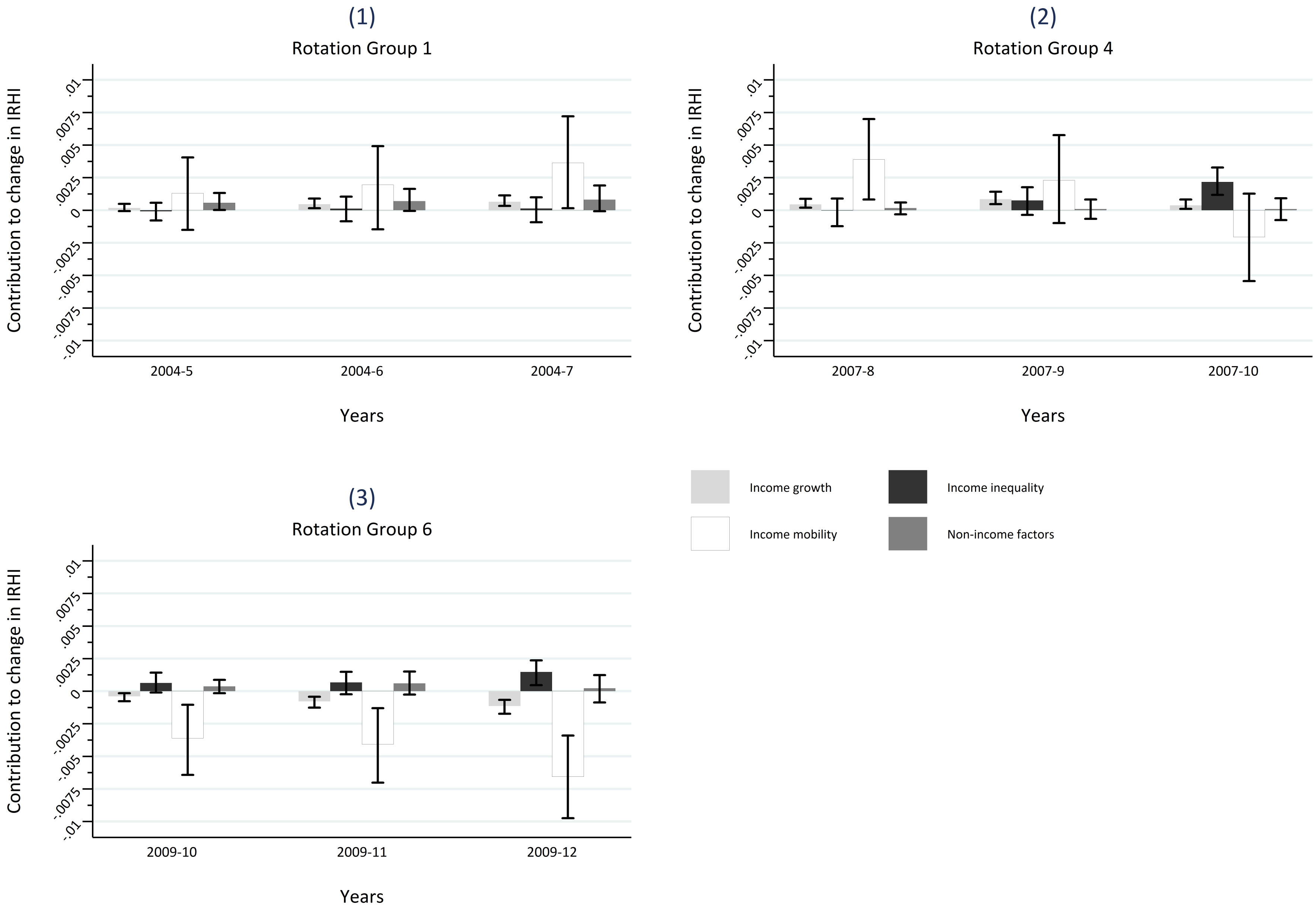
Regional health differences, by contrast, are very small. Figure 4.3 shows decomposition results for rotation groups 2004-2007, 2007-2010 and 2009-2012 with 95% confidence intervals.

4.4.2 2004-2007 Results.

Between 2004 and 2007, IRHI rose significantly. Panel (1) of Figure 4.3 shows that income growth is important in understanding this rise. The income growth term, although small, indicates that health responsiveness to proportional income growth was larger for those with higher income in 2004. Despite being the largest term in all years, income mobility only becomes significant in the 2004-2007 comparison. This implies that income re-ranking occurring prior to 2007 was not systematically related to age, gender or location, while the elderly were on average (and just borderline significantly) more likely to experience negative income re-ranking between 2004 and 2007.[23] As the elderly combine this move down the income ladder with the lowest predicted health, this led to a rise in IRHI. The evolution of income inequality and the non-income factors are unimportant for the IRHI change in this period.

[23]The fact that term 3 is so large in magnitude but still insignificant indicates that there are a small amount of very large and influential income re-rankings occurring. An individual moving from the bottom of the income distribution to the top in turn affects the rankings of the rest of the sample as well. Term 3 therefore changes dramatically when this individual is left out of the sample in a bootstrap replication.

Figure 4.3: Main Decomposition Results



Notes:

1. Bars indicate 95% confidence intervals.

4.4.3 2007-2010 Results.

IRHI grew significantly between 2007 and 2008, but returned in the subsequent two years to its 2007 level. The decreasing trend turns out to be almost entirely driven by the changing association between the age dummies and the income rank, while region is relatively unimportant (see income mobility term in panel (2) of Figure 4.3 and panel (1) of Figure 4.4). Closer inspection reveals that it is mainly influenced by the older, unhealthier, age groups. While initially, during the period of income growth prior to 2008, the elderly were falling in income rank, there is a reversal after 2008. Panel (2) of Figure 4.4 shows that this was especially true for those over 75. The income rank of the older age groups, with poorer health, increased contributing to the fall in IRHI.

Also significant between 2007 and 2010 is the contribution of income inequality. This suggests that the health effects of income gains – over and above proportional income growth – led to a rise in IRHI. Income growth is positive in each wave and remains small but significant. As average income falls in the final 2010 wave, so does the magnitude of income growth.

4.4.4 2009-2012 Results.

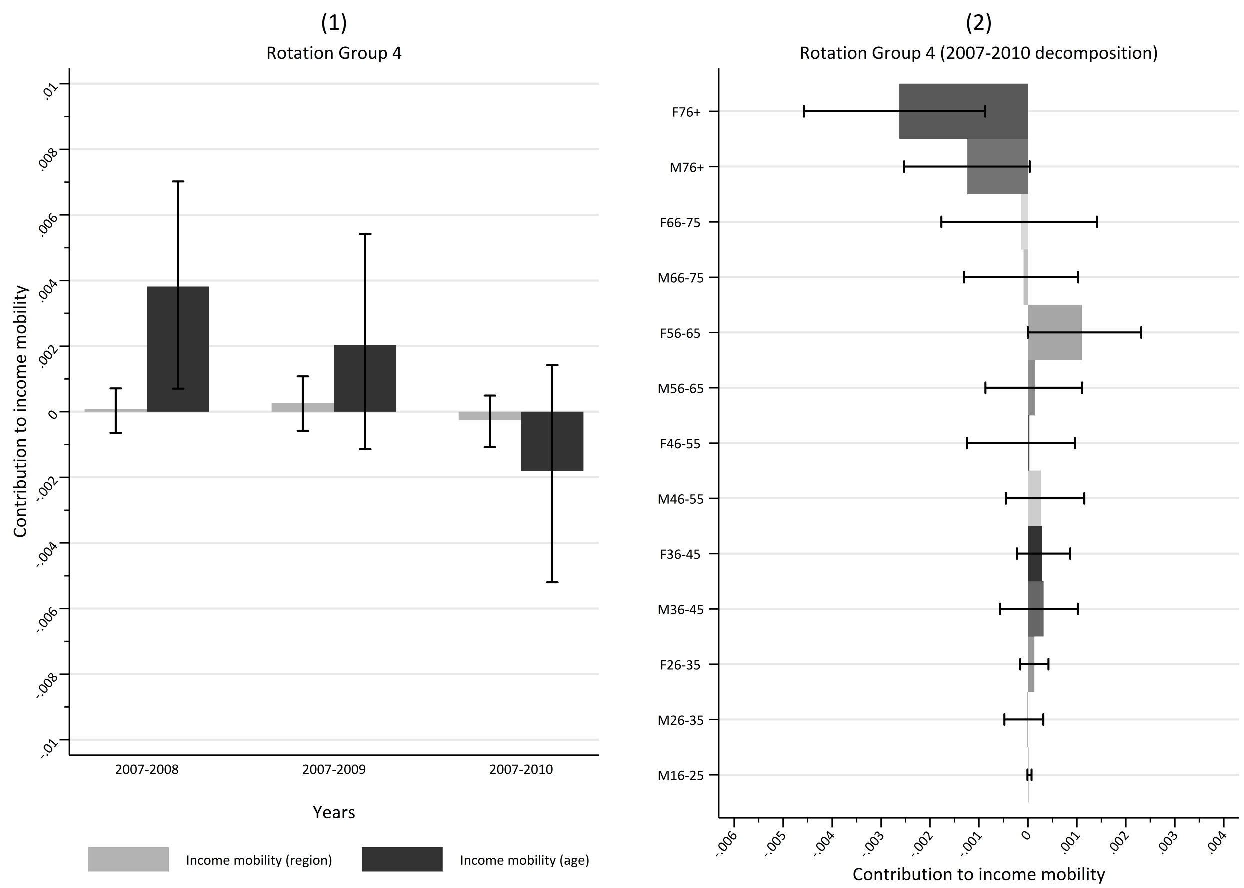
The final 4-year rotation group of the EU-SILC entirely reflects the crisis years. This is the period in which the largest drop in IRHI occurs and the trends observed in the 2007-2010 decomposition also emerge here. The significant fall in IRHI is primarily due to income mobility, which is largest in magnitude and significant in all years. Panels (1) and (2) of Figure 4.5 demonstrate that it is the experience of certain older age groups – men and women aged 66 and above – which is the largest contributor to the decrease. By contrast, the younger and healthier individuals have a small but positive contribution. This leads to a fall in IRHI as those with poorer health became relatively richer.

The 2009-2012 decomposition also reveals that both income growth and income inequality are significant drivers in the change of IRHI. The negative sign of income growth reflects the fact that had the average income fall between 2009 and 2012 been applied to everyone, those with high incomes would have had a larger fall in health than those with low incomes. Income inequality is positive however, indicating that the fall in income was disproportionately felt by the poor. Still, the overall effect of these terms relating to health responsiveness to income is small compared to income mobility.

4.5 Discusson

The decomposition results reveal two very different trends in IRHI before and after the crisis. Prior to 2009 there was a trend of increasing inequality which was mostly driven by income growth but

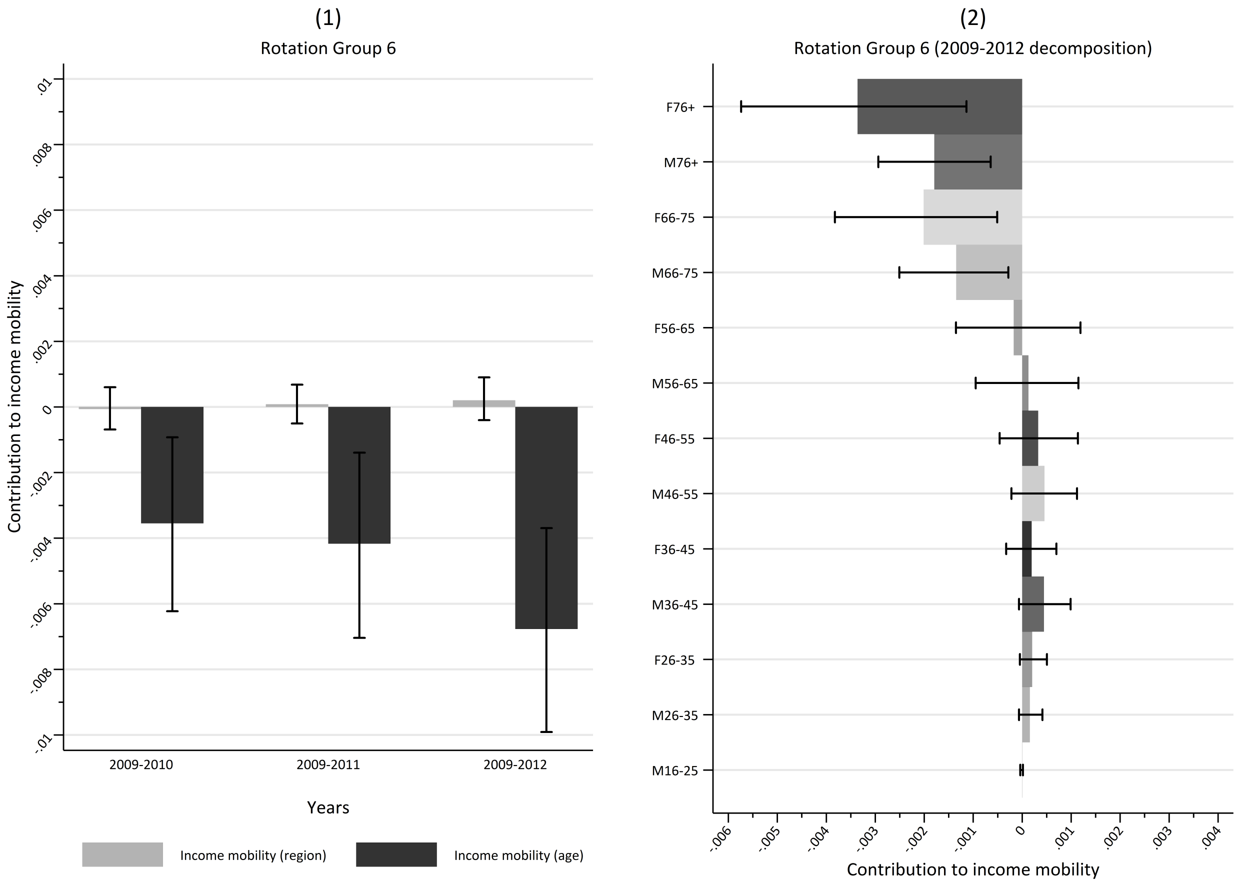
Figure 4.4: Decomposition of Income Mobility for Rotation Group 4



Notes:

1. Bars indicate 95% confidence intervals.

Figure 4.5: Decomposition of Income Mobility for Rotation Group 6



Notes:

1. Bars indicate 95% confidence intervals.

also by income mobility, with the elderly slightly moving down on the income ranks. After the start of the 2008 financial crisis we observe a sharp fall in IRHI. Income mobility is the main driver of this change: young and middle-aged healthier groups experienced a greater income drop, while on average the incomes of the elderly were less affected. This caused shifts in the income ranks in favour of the older, less healthy group, leading to a decrease in IRHI.

Further decomposing the contribution of each regional-demographic group to income mobility reveals the relative importance of three distinct underlying mechanisms. Indeed, each sub-term depends on three elements – the partial association between the group and health (β^k), the number of individuals in that particular group, and the changes in income ranks between the two periods for these individuals ($z_{i2} - z_{i1}$). Any changes in income mobility result from some combination of these elements. Tables 4.4 and 4.5 presents results for each of these three elements per demographic group and region, respectively.

Column 1 of these tables report the interval regression coefficients and Columns 2 to 7 the percentage shares of each regional-demographic group for the first and final years of each rotation group, while columns 8 to 10 report the income re-ranking for each regional-demographic group. The results for income re-ranking are obtained by running a simple no-constant OLS using the regional-demographic variables as explanatory variables for the change in individual z-scores between the two periods.[24] A positive coefficient implies a rise in income rank between the two periods.

Table 4.4 confirms that the income re-ranking of the elderly, in particular after the onset of the crisis, is most important for understanding changes in IRHI due to income mobility. Between 2004 and 2007, there was little re-ranking taking place, although the negative coefficients for the elderly indicate that, if anything, the elderly were slightly losing relative to young. In the final rotation group however the coefficients of the 65+ have become highly significant and switched sign. This, combined with the comparatively large negative coefficient of the 65+ in the health regression, and the sizable and increasing number of individuals in this category, leads to a large fall in income related health inequality.[25]

The primary reason that the elderly's incomes were better protected during the crisis appears to be the old-age pension system. In Spain, the vast majority of pensioners receive their incomes from contributory pensions based on earnings prior to retirement (OECD, 2013). As a consequence, cur-

[24] The change in z-scores is bounded between -2 and 2 since the z-scores have been normalized between -1 and 1. For example, the most extreme case of an individual going from the highest to the lowest rank would lead to $z_{i2} - z_{i1} = (-1) - 1 = -2$.

[25] One may question whether IRHI due to natural ageing is interesting or important, since ageing is an unavoidable biological process. In this case, the decomposition method can be viewed in different ways. If we are interested in the evolution of total IRHI then the sum of all 4 terms should be considered. If we wish to exclude the effect of natural ageing then we should exclude the non-income factors term. If we wish to narrow our focus further, and ignore that part of the evolution of IRHI that is due to the mobility of different age groups then the income mobility term should also be excluded.

Table 4.4: Age/Sex Factors Influencing Term 3

Variable	Coefficient	Share of Individuals in Age-sex Category (%)						Re-ranking Coefficient[a,b]		
		Rotation 1		Rotation 4		Rotation 6				
		2004	2007	2007	2010	2009	2012	2004-2007	2007-2010	2009-2012
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female										
16-26	Reference	6.21	3.75	6.03	3.93	5.4	3.53	0.0152	0.0347	-0.0049
26-36	-0.0146***	8.81	8.19	9.27	8.41	8.98	8	0.000912	-0.0143	-0.023
36-46	-0.0286***	11.12	10.92	10.05	9.98	9.66	9.37	0.0106	-0.012	-0.00803
46-56	-0.0515***	8.65	9.55	8.4	9.27	8.65	9.28	-0.00071	-0.00066	-0.00675
56-66	-0.0816***	7.05	7.64	6.71	7.01	7.83	7.83	-0.00994	-0.0181	0.00274
66-76	-0.114***	6.52	6.4	5.82	6.19	5.27	6.22	-0.00335	0.00199	0.0279*
75+	-0.156***	4.75	6.67	5.02	6.5	5.49	7.03	-0.00251	0.028**	0.0354***
Male										
16-26	0.00117	5.97	4.11	5.98	3.94	6.12	4.45	0.0141	0.0272	-0.00259
26-36	-0.0125***	8.29	7.29	10.76	9.81	9.47	7.87	0.015	0.00154	-0.0222
36-46	-0.0272***	10.01	9.81	10.28	10.39	9.78	10.47	-0.00444	-0.0163	-0.0199
46-56	-0.0407***	8.22	8.9	7.79	8.69	8.45	8.78	0.00478	-0.00821	-0.0138
56-66	-0.0667***	6.53	7.12	5.97	6.38	7.19	7.61	-0.0136	-0.00323	-0.00266
66-76	-0.0853***	5.21	5.79	4.9	5.09	4.55	4.98	-0.0129	0.00211	0.0286*
75+	-0.118***	2.65	3.87	3.03	4.4	3.17	4.55	-0.00728	0.0256*	0.0352**

Notes:
1. [a]Re-ranking coefficient refers to a no-constant regression where change in rank is regressed on the demographic variables in the final period
2. [b]The coefficients are not jointly significant for the 2004-2007 model, while they are for the 2007-2010 and 2009-2012 model.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.5: Regional Factors Influencing Term 3

Variable	Coefficient	Share of Individuals in Age-sex Category (%)						Re-ranking Coefficient^{a,b}		
		Rotation 1		Rotation 4		Rotation 6				
		2004	2007	2007	2010	2009	2012	2004-2007	2007-2010	2009-2012
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Region										
Madrid	Reference	4.13	4.14	13.14	13.3	12.22	12.28	0.0455*	-0.025	-0.00799
Galicia	-0.0267***	7.7	7.7	7.27	7.32	6.43	6.43	-0.00405	0.023	0.00503
Asturias	-0.0048	3.36	3.35	3	2.98	2.71	2.7	0.00533	0.00578	0.00896
Cantabria	0.00449	0.9	0.9	1.28	1.29	1.44	1.42	0.0148*	-0.0149	-0.00627
País Vasco	0.0023	3.69	3.63	3.66	3.67	3.94	3.95	-0.00055	0.0417***	0.0148
Navarra	0.00135	2.01	2.01	1.65	1.65	1.3	1.3	0.00288	-0.00156	0.02*
La Rioja	-0.000513	0.83	0.83	0.69	0.69	0.88	0.96	0.00146	0.00245	-0.00226
Aragón	-0.000675	3.44	3.41	3.36	3.35	3.23	3.23	-0.00034	0.0285**	0.0261*
Castilla y León	-0.0012	7.13	7.29	6.5	6.49	6.39	6.31	0.00815	0.0133	0.00948
Castilla-La Mancha	-0.000617	4.74	4.67	4.89	4.82	5.17	5.2	-0.0287**	-0.0122	-0.0146
Extremadura	0.00203	3.54	3.42	2.7	2.7	2.75	2.82	-0.0152*	0.0274***	0.0113
Cataluña	0.000396	16.9	16.9	15.09	14.93	15.12	15.19	-0.0153	-0.0109	0.0144
Valenciana	-0.00439	12.6	12.69	10.11	10.13	12.44	12.22	0.0222	-0.0620***	-0.0222
Baleares	0.00554*	2.28	2.25	2.19	2.19	1.92	1.87	-0.00624	0.0171	0.0133
Andalucía	-0.00871***	18.36	18.47	17.44	17.41	16.33	16.39	0.0128	0.0306*	-0.015
Murcia	-0.00974***	4.52	4.46	3.11	3.18	3.41	3.41	-0.0123	-0.022	-0.0267
Ceuta y Melilla	-0.0148***	0.29	0.29	0.24	0.24	0.27	0.27	0.0052	-0.00152	-0.0028
Canarias	-0.0119***	3.57	3.57	3.7	3.68	4.06	4.06	-0.0298***	-0.0344*	0.012

Notes:
1. ^{a}Re-ranking coefficient refers to a no-constant regression where change in rank is regressed on the demographic variables in the final period
2. ^{b}The coefficients are not jointly significant for the 2004-2007 model, while they are for the 2007-2010 and 2009-2012 model.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

rent economic conditions have little immediate effect on retiree incomes. Moreover, any potential changes to pension benefits are delayed by political processes and reforms are not applied retrospectively. Thus, in spite of a series of reforms that took place during the last decade in Spain, existing pensioners' incomes have remained relatively untouched.[26]

4.5.1 Role of Labour Market Status and Occupation.

Our results thus far indicate that the pre-crisis rise and post-crisis fall in IRHI were largely related to differential income mobility. In this section we explore how income mobility is associated with labour market status and occupation.[27]

Again, we use OLS regression to analyse the correlation between labour market status and changes in the income ranks (see Table 4.6). Prior to the crisis (column 1), we see that the changes in the z-scores are not significantly different across labour market states (except for the self-employed), but during the crisis years (column 2) every group, except the employed and unemployed, on average, moves up in the income ranking. Interestingly, the self-employed, the group with the greatest drop in the income ranks between 2004 and 2007, has gained the most after 2009. The retired and disabled groups also experienced gains, both of which receive "sticky" benefits that were not immediately affected by current economic conditions.

Columns 3 and 4 repeat the exercise for employed individuals only to examine differences between occupations for those employed in the first wave of each rotation group, (in 2004 and 2009). We do not observe large differences in re-rankings across occupations in the pre-crisis years (column 3), but during the crisis years (column 4), all occupation groups fell relative to the Manager group. The largest significant drop occurs in the Elementary Occupation group, which contains manufacturing, mining and construction labourers. These findings are in line with previous evidence showing that it was those in the construction sector whose incomes fell the most after the onset of the crisis in Spain Bentolila et al. (2012).

4.6 Conclusion

We examine the evolution of IRHI in Spain both before and during the Great Recession, and decompose IRHI changes into four separately interpretable terms, reflecting the contribution of (i) income

[26]For a comprehensive overview of recent reforms of Spanish old-age and disability pensions see García-Gómez et al. (2012).

[27]Neither our model of health nor our decomposition accounts for individuals' labour market status. We have repeated the decomposition with the inclusion of labour market status and the results are extremely similar to those presented here. This is because once age is controlled for labour market status has very little correlation with health, and consequently can explain only very little.

Table 4.6: Labour Market Status/Occupational Re-ranking

Variable	**Re-ranking Coefficient**^{a,b}			
	2004-2007 (1)	**2009-2012 (2)**	**2004-2007 (3)**	**2009-2012 (4)**
Labour Market Status				
Employed	0.00472	-0.0142*		
Self Employed	-0.0324***	0.0447***		
Unemployed	0.0109	-0.0675***		
Other	0.00397	0.0152*		
Retired	-0.00679	0.0287***		
Disabled	0.000519	0.0287		
Occupation				
Managers			-0.0222	0.0658***
Military			0.0553	0.0136
Professionals			0.00953	0.0155
Technicians			-0.0197	0.025
Clerks			-0.00476	-0.0314
Service & Sales			0.0263**	0.00399
Agricultural			-0.0217	-0.0513
Trade			-0.012	-0.0245
Machine Operators			-0.00249	-0.0286
Elementary Occupation			-0.00456	-0.0452**

Notes:

1. ^{a}Re-ranking coefficient refers to a no-constant regression where change in rank is regressed on economic status/occupation in the first period.
2. ^{b}The coefficients are jointly significant for the all models, except for the 2004-2007 occupation regression, where they are only jointly significant at the 0.10 level.
3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

growth, (ii) income inequality changes, (iii) income mobility and (iv) changes in non-income terms. Our findings are as follows.

First, while our approach is only informative for health changes resulting from changes in the explanatory variables, our findings suggest that health inequality by income in Spain was rising in the four years of economic growth prior to the start of the crisis, but this rise was modest. By contrast, after 2008, it started falling at a faster pace. Second, there appear to be two reasons for this modest rise in IRHI prior to 2008 – income growth and to a lesser extent income mobility – suggesting that the health benefits associated with income growth were disproportionately concentrated amongst the already rich; and that the elderly, often in poorer health, fell slightly on the income ranks leading to increased disparities. Third, the falling health disparities by income mainly derived from the uneven distribution of the income consequences of the crisis. The incomes of younger, healthier groups were affected much more by rising unemployment than the incomes of the over 65s which mainly consist of pensions. Since contributory pensions are "sticky" in Spain and therefore relatively unaffected in the first years of the crisis, pensioners improved their relative position in the income distribution substantially. Fourth, we study the role of labour market participation status and occupation and find that, in line with others studies, it was primarily the income deterioration of the unemployed and the employed, especially those in the construction sector, that was responsible for their fall in the income ranking.

While the great recession caused a substantial deterioration in income, health policy makers can perhaps take solace in the fact that the Spanish pension system - at least in the short run - appears to have shielded some of the most vulnerable individuals. The EU has devoted special attention to reducing health inequalities and for decades countries have attempted to reduce pervasive and persistent health disparities in periods of economic growth. Ironically, our study reveals that the recent crisis has perhaps done more to cut back inequality than many years of pro-poor health policy making. This may be somewhat surprising, given the initial predictions of many observers and in light of media reports of crises hitting the most vulnerable population segments first. But in reality it can be understood as a logical outcome in the presence of sticky pensions and other welfare benefits in the immediate aftermath of a financial crisis. While employment rates and earnings levels are less protected in the short run, also pension and other benefits may be curtailed in the longer run as a consequence of fiscal constraints. It also remains to be seen whether the post-crisis evolution of income-related health inequality has been similar in other European countries with less sticky pension and other benefits.

4.A Appendix

Table A.4.1: IRHI Change Within Rotation Groups

Rotation Group	**IRHI Change**		
1	2004-2005	2004-2006	2004-2007
	0.0019	0.003	0.0047*
2	2005-2006	2005-2007	2005-2008
	0.0031*	0.00071	0.0041*
3	2006-2007	2006-2008	2006-2009
	0.0019*	0.0016	0.0027*
4	2007-2008	2007-2009	2007-2010
	0.00447*	0.0037*	-0.00001
5	2008-2009	2008-2010	2008-2011
	0.00193	-0.00273	-0.00478*
6	2009-2010	2009-2011	2009-2012
	-0.003*	-0.0038*	-0.0064*

Notes:
1. Table shows changes in IRHI compared to the base year for each rotation group.
2. * $p < 0.05$

Table A.4.2: Decomposition Results for Rotation Group 1

	2004-2005		2004-2006		2004-2007	
IRHI Change	0.00196		0.00339		0.00483	
Income Growth	0.00011		0.00065		0.00084	
Income Inequality	-0.00016		0.00003		-0.00008	
Income Mobility	0.0014		0.00216		0.00354	
Non-income Factors	0.00061		0.00073		0.00094	
			Individual Contribution			
	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors
M 16-26	-0.00001	0	-0.00002	0	0.00001	0
F 26-36	-0.00022	0.00018	-0.00012	0.00012	-0.00004	0.00025
M 26-36	0.00002	0.00016	-0.00026	0.00036	-0.00016	0.00047
F 36-46	-0.00006	0.00008	0.00011	-0.00014	-0.00051	-0.00007
M 36-46	-0.00008	-0.00028	-0.00025	-0.00063	0.00015	-0.00076
F 46-56	0.00076	-0.00065	0.00008	-0.00069	0.00013	-0.00102
M 46-56	-0.00032	0.00013	0.00015	0.00034	-0.00019	0.00026
F 56-66	0.00014	-0.00064	0.00027	-0.00097	0.00075	-0.00158
M 56-66	0.00062	-0.00068	-0.00033	-0.00145	0.00097	-0.00177
F 66-76	-0.00022	0.00021	0.00102	-0.00032	0.00038	-0.00113
M 66-76	0.00045	-0.00015	0.00094	-0.00043	0.00097	-0.00083
F 75+	0.00012	0.00165	-0.00027	0.00313	0.00038	0.00495
M 75+	0.00023	0.0006	0.00049	0.00143	0.00052	0.00215
Galicia	-0.00012	0	0.00043	0	0.00028	0.00004
Asturias	-0.00001	0	-0.00001	0	-0.00002	0
Cantabria	0.00002	0	0.00002	0	0.00002	0
País Vasco	0	0	0	0	0	0.00001
Navarra	0.00001	0	0.00001	0	0	0
La Rioja	0	0	0	0	0	0
Aragón	0	0	0	0	0	0
Castilla y León	-0.00002	0	-0.00001	0	-0.00002	0
Castilla-La Mancha	0.00001	0	0	0	0.00001	0
Extremadura	0	0	-0.00002	0	-0.00002	0
Cataluña	0	0	0	0	-0.00001	0
Comunidad Valenciana	0.00001	-0.00001	0.00002	-0.00001	-0.00012	-0.00001
Baleares	-0.00003	-0.00001	-0.00005	0	-0.00002	0.00001
Andalucía	-0.00021	0.00001	-0.0002	-0.00001	-0.00021	0
Murcia	0.00015	0	-0.00002	-0.00001	0.0001	-0.00001
Ceuta y Melilla	-0.00002	0	-0.00003	0	-0.00002	0
Canarias	0.00018	0	0.00023	-0.00002	0.00021	-0.00002

Notes:
1. Table shows the full decomposition for rotation group 1, and the individual contribution of each variable for the Income Mobility and Non-Income Factors terms.
2. Madrid, F 16-26 and Employed used as control groups.

Table A.4.3: Decomposition Results for Rotation Group 4

	2004-2005		**2004-2006**		**2004-2007**	
IRHI Change	0.00487		0.00413		0.00078	
Income Growth	0.00078		0.00135		0.00084	
Income Inequality	-0.00017		0.00081		0.00205	
Income Mobility	0.00374		0.00215		-0.00217	
Non-income Factors	0.00018		0.00006		0.00007	
			Individual Contribution			
	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors
M 16-26	0	-0.00002	0.00001	-0.00001	0.00001	-0.00002
F 26-36	-0.00015	-0.0001	0.00002	0	0.00012	0.00002
M 26-36	-0.00045	-0.00003	-0.00029	0.00002	0	0.00011
F 36-46	-0.0003	0.00005	0.00007	0	0.00018	-0.00015
M 36-46	0.00027	-0.00015	0.00031	-0.00028	0.00023	-0.00054
F 46-56	0.00086	-0.00018	-0.00013	-0.00019	0.00017	0.00008
M 46-56	0.00026	-0.0002	0.00032	0.00011	0.0004	-0.00006
F 56-66	0.00068	0.00027	0.00048	-0.00022	0.00091	-0.00094
M 56-66	0.00053	-0.00016	0.00037	-0.0003	0.00012	-0.00065
F 66-76	0.0008	-0.00097	0.00151	-0.00137	-0.00026	-0.00103
M 66-76	0.00058	-0.00032	0.00077	-0.00129	-0.00012	-0.00124
F 75+	0.00006	0.00134	-0.00121	0.002	-0.00243	0.00257
M 75+	0.0005	0.00066	-0.00029	0.0016	-0.00116	0.00196
Galicia	0.00013	-0.00003	0.00012	0.00001	-0.00044	-0.00002
Asturias	0.00001	0	0.00002	-0.00001	-0.00001	0
Cantabria	0	0	-0.00001	0	-0.00002	0
País Vasco	0.00003	0	0.00005	0	0.00004	0
Navarra	0	0	0	0	0	0
La Rioja	0	0	0	0	0	0
Aragón	0	0	-0.00001	0	-0.00001	0
Castilla y León	0.00001	0	-0.00001	0	-0.00002	0
Castilla-La Mancha	0	0	0.00001	0	0	0
Extremadura	0.00001	0	0.00002	0	0.00001	0
Cataluña	0.00001	0	0	0	-0.00001	0
Comunidad Valenciana	0.00009	0	0.00015	-0.00001	0.00019	-0.00001
Baleares	0.00002	0	0.00004	0	0.00002	0
Andalucía	-0.00028	0.00001	-0.00033	0.00001	-0.0003	0.00001
Murcia	0.00002	0.00001	0.00008	0.00001	0.00008	0.00001
Ceuta y Melilla	0.00001	0	0.00001	0	0	0
Canarias	0.00004	0	0.00006	-0.00003	0.00012	-0.00004

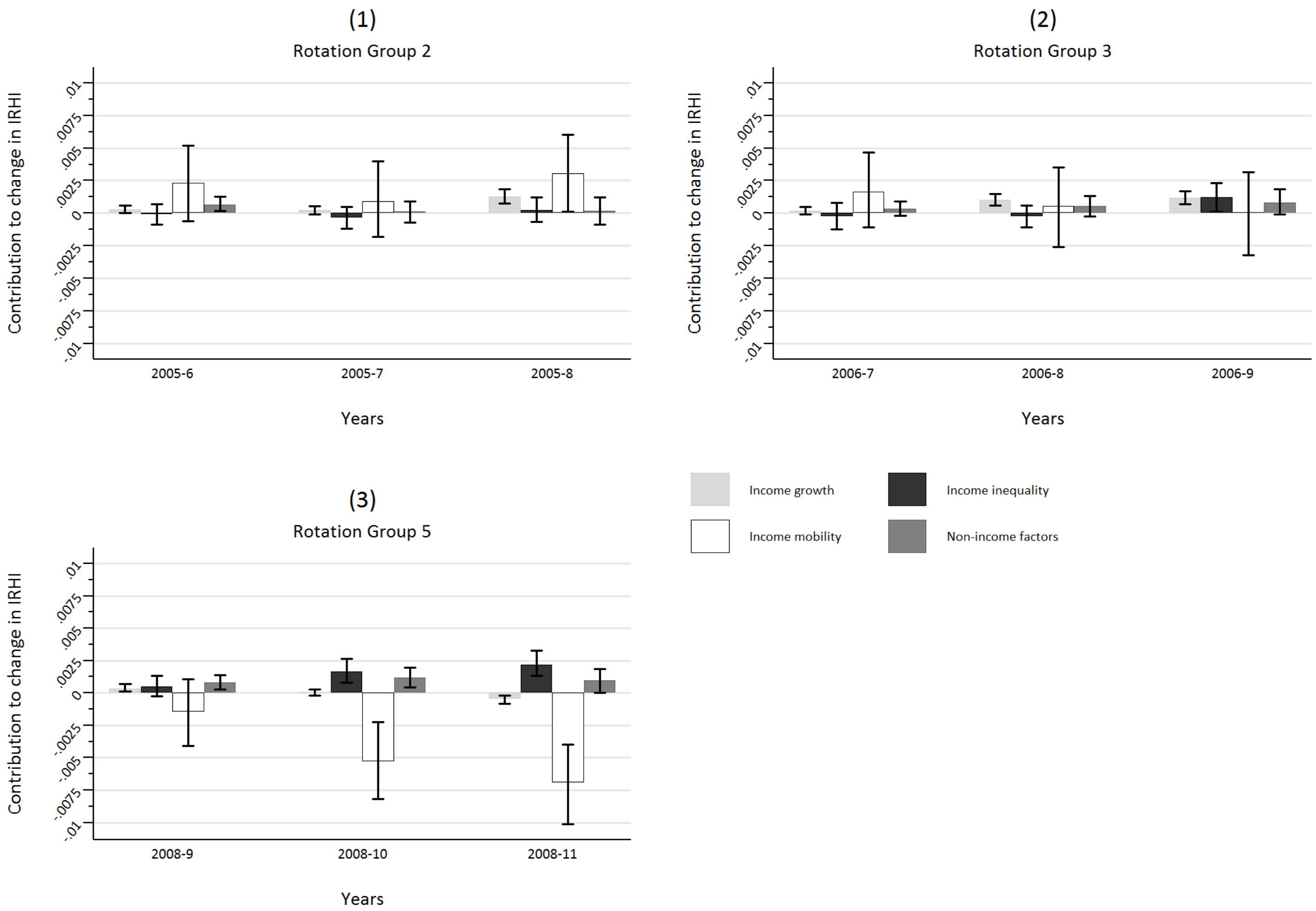
Notes:
1. Table shows the full decomposition for rotation group 4, and the individual contribution of each variable for the Income Mobility and Non-Income Factors terms.
2. Madrid, F 16-26 and Employed used as control groups.

Table A.4.4: Decomposition Results for Rotation Group 6

	2004-2005		**2004-2006**		**2004-2007**	
IRHI Change	-0.00244		-0.00367		-0.00643	
Income Growth	-0.00043		-0.00119		-0.00182	
Income Inequality	0.00097		0.00103		0.00153	
Income Mobility	-0.0033		-0.00385		-0.00611	
Non-income Factors	0.00032		0.00059		0.00029	
			Individual Contribution			
	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors	Income Mobility	Non-inc. Factors
M 16-26	0.00001	0	0	0	0	0
F 26-36	-0.00006	0.00015	0.00007	0.00032	0.00018	0.00051
M 26-36	0.00016	0.00007	0.00009	0.00013	0.00011	0.00021
F 36-46	0.00013	-0.00006	0.00007	-0.00011	0.00023	-0.00042
M 36-46	0.00038	-0.00003	-0.00002	0.00003	0.00042	0.00004
F 46-56	0.00033	-0.00029	0.00031	-0.00002	0.00031	-0.00015
M 46-56	0.00025	-0.00018	0.00029	-0.00032	0.0004	-0.00046
F 56-66	-0.00102	-0.00037	-0.00044	-0.00173	-0.00012	-0.00214
M 56-66	-0.00033	0.00017	0.00083	-0.00032	0.00022	-0.00073
F 66-76	-0.00015	-0.00026	-0.00142	-0.00051	-0.00191	-0.00076
M 66-76	-0.00013	-0.00095	-0.00043	-0.0014	-0.00118	-0.00166
F 75+	-0.00215	0.00121	-0.00223	0.00284	-0.00321	0.0037
M 75+	-0.00073	0.00088	-0.00099	0.00169	-0.00176	0.00217
Galicia	-0.00042	0	-0.00042	-0.00001	-0.00009	-0.00001
Asturias	0	0	-0.00001	0	-0.00002	0
Cantabria	0	0	0.00001	0	0	0
País Vasco	0.00001	0	0.00003	0	0.00002	0
Navarra	0	0	0	0	0	0
La Rioja	0	0	0	0	0	0
Aragón	-0.00001	0	-0.00001	0	-0.00001	0
Castilla y León	0.00002	0	0.00001	0	-0.00001	0
Castilla-La Mancha	0	0	0.00001	0	0	0
Extremadura	0	0	0.00001	0	0.00001	0
Cataluña	0	0	0.00001	0	0.00001	0
Comunidad Valenciana	0.00015	-0.00001	0.00012	-0.00002	0.00007	-0.00003
Baleares	0.00003	0	0.00002	-0.00001	0.00001	-0.00001
Andalucía	0.00011	0	0.00021	0.00001	0.00015	0.00002
Murcia	0.00012	0	0.00013	0	0.0001	0
Ceuta y Melilla	0.00003	0	0	0	0.00001	0
Canarias	-0.00006	0	-0.00009	0.00002	-0.00007	0.00002

Notes:
1. Table shows the full decomposition for rotation group 6, and the individual contribution of each variable for the Income Mobility and Non-Income Factors terms.
2. Madrid, F 16-26 and Employed used as control groups.

Figure A.4.1: Unemployment and GDP growth in Spain



Notes:

1. Figures shows the decomposition terms for rotation groups 2, 3, and 5.

Chapter 5

Every Crisis Has a Silver Lining? Unravelling the Pro-Cyclical Pattern of Health Inequalities by Income

Joint work with Pilar García-Gómez, Eddy Van Doorslaer and Tom van Ourti

5.1 Introduction

It is well known that those with higher incomes enjoy longer and healthier lives than those with lower incomes. These inequalities, which are widespread and persistent, have presented a challenge to policy makers and researchers. Both the Centre for Disease Control in the US and the European Commission have highlighted the need to reduce disparities in health, and have devoted resources to doing so (Frieden, 2013; European Commission, 2009).

However, despite these concerns, important gaps remain in our understanding of these inequalities. Firstly, relatively little is known about how Income Related Health Inequality (IRHI) in Europe has changed since the Great Recession. While changes in the income distribution have been well documented (Jenkins et al., 2012), comprehensive cross-country evidence on changes in the distribution of health by income before, during and after the crisis is lacking.[1] Without precise estimates of IRHI over this period, we currently miss important information that is necessary to address these inequalities.

Secondly, evidence is also lacking on the relative importance of (changes in) different income sources for (changes in) IRHI. We distinguish between the two most important sources of income:

[1] Ásgeirsdóttir and Ragnarsdóttir (2013) study differences in IRHI for 26 European countries in 2007. However, this cross-sectional approach is uninformative about the evolution of IRHI between 2004 and 2013.

market incomes (like wages), and government transfers (like old-age and unemployment benefits) and separate their influence on IRHI. Why is it plausible that changes in these different income sources have differing IRHI consequences? First, because their distribution across age and health groups differs, and secondly, because they tend to vary in opposite directions in times of recession and growth. The distinction is also important because of its implications for policy; governments are able to manipulate transfers such as unemployment benefits more directly than, for example, wages. The crisis induced heterogeneous labour market effects across nearly all European countries and governments responded differently with a range of austerity measures, primarily relating to unemployment and pension benefits. Further, if there is a distinct role of transfer income for IRHI changes, then it is important to shed light on some of the – perhaps unintended – IRHI consequences that policies governing these transfers may have had.

Our contributions are fourfold. First, we document trends in IRHI in 7 European countries between 2004 and 2013 – both before and after the financial crisis.[2] Second, we develop a novel decomposition method that identifies the separate roles of government transfers and market earnings on the evolution of IRHI. Third, by means of the decomposition, we unravel the most important drivers of the distinctive patterns that we observe for IRHI pre- and post-crisis. Lastly, we provide descriptive evidence on the role that specific pension policies, and the austerity measures enacted in Greece, have had on IRHI.

We add to the literature using rank-dependent, concentration index-type measures to compare health inequalities by income across countries that started with Van Doorslaer et al. (1997). Subsequent contributions have employed a series of decomposition methods and measurement corrections that provided additional insight into the drivers of cross-country differences (Van Ourti et al., 2009,?). These and other European comparative studies report substantial pro-rich inequalities in health in Europe, and they highlight the role of changes in the income ranks, in addition to health, for inequality trends.

Coveney et al. (2016) used a decomposition of concentration indices to study IRHI changes in Spain between 2004 and 2012. Though IRHI was initially growing between 2004 and 2008 as the Spanish economy grew, large reductions in inequality occurred after 2008.[3] A decomposition analysis of these trends reveals that IRHI was primarily driven by the income position of the relatively unhealthy elderly groups. In "good" economic times, the income position of the young tended to rise faster than that of the elderly, increasing the income gap between the healthy and the unhealthy and subsequently leading to increases in IRHI. During bad economic conditions, incomes of the young

[2] Austria, Belgium, France, Greece, Italy, Portugal and Spain.

[3] This is in line with Regidor et al. (2014) who – using a different methodology – conclude that all-cause mortality declined more rapidly during the economic crisis among groups with low socioeconomic status.

fell while incomes among the elderly tended to be far more stable, leading to decreases in IRHI. While these findings hint at the distinct roles played by government transfers versus income from labour, the decomposition methods used did not explicitly allow for this distinction. Further, studying a range of European countries, with differing levels of exposure to the crisis as well as a range of different pension and other policies, provides further insights into the determinants of trends in IRHI.

We do not aim to add to the literature that started with Ruhm (2000), linking health and economic conditions, aiming to identify a causal effect of the crisis or income on health. Rather, our decomposition illustrates how changes in transfer and market incomes are related to changes in the association between income and health, and thus IRHI. By following cohorts of individuals over time in relation to the underlying income and health distributions, our approach also differs from the cross-country comparisons of Mackenbach and co-authors (Mackenbach et al., 1997, 2008), which document levels and trends in socio-economic inequalities in health (mostly education- and occupation-related) for a large number of European countries.

Our findings are as follows. First, we find that IRHI trends are interwoven with macroeconomic conditions. Documenting annual IRHI changes across 7 European countries between 2004 and 2013, we find differential trends that imply IRHI is pro-cyclical: perhaps surprisingly, inequalities tend to increase in good economic times and fall in bad times. Between 2004 and 2008, a time of relatively steady economic growth in Europe, IRHI was on average relatively flat, though it significantly increased in countries with substantial economic growth, such as Greece and Spain. Between 2008 and 2012, IRHI fell in countries that were most affected by the crisis, namely Greece, Spain and Portugal. IRHI in countries that did not experience severe economic consequences as a result of the crisis, such as France and Austria, remained stable or increased slightly.

Second, by decomposing these changes, we find that in general the two main sources of household income – market income and government transfers – have opposite effects on IRHI. Market income growth is associated with increasing health inequalities, while rising government transfers tend to reduce them. This stems from the fact that market incomes are on average afforded to the healthy, while government transfers, especially pensions, are on average afforded to both the unhealthiest and poorest individuals in the population.

Third, related to the first and second finding, we show that the pro-cyclical pattern of IRHI can largely be explained by the differing importance of government transfers and market incomes in good economic versus bad economic conditions. The economic crisis led to differential income changes by age-group, and thus by health status. Thus, if income from work grows more (less) than pensions during good (bad) economic times, IRHI grows (falls), in particular when the relative income position of the (unhealthy) very elderly is affected.

Lastly, we present descriptive evidence that both household structure and policies governing pensions across countries have a measurable impact on IRHI. Households where intergeneration sharing of pensions is high, as well as pension reforms Greece enacted as part of the austerity measures in 2010 and 2011, both appear to have dampened the IRHI reducing effect of government transfers. Further, given the importance of the income position of the very elderly in determining IRHI trends, we conclude that policies governing the generosity of pensions for this group, such as indexation policies, can play a role in governing these trends.

5.2 Decomposition of Changes in Income-Related Health Inequality

Our decomposition method is based on an extension of the method used in Baeten et al. (2013). In this section we describe the approach for a balanced cohort of n individuals that we observe at the start (period 1) and end (period 2) of a time interval.

5.2.1 Health Inequality Measurement.

To measure health inequalities we use the corrected concentration index (CCI) (Erreygers, 2009) which satisfies the mirror condition and is insensitive to equal health additions (absolute inequality) (Erreygers and Van Ourti, 2011). When health is bounded between 0 and 1, the index can be written as:

$$CCI(h_t|y_t) = \frac{8}{n^2}\sum_{i=1}^{n} z_{it}h_{it} \tag{5.1}$$

where h_t and y_t are the health and income distribution in period $t = 1$ or 2, h_{it} describes the health level of individual i and z_{it} is a weight that depends linearly on the income rank of individual i with individuals ranked from poor $(i = 1)$ to rich $(i = n)$, *i.e.* $z_i = ((2i - n - 1))/n$. This income weight takes the value 0 for the individual with median income, and increases linearly with income rank.

5.2.2 Health Model.

We use a simple descriptive model that links health linearly and additively to its associated factors:

$$h_{it} = \alpha + \theta(y_{it}) + x_{it}^{'}\beta \tag{5.2}$$

where α is an intercept parameter; $\theta(y_{it})$ is a non-linear function of income; x_{it} represents a vector of K non-income variables (in our analysis, these are a set of age-sex and region dummies), and β is its associated parameter vector reflecting partial associations. The exact functional form for $\theta(.)$ pre-determines the sign and magnitude of some parts of our decomposition. Therefore we use a flexible functional form in the empirical application.

5.2.3 Decomposition of IRHI Change.

Our interest lies in decomposing changes in IRHI. Taking the change in the CCI between two periods and combining equation Equation (5.1) and Equation (5.2) leads to:

$$
\begin{aligned}
&CCI(h_2|y_2) - CCI(h_1|y_1) \\
&= \frac{8}{n^2}\Big[\sum_{i=1}^{n} z_{i2}h_{i2} - \sum_{i=1}^{n} z_{i1}h_{i1}\Big] \\
&= \frac{8}{n^2}\sum_{i=1}^{n}\Big\{\Big[z_{i2}\theta(y_{i2}) - z_{i1}\theta(y_{i1})\Big] + \beta\Big[z_{i2}x_{i2}^{'} - z_{i1}x_{i1}^{'}\Big]\Big\}
\end{aligned}
\tag{5.3}
$$

Equation (5.3) shows that we can disentangle the change in IRHI into a part due to changes in the association between the income rank and the non-linear income function $(z_{i2}\theta(y_{i2}) - z_{i1}\theta(y_{i1}))$ and a part due to changes in the association between the income rank and the non-income factors $(z_{i2}x_{i2}^{'} - z_{i1}x_{i1}^{'})$.[4]

Because the aim is to separate the role of different income sources for the change in IRHI, we distinguish between total income (y_{it}) as the sum of market incomes (y_{it}^{M}) and government transfers (y_{it}^{T}), i.e. $y_{it} = y_{it}^{M} + y_{it}^{T}$. Income weights can then be defined separately for each source. Weights associated with total income (z_{it}) and market income (z_{it}^{M}) are defined in the standard way described above. The income weights associated with transfers are defined as the difference between an individual's total income rank and market income rank:

$$z_{it}^{T} = z_{it} - z_{it}^{M} \tag{5.4}$$

An individual's transfer income rank thus not necessarily coincide with the rank of y_{it}^{T}, but measures the number of steps on the income ladder that separate total from market income. In our descriptive setting this coincides – as is common in the income redistribution literature (Plotnick, 1981; Lambert, 2001) – with interpreting market income as the income that would prevail in the absence of government transfers, or in other words with the redistributive effect of government transfers.

[4] An additional assumption is that there is no structural change in the health equation across periods.

Combining our model for health (Equation (5.2)), our definition of transfer income weights (Equation (5.4)), and after manipulating the terms in the final line of Equation (5.3), the change in IRHI between periods 1 and 2 can be expressed as the sum of 5 terms:

$$
\begin{aligned}
&CCI(h_2|y_2) - CCI(h_1|y_1) \\
&= \frac{8}{n^2} \sum_{i=1}^{n} \Bigg\{ \underbrace{(z_{i2}^M - z_{i1}^M) \sum_{j=1}^{k} x_{ji2}\beta_j}_{\text{market-related income mobility}} + \underbrace{(z_{i2}^T - z_{i1}^T) \sum_{j=1}^{k} x_{ji2}\beta_j}_{\text{transfer-related income mobility}} \\
&\quad + \underbrace{z_{i2}^M \theta(y_{i2}^M) - z_{i1}^M \theta(y_{i1}^M)}_{\text{market-related inequality } \Delta} + \underbrace{\left[z_{i2}\theta(y_{i2}) - z_{i2}^M \theta(y_{i2}^M)\right] - \left[z_{i1}\theta(y_{i1}) - z_{i1}^M \theta(y_{i1}^M)\right]}_{\text{transfer-related inequality } \Delta} \\
&\quad + \underbrace{z_{i1} \sum_{j=1}^{k} \beta_j (x_{ji2} - x_{ji1})}_{\text{ageing and migration}} \Bigg\}
\end{aligned}
\tag{5.5}
$$

5.2.4 Explanation of Decomposition Terms.

We term the first two expressions in Equation (5.5) *market-related income mobility* and *transfer-related income mobility* respectively. *Market-related income mobility* measures the association between changes in the market income weights/ranks and non-income related health in the second period. The expression between brackets captures the change in an individual's market income weights/ranks between period 1 and 2, and will be positive (negative) if an individual has moved up (down) in the market income ranks. The second part of the term captures the non-income related health of the individual in the second period. The *transfer-related income mobility* term is identical, except for the use of transfer income weights. Both income mobility terms are more positive (negative) when upwardly (transfer/market) income mobile individuals have better (worse) non-income health in period 2, or vice versa.

Note that if the non-income variables consist of multiple variables that enter the health equation additively, then the mobility terms comprise a summation of different sub-terms. This holds, for example, if one uses a set of age-sex and region dummies as we do. This allows one to separate the aggregate mobility effect into the contribution per age-group and region category. Summing the total transfer and market mobility terms gives the total income mobility.

The third expression in Equation (5.5) is termed *market-related inequality change*. It measures the consequences for IRHI of the change in the distribution of market incomes. $\theta(y_{i2}^M)$ denotes the health level in the second period that corresponds to y_{i2}^M conditional on the non-income factors. The first product therefore measures market related inequality in the conditional health levels. This is simply the CCI for market income related health in the second period. The second product in the expression is identical, but refers to the first period. The difference between these two corrected concentration indices therefore captures how changes in the distribution of market incomes between the two periods were associated to changes in IRHI, both by their association with health through the $\theta(.)$ function, and via the re-ranking of individuals on the market income scale. For a monotonically increasing $\theta(.)$ function, market-related inequality change will indicate rising (falling) IRHI when the rich (poor) predominantly experience income improvements (deteriorations).

The fourth expression in Equation (5.5) is the *transfer-related inequality change*. Term $[z_{i2}\theta(y_{i2})-[z_{i2}^{M}\theta(y_{i2}^{M})]$ captures the degree to which transfer incomes change the association between income weight/rank and income-related health in the second period; the second term measures this effect in the first period. Both terms thus reflect whether transfer incomes result in a more or less equal distribution of income-related health, or the extent of the redistributive effect of transfer incomes in the separate periods. Their difference is a measure of how this effect has changed over time, and its consequence for the evolution of IRHI. Summing *market-related inequality change* and *transfer-related inequality change* gives the change in the CCI for total income-related health between periods 1 and 2.

Finally, any remaining change in IRHI is captured by the *ageing and migration term*. It indicates how changes in non-income related health, due to their association with initial income weights/ranks, have led to changes in IRHI. As our non-income variables are age-sex and region dummies, it therefore measures the impact of ageing and within-country migration on IRHI. As these phenomena have consequences for health, the degree to which they are associated with income ranks may affect IRHI. This term mainly acts as a control, allowing us to study changes in IRHI net of ageing and migration effects.

5.3 Empirical Analysis

5.3.1 Data.

We use the European Union Survey on Income and Living conditions (EU-SILC), a European-wide survey designed primarily to collect labour and income related data. It is well suited to our analysis for several reasons. First, it provides a detailed breakdown of the sources of disposable household income, which is crucial to measuring the separate effects of government transfers and market income on IRHI trends. Secondly, individuals are asked to rate their self-assessed health (SAH), which is used as our health measure.

Our selection of countries is based on data availability and quality in the EU-SILC. We require that countries have adequate income and health data for the whole 2004-2013 period. Appendix Table A.5.1 provides information on the available information for the 29 EU-SILC countries and the selection criteria used for inclusion. This leaves us with the following 7 countries: Austria, Belgium, France, Portugal, Italy, Greece and Spain.[5] Crucially, the latter 4 countries are of particular interest because they were more affected by the 2008 financial crisis.

The EU-SILC is a rotating panel. A new random sample (referred to as a rotation group) is drawn every year, followed for 4 years and then dropped. Therefore, at any point, each country has 4 concurrent panel samples. There are 7 rotation groups in our study period, i.e. $2004-2007, \ldots, 2010-2013$.[6] We use balanced data from all 7 rotation groups to estimate our model for health (Equation (5.2)).[7][8] Table 5.1 gives the number of observations per rotation group and country. Due to changes in data collection methods, the income data for France from 2009 onwards are not comparable to earlier waves. We therefore ignore the 2007-2010 period for France.[9]

Table 5.1: Observations Per Wave, Rotation Group, Country in EU-SILC Dataset

Rotation group	1	2	3	4	5	6	7
Period	2004-2007	2005-2008	2006-2009	2007-2010	2008-2011	2009-2012	2010-2013
Observations							
Austria	2,294	1,923	1,901	1,894	1,882	2,200	2,163
Belgium	1,315	2,126	1,966	1,987	1,670	1,762	1,886
Greece	2,221	2,113	2,498	2,238	2,756	2,479	2,244
Spain	4,136	4,918	5,046	5,474	5,522	5,177	4,571
France	1,433	2,323	2,360	2,359	2,399	2,295	2,316
Italy	8,240	7,898	7,673	7,709	6,730	5,652	4,986
Portugal	1,945	1,758	1,770	2,014	2,101	2,612	2,571

Notes:
1. Table shows for each rotation group the period spanned and the number of balanced observations (observed for the whole 4 year period) for each country.

[5] Our selection criteria is that a country is represented in all 7 rotation groups. Furthermore, although many of the Nordic countries – Finland, Iceland and Sweden – are present in rotation groups, their use of register-based data collection methods leads to many missing values of the SAH variable raising concerns of attrition bias. Sample sizes in some of these countries are too low for reliable analysis. For instance, there are only 13 women above the age of 75 in the 2004 sample in Iceland. Appendix Table A.5.1 details the selection criteria per country.

[6] The French EU-SILC uses longer rotation groups, but for comparability with other countries we shorten them to 4 years.

[7] We symmetrically drop the top and bottom 1% of total income to remove potential outliers.

[8] Our restriction to balanced panels excludes the possibility of attrition bias. However, trends of IRHI computed when using all data - not just a balanced panel – are extremely similar to those we find here, suggesting attrition bias is not driving our results.

[9] The data collection method for certain components of income in France, namely "*interest, dividends and profit from capital investments in unincorporated business*", went from being survey-based to register-based in 2009. The average value of this component increased by almost €3,000, and led to a dramatic rise in average incomes. It is not possible to distinguish between "real" increase in the component and inflation due to more accurate collection methods.

5.3.2 Income and Health Measurement.

The EU-SILC provides, per person and household, a detailed breakdown of the components of annual household income. We separate total income into what we term market income and transfer income. An individual's market income is defined as the equivalized value of disposable household income before all social transfers, and transfer income as the equivalized value of the sum of all household social transfers.[10] The income reference period is the previous calendar year. Table 5.2 lists the EU-SILC components that make up household market and transfer income.

Table 5.2: Income Components of Transfer and Market Incomes

Transfer Income	**Market Income**
• Unemployment benefits	• Gross employee cash or near cash income
• Old-age benefits	• Company car
• Survivor benefits	• Gross cash benefits or losses from self-employment
• Sickness benefits	• Pensions received from individual private plans
• Disability benefits	• Income from rental of a property or land
• Education-related allowances	• Regular inter-household cash transfers received
• Family/children related allowances	• Returns from unincorporated business
• Social exclusion not elsewhere classified	• Income received by people aged under 16
• Housing allowances	
	Minus
	• Regular taxes on wealth
	• Regular inter-household cash transfer paid
	• Tax on income and social insurance contributions

Notes:
1. Table shows the makeup for our definitions of Transfer and Market incomes as used in the EU-SILC survey. Importantly, old-age benefits captures all benefits that provide a replacement income once an individual retires or reaches a certain age. This includes public pension payments, care allowances, disability cash benefits, lump sum payments at the time of retirement and other cash benefits. It does not include any payments from private pension plans, which enter the market income definition.

What is the relative importance of each of these components? Public pensions form the largest share of transfer income, and employee income (income from work) for market incomes.[11] When using the term "pensions" we are referring to what EU-SILC terms "old age" benefits. These include the collection of all social payments to the elderly that are designed to provide a replacement income when a person has reached a certain age.[12]

[10]Our equivalization procedure involves dividing household income by the square root of the number of individuals living in the household in the current period.

[11]Appendix Table A.5.2 shows the percentages of transfer and market incomes that are made up of pensions and wages respectively, per rotation group and country.

[12]This includes public pension payments, care allowances, disability cash benefits, lump sum payments at the time of retirement and other cash benefits. It does not include any payments from private pension plans, which enter the market income definition. See the EU-SILC guidelines documentation for further details. Our data shows private pensions are not an important part of transfers for these countries. On average across all rotation groups and countries, payment from private plans are less than 1% of old age benefits. Per country, the average fraction of private payments to pension payments is never higher than 3%.

The other key variable in our analysis is the self-assessed health variable. Individuals are asked the following question: "*How is your health in general? Is it: (1) very good, (2) good, (3) fair, (4) bad, (5) very bad?*"

5.3.3 Implementation of Decomposition.

The first step in the decomposition procedure is to calculate, per country, rotation group and year, IRHI using the CCI. The CCI requires a ratio-scaled health measure (Erreygers and Van Ourti, 2011). In order to transform the ordinal SAH measure in EU-SILC to a ratio-scaled measure, we use an interval regression with the threshold values imposed from external data (Van Doorslaer and Jones, 2003).[13] The variables included in these regressions are age/sex dummies,[14] region dummies,[15] and a second degree income polynomial, in line with the widely observed concave shape of the health-income gradient. This predictive set of variables is parsimonious, yet is strongly associated with health. The interval regression is run separately for each country and serves a dual purpose: (i) they produce a ratio-scaled predicted health score between 0 and 1, and (ii) they provide the non-income and income coefficients (β_j and $\theta(.)$) used in the decomposition (Equation (5.5)).[16] The regression results for each country are shown in Appendix Table A.5.3.

For each country, we then take 3 rotation groups (2004-2007, 2007-2010, 2010-2013), and calculate and decompose the change in the CCI from the first year (the base year). We only present the decomposition with respect to the last year of the rotation group because intermediate decompositions are similar in sign and relative magnitude within rotation groups.[17] In order to allow for statistical inference on IRHI levels, IRHI changes and the decomposition terms, we bootstrap the entire procedure 1,500 times.

5.4 Results and Discussion

This section first examines the general trends in IRHI in the 7 countries under study between 2004 and 2013. We then separately study the role of the *mobility* and *health inequality* terms in IRHI changes before and after the financial crisis in 2008. Next we compare cross country differences in the transfer

[13] Our imposed thresholds are from the empirical distribution function of the health utility index in the Canadian National Population Health Survey 1994-1995.

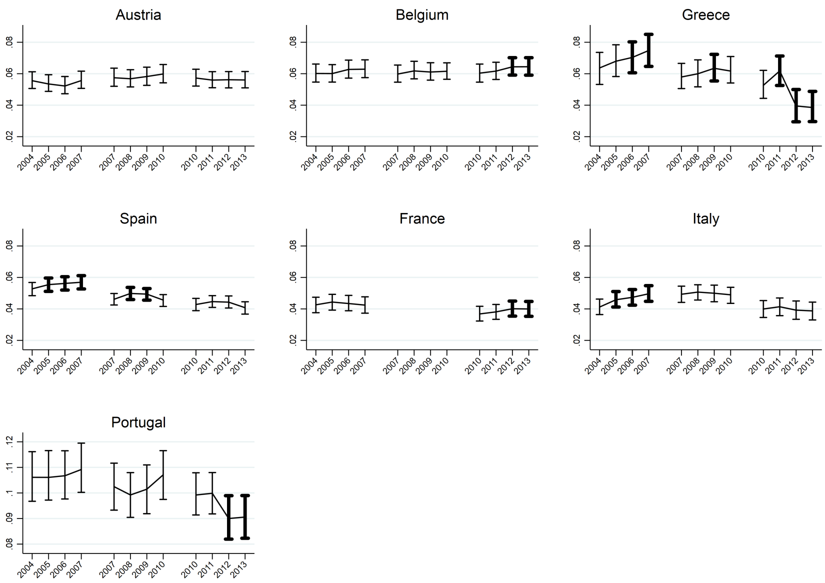
[14] The age/sex dummies divide age into the following categories, separately for men and women: 16 to 25 years, 26 to 35 years, 36 to 45 years, 46 to 55 years, 56 to 65 years, 66 to 75 years, and more than 75 years of age.

[15] Regions in EU-SILC are recorded at the NUTS II level. For Portugal and Belgium however this information is missing and we use urbanization dummies (dense, medium and thinly populated areas) instead.

[16] While we only explicitly decompose the 2004-2007, 2007-2010 and 2010-2013 rotation groups, to maximize sample size the observations for all 7 rotation groups spanning 2004 to 2013 are included in the health regressions.

[17] An exception is Greece in the 2010-2013 decomposition, which we explore in more detail below. The full decomposition results per comparison and per rotation group are available upon request.

Figure 5.1: IRHI Trends



Notes:

1. Figure shows, for each country, IRHI per year, per rotation group. Note the different scale of Portugal. Bold bars years in which change in CCI compared to base year is significant ($p < 0.05$). Y-axis: value of the CCI.

mobility terms and pension policies. Finally, the role of the austerity measures enacted in Greece on IRHI is explored.

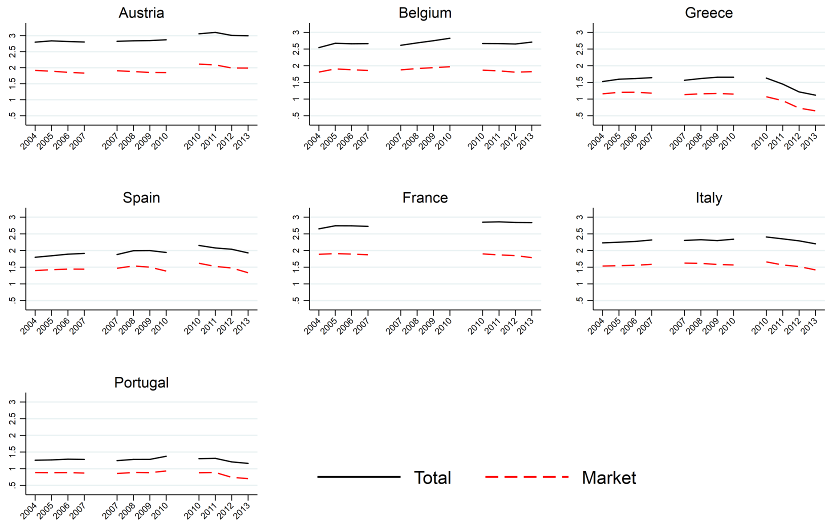
5.4.1 IRHI Trends Across 7 European Countries.

Figure 5.1 shows how IRHI, as measured by the CCI and calculated using predicted health, has evolved between 2004 and 2013 for the 7 countries under study. The separate lines represent the three rotation groups used to span the period. The black bars show 95% confidence intervals. While the confidence intervals in Figure 5.1 are informative about the sampling variability of the yearly point estimates of IRHI, our interest lies in examining the changes of IRHI between different periods. It is therefore useful to know if the changes in IRHI with respect to the base year are statistically significant, which is signified by the bold bars.[18]

We note both geographical and time patterns in the IRHI trends. Before 2008, IRHI was either increasing or showed no significant movements, whereas between 2010 and 2013 some countries experienced dramatic decreases. There seems to be a geographical pattern; IRHI in southern EU countries was initially rising before beginning to fall after approximately 2008. Continental countries

[18] We do not check the statistical significance of changes across rotation groups. While they might be relevant (e.g. comparing the change in IRHI from the onset of the financial crisis), we only observe the same set of individuals over a period of 4 years.

Figure 5.2: Equivalent Household Income



Notes:

1. Figure shows, for each country, average total and market income per year, per rotation group.

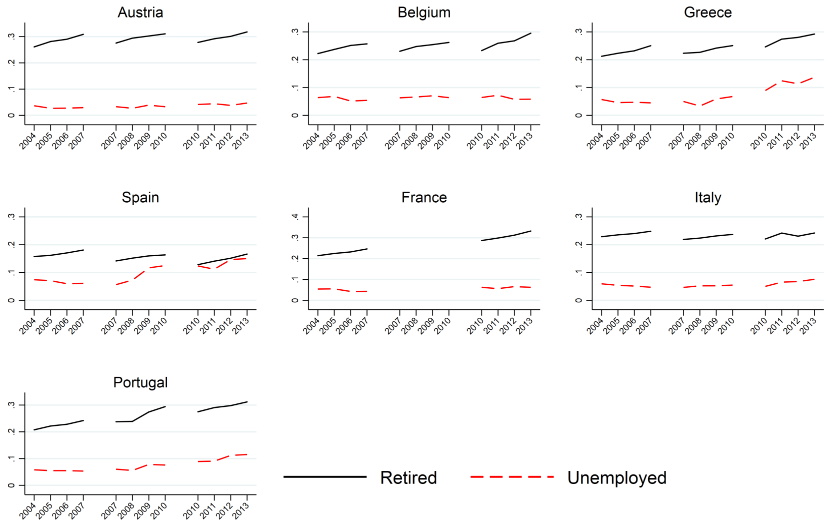
saw much smaller changes in IRHI, and both Belgium and France experienced significant increases in IRHI in the 2010-2013 period.

The above suggests distinctive trends in IRHI before and after the “Great Recession”. Figures 5.2 to 5.4 respectively show, for our sample of analysis, the trends in average total and market equivalized household incomes, unemployment and retirement rates, and the generalized Gini coefficient, calculated using both market and total income. Figure 5.3 reveals an increase in unemployment for Spain, Portugal and Greece between 2009 and 2010, where it continues to rise throughout the 2010-2013 period. Unemployment increases are only noticeable in Italy in the 2010-2013 period, while the continental countries appear unaffected. Patterns in equivalized income are less obvious, and appear mostly in the last rotation group in the southern countries, especially Greece.

Not surprisingly, in periods of economic growth the generalized Gini coefficient of market income tends to increase. The addition of transfers leads to lower absolute income inequality. For countries that suffered noticeable household income declines after the economic crisis, the generalized Gini index decreased.

Given these trends, we distinguish between 3 different periods in our analysis. Following Jenkins et al. (2012), we consider the 2004-2007 period to be the pre-crisis period; a time of relatively normal growth for the 7 countries. We term the rotation group spanning 2007-2010 the crisis period. Finally, the post-crisis rotation period (2010-2013) is when consequences of the crisis are most obvious in

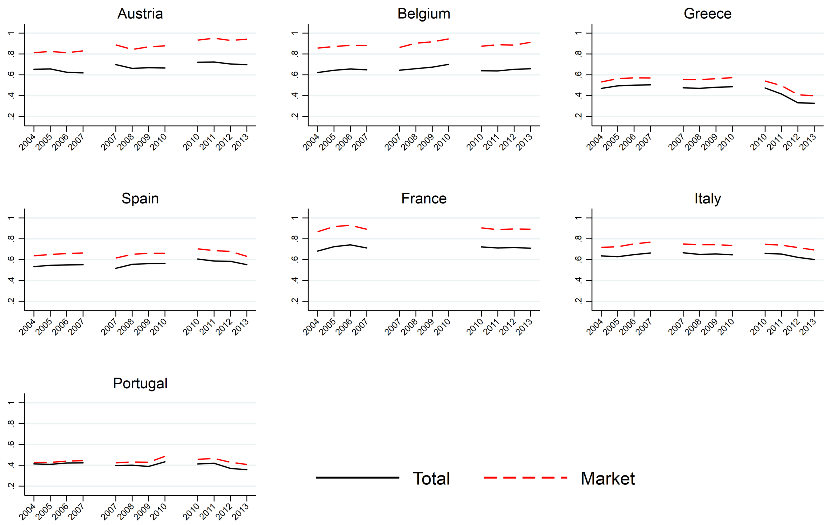
Figure 5.3: Activity Status Trends



Notes:

1. Figure shows, for each country, the average proportion of retired and unemployed individuals per year, per rotation group.

Figure 5.4: Generalized Gini Trends



Notes:

1. Figure shows, for each country, the value of the generalized Gini for total and market income per year, per rotation group.

southern countries, while large effects for household income, inequality and employment are absent for the continental countries.

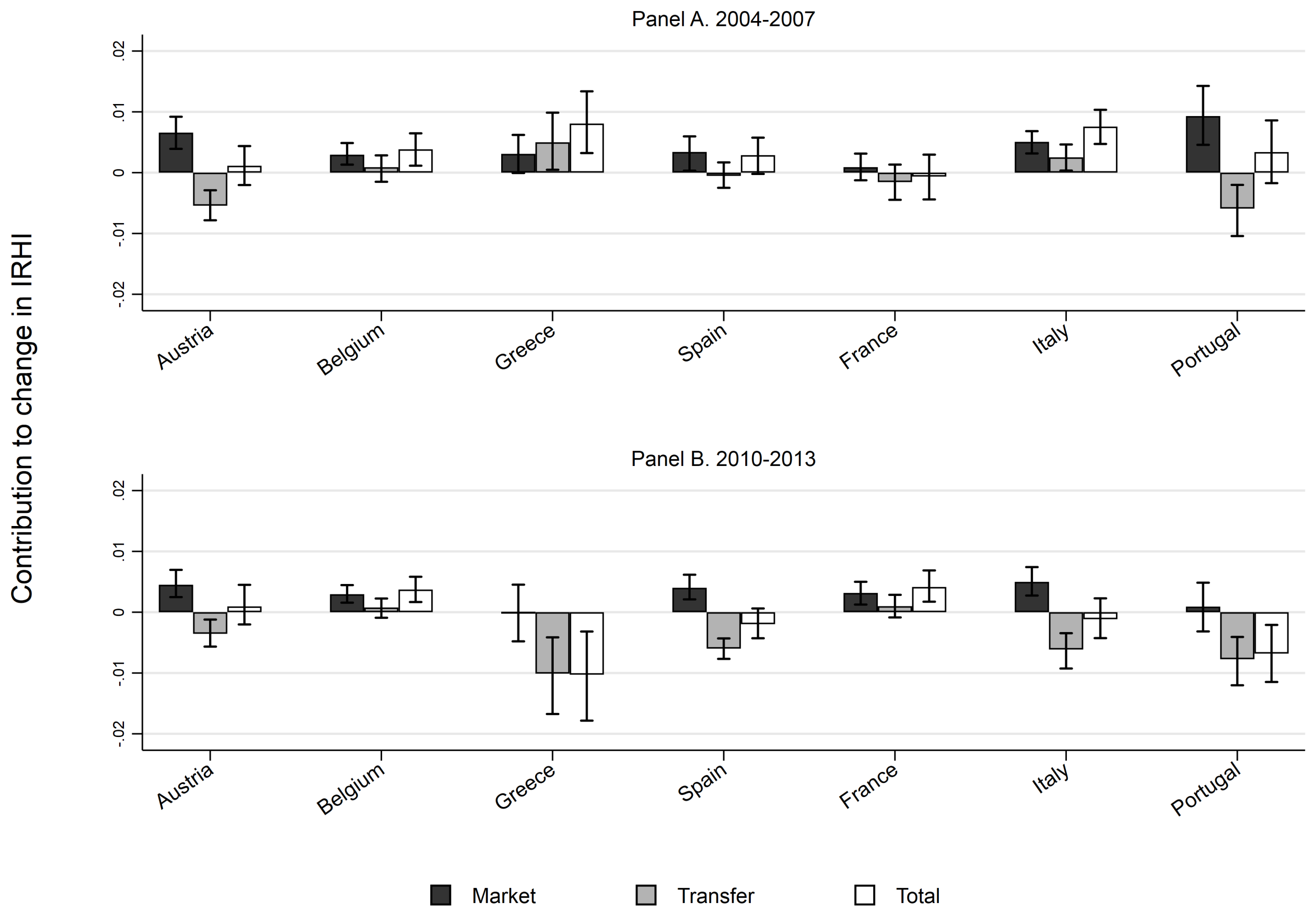
Comparisons of household statistics in Figures 5.2 to 5.4 and IRHI in Figure 5.1 reveal that IRHI trends differ in good and bad economic times. Years in which countries experience steady income growth – such as Greece, Spain and Italy in the pre-crisis period – coincide with significant increases in IRHI. Average income drops – and increasing unemployment – appear to be linked to decreases in IRHI; notably in Greece and Portugal in the post-crisis period. Our subsequent analysis is motivated by these observations: why does IRHI follow a pro-cyclical pattern, with significant increases (decreases) during times of economic growth (recession)? The decomposition will focus on the pre-crisis (2004-2010) and post-crisis (2010-2013) periods. This is because these periods encapsulate clear phases of economic growth or decline for countries, while the crisis period (2007-2010) often includes mixed periods of both.[19]

5.4.2 Decomposition Results.

Figures 5.5 and 5.6 depict the estimated income mobility and inequality change terms, respectively. The ageing and migration term proves to be unimportant for explaining IRHI evolution.[20] Panels A and B in Figure 5.5 show the results for, respectively, the pre-crisis and post-crisis rotation groups for all countries. The leftmost cluster of bars in panel A shows (in order from left to right) the contribution that market-related mobility (black), transfer-related mobility (grey) and total income mobility (white, and the sum of the previous two terms) had on IRHI changes in Austria between 2004 and 2007. The remaining clusters/panels have a similar interpretation for the different countries and rotation groups. In Figure 5.6 each cluster of bars shows, per country, the effect that market-related inequality change, transfer-related inequality change and total inequality change (sum of the previous two terms) had on IRHI change in that rotation group.

[19] We present the decomposition results of the 2007-2010 period in Appendix Figure A.5.1.
[20] See Appendix Figure A.5.2.

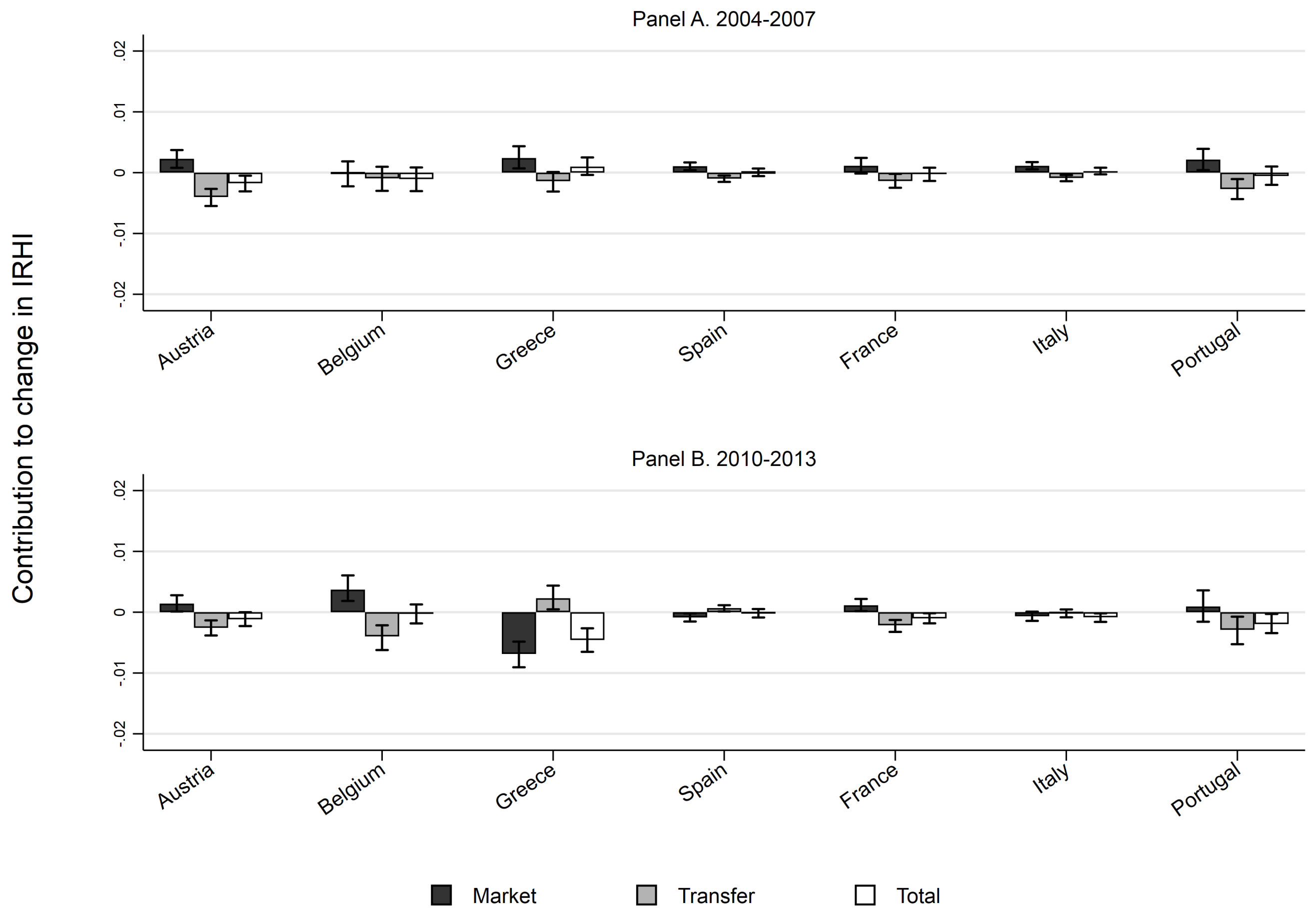
Figure 5.5: Income Mobility Terms



Notes:

1. Figure shows, for each country, the value of the generalized Gini for total and market income per year, per rotation group.

Figure 5.6: Income Inequality Terms



Notes:

1. Figure shows, for each country, the value of the generalized Gini for total and market income per year, per rotation group.

The mobility terms are much larger in magnitude than the inequality terms, and thus appear to be the most important determinant of IRHI change. The reason for this is that the association between ageing and health is stronger than the association between income and health.[21]

5.4.3 Mobility Terms.

Figure 5.5 reveals that, across countries and periods, market mobility tends to be positive and sizable. In comparison – though usually negative – the size and sign of transfer mobility is more varied, and therefore it is often this term which leads to differences in the total mobility term across country-period comparisons.

Recall that the mobility terms can be further split into per-age/sex groups and per-region contributions (see Equation (5.5)). Doing so gives an indication of which age/sex group's income movements are influencing the direction of the separate mobility terms, and therefore gives insight into the patterns in Figure 5.5. While we don't refer to these more detailed results explicitly in the main text, they inform much of the following discussion, and can be found in the appendix for each country, mobility term and for both the pre- and post-crisis period.[22]

The reason for the IRHI increasing effect of market mobility is that improvements in market incomes help the youngest, and therefore healthiest, groups to climb the income ladder, thus increasing the disparities in health by market income. Greece and Portugal are amongst the only countries to not experience significantly positive market mobility in the post-crisis period, as market incomes no longer grew in this period and the very elderly were least affected by shrinking market incomes as they rely more than any other age group on pension incomes. Given the variation in transfer mobility, we distinguish between the following distinctive patterns.

First, one can distinguish between two types of periods and countries: (i) those in which transfer mobility fully compensates for the increase in IRHI caused by market mobility, such as in Austria (pre- and post-crisis), Portugal (pre-crisis period) and Spain and Italy (post-crisis), and (ii) periods and countries in which transfer mobility is close to zero, such as in Belgium and France (pre- and post-crisis), and Spain (pre-crisis). Our results show that transfer mobility tends to be IRHI reducing, as transfers mainly consist of pensions, which disproportionately benefit older and relatively less healthy groups. However, the crucial difference between the above two patterns is the income position of the very elderly (75+) compared to young age groups. We return to this observation below.

[21] For example, in Portugal, the country with the largest predicted health difference between the oldest and the youngest individuals, the difference between the individuals with the minimum and the maximum income-related health value is roughly the same as the difference in predicted health between a 16-25 year old and a 56-65 year old man. See Appendix Table A.5.3.

[22] See Appendix Table A.5.4-Table A.5.10 for the results per age/sex group. The results per region are suppressed as they are small and not important to the decomposition, but are available upon request from the authors.

Second, there are countries for which transfer mobility is large and positive, such as Italy and Greece pre-crisis. Further decomposition of these terms reveals that this can be attributed to household structure. Rather than solely being enjoyed by the old, younger people in Italy and Greece also benefited from the large increase in pension incomes between 2004 and 2007. This is due to young individuals in these countries continuing to live at their parent's home, and therefore benefiting from their parent's (or grandparent's) influx in pension income upon the retirement of the elderly members of the household. This increase in transfer income for the young and the just-retired, and to the exclusion of the very-elderly, led to increasing income disparities between the healthy and the unhealthy, and therefore increased IRHI.

Lastly, there is a remarkable pattern for the southern countries post-crisis, whereby transfer mobility is large and negative in the final rotation group. This is most noticeable in Greece and Portugal, where this term "over-compensated" for market mobility. In such cases total income mobility is negative, and leading to decreases in IRHI between 2010 and 2013. This is due to the "stickiness" of pensions relative to income from work – while the crisis led to a significant fall in the incomes of the young , the incomes of elderly (and, on average, unhealthier) pensioners were less affected. This generated a drop in IRHI.

5.4.4 Market and Transfer Inequality Change.

The *market inequality change* term tends to be positive in most countries and periods as market incomes tend to become more unequally distributed over time (see Figure 5.6).[23] This occurs primarily for two reasons. Firstly, wage growth for the employed is typically positive. Second, as shown in Figure 5.3, the number of retirees in our panels – those who have much lower market incomes – gradually increases over time. Both of these phenomena lead to growing inequality in market incomes, and therefore also growing inequality in (predicted) market-income related health.

By contrast, the *transfer-related inequality change* terms tend to be negative, leading to IRHI decreases. This reflects two facts. First, the redistributive effect of transfers was negative in each year, i.e. market income-related health inequalities $(z_{it}^M \theta(y_{it}^M))$ were always larger than total income-related health inequalities $(z_{it}\theta(y_{it}))$. Second, the redistributive effect became larger (i.e. more negative) over time in most countries. We further find that the *transfer-related inequality change* terms compensate, in most countries and periods, the increase in *market inequality change* such that the changes in total *income inequality* change are usually close to zero.

[23]Portugal (post-crisis) is the only exception which combines an increase in market inequality change with a decrease of the generalized Gini index (see Figure 5.4). This happens because the partial association between income and health – the $\theta(.)$ function – is steeper and more concave in Portugal than other countries (see Appendix Table A.5.3), and because average incomes were declining at the same time.

The most important social transfers, in terms of change in the redistributive effect, are pensions. To demonstrate this, we repeat our decomposition and redefine transfer income to only include income from "old age benefits" and "survivor benefits",[24] attributing the remaining transfer components to market income. The results are shown in the Appendix Figure A.5.3. Although the magnitudes change slightly for some countries, we observe much the same pattern as in Figure 5.6. This confirms that old-age and survivor benefits are the primary source of the increasingly redistributive effect of transfers over time.

In addition to transfers lowering IRHI through the mobility terms, our results identify a secondary IRHI reducing mechanism of transfers, and more specifically pensions. The $\theta(.)$ function describes the association between income and health conditional on age (and gender and region). Pensions reduce IRHI rises by providing income to, on average, poorer individuals, thereby reducing disparities in market income related health by improving the poor's (age-independent) health and income rank. Those that benefit from pensions are in worse health, not only because they are old, but also because their market incomes provide little market income-related health.

The smaller association between income and health relative to the association between age and health means that the pro-cyclical pattern seen in the *mobility* terms is less obvious for the *inequality change* terms. For each of the crisis countries, the effect of total-income *inequality change* does in fact switch from positive to negative (Greece, Spain, Italy) or becomes more negative (Portugal), when comparing the results from 2004-2007 to 2010-2013. However, with the exception of Greece these changes are quantitatively unimportant.

5.4.5 Pension Policies and IRHI.

The decomposition results highlight the particular importance of *transfer mobility* in determining the trends in IRHI: the income position of the old and very elderly (75+) as compared to younger age groups turned out to be crucial. A natural next step, and the goal of this section, is to check whether different trends in transfer mobility are related to differences in pension policies across countries. This is a first step towards understanding the role of pension policies for IRHI trends, although we acknowledge that the selection of 7 countries (see data section) inevitably restricts the scope of our analysis.

[24]The EU-SILC data is organized in such a way that separating these components, and focusing only on old age benefits, is impossible for the early rotation groups, such as the 2004-2007 rotation groups. However, for the years in which we can separate these components we find that the average contribution across all countries of old age benefits (pensions and other lump cash benefits afforded to those who have reached the required age) to household income is approximately €7,000 while for survivor benefits the average contribution is €300.

We restrict the analysis to the pre-crisis 2004-2007 period because – in contrast to later periods – this was a time of relatively normal economic growth. The severity of the crisis (and the policy reaction to it) in the post-crisis period differed substantially across countries. Restricting the period to a time of similar economic growth across countries facilitates a clearer cross country comparison of pension policies.

Panel A of Table 5.3 shows the average change in levels and ranks of transfer incomes between 2004 and 2007 for different age groups (as defined in 2007). *Transfer mobility* in Portugal and Austria led to large reductions in IRHI because the very elderly enjoyed gains in transfer income (ranks) relative to working-age groups, especially the young. The effect of *transfer mobility* is muted in Belgium, Spain and France because the very elderly's relative transfer income position has stagnated. Finally, as discussed in the previous sections, in both Greece and Italy the transfer incomes of the young largely outperform those of the 65+ (Italy) and 75+ (Greece) age groups. Therefore, even in situations where newly retired individuals in the 56-65 group are compensated for their loss in market incomes, it is apparently not always sufficient for the net effect of *transfer mobility* to be IRHI reducing. Because of the large drops in predicted health as individuals' age, any relative income losses for those in the 66-75 and especially those in the 75+ age categories will have IRHI increasing effects.

Table 5.3: Pension Scheme Characteristics and Income Redistribution

	Austria	Belgium	Greece	Spain	France	Italy	Portugal
Panel A. Change in transfer income levels and ranks 2004-2007							
Change in transfer income levels 2004-2007 per age group (€)							
16 - 35	-314	323	486	81	988	358	-87
36 - 55	-72	352	327	519	730	-34	187
56 - 65	4635	3187	3913	2341	2183	1602	1264
66 - 75	579	450	283	1241	1003	68	952
76+	1950	-101	240	357	3	346	179
Change in transfer income ranks 2004-2007 per age group (€)							
16 - 35	-21	6	-13	-13	5	7	-19
36 - 55	-16	-6	-8	0	-3	-5	-4
56 - 65	68	39	67	33	36	27	23
66 - 75	-11	-16	-22	1	-9	-20	24
76+	24	-28	-10	-15	-42	-13	-4
Panel B. Pension generosity indicators (2007)							
Replacement rate (lifetime)^{i}	0.91	0.63	1.1	0.85	0.63	0.78	0.69
Replacement rate (current)^{ii}	0.62	0.44	0.4	0.48	0.6	0.49	0.47
Pension wealth^{i}	9	5.6	13	10	8.1	8.4	7.9
Panel C. Pension generosity indicators (2007)							
Indexation policies (2004-2007)^{i}	Disc. prog	Price*	Disc.	Price	Price	Price prog.	Price prog.

Notes:

1. Panel A. shows the changes in transfer income levels and corresponding ranks between 2004 and 2007 for 5 different age groups.
2. Panel B. shows 3 different sets of pension generosity indicators; the replacement rate as a ratio of the pension entitlement over the average annual life-time income of a hypothetical man who earned the mean income, the replacement rate as a ratio of the current income of pensioners over the current (2007) earnings of those aged between 50 and 59, and pension wealth, measured as the number of years of average annual income one can expect to receive upon retirement, taking into account life-expectancy, indexing rules and retirement age.
3. Panel C. classifies the different indexation methods used across countries; “disc”= discretionary increases, “price” = price indexed, “price*” = price index for a select basket of goods, “prog.” = indexed or discretionary increases favour poor pensions with larger increases.
4. (i) Source: OECD (2005, 2007). (ii) Source: Eurostat (2018).

The phenomenon of relatively lower incomes of the very elderly (75+) compared to more recent retirees is observed across Europe and elsewhere (OECD, 2008). In our setting, the key factor is how transfer incomes change the income position of the very elderly (i.e. how much they move up or down the income ladder), relative to others in the population – especially young groups. Differential growth in pensions compared to transfer payments to other age groups can be driven by several factors. First, if there is a gradual rise in pension contributions leading to more generous pensions for newly retired, the relative losses of the already-retired will be larger. This will be amplified if pension incomes from the recently retired are shared with younger household members while no similar sharing mechanism holds for the very elderly. Secondly, the indexation policy of pensions matters. If indexation is pegged to inflation, the real value of pensions will not increase. However, some countries use other indexing rules such as pegging pensions to average earnings, or "progressive" indexation, with smaller pensions enjoying higher proportional increases (OECD, 2009). Third, the age at which an individual retires will have consequences for the *transfer mobility* term. Because poor health rises sharply with age, the IRHI reducing effect of an increase in pension income at retirement will be larger the later-in-life an individual retires.[25]

In order to get a sense of the role of the different institutional settings, panel B of Table 5.3 shows a number of different indicators of pension generosity in 2007; while panel C shows the type of indexing rules in effect per country during the 2004-2007 period. There is little evidence to suggest that generosity is related to the IRHI reducing effect of pensions. Austria consistently has the highest level of pension generosity, while Portugal has amongst the lowest, yet the *transfer mobility* for both of these countries significantly reduced IRHI. Belgium, France and Spain occupy different ranks according to the generosity measure used, yet for all of these countries *transfer mobility* is close to zero. As emphasized earlier, it is the *relative* changes in transfer income rank which are crucial to determining *transfer mobility*, and for those who are already retired pension generosity may do little to increase their income rank.

Indexation policy, shown in panel C of Table 5.3, may also be important for differences in *transfer mobility*. Austria, which employs yearly discretionary increases in pensions, has adjusted pensions between 2004 and 2007 in a progressive manner (Whitehouse, 2009). Benefits rose with prices up until the median pension, while all pensions above the median were increased by a flat amount. Portugal had a similar progressive indexation. Belgium, France and Spain, applied pension indexation with prices, without any progressivity adjustments (OECD, 2007). Because the very elderly tend to have lower pension benefits, any progressivity in indexation will naturally favour them, thus plausibly

[25]However, due to large age brackets, our decomposition may fail to pick up the IRHI effect of small differences in the retirement age between countries. Moreover, the income position of the very elderly might not be directly affected by this; but only indirectly as compared to the newly retired.

increasing pension incomes ranks for these groups relative to younger groups, and reducing IRHI (OECD, 2009).

The age of new retirees in Austria between 2004 and 2007 is also important for *transfer mobility*. The large increase in transfer income for those aged above 75 is in part due to a number of newly retired pensioners who, in 2007, were aged 75 or more. According to calculations using our data, approximately 10% of the new retirees in this period fit this description in Austria. Thus, the influx of transfer income to these old age and relatively poor-health individuals increased the IRHI reducing effect of the rank mobility from transfers.

5.4.6 Greek Austerity Measures and IRHI.

The most drastic policy changes in this period were enacted in Greece. In exchange for two bailout packages in 2010 and 2011, the Greek government introduced a wide ranging set of austerity measures. Among these were cuts in social transfers such as pensions and unemployment benefits, taxation of pensions above €1,400 a month by 5-10%, and freezing mandatory increases in public pensions between 2011 and 2015.[26]

As mentioned above, the pattern for the inequality change term for Greece between 2010 and 2013 is noticeably different from other countries, as the transfer term is positive while the market term is negative. The decrease in absolute income inequality deriving from the large drop in income from work over this period means that market inequality change is negative, leading to reductions in IRHI. The positive sign for transfer inequality change indicates that the reduction in inequality between 2010 and 2013 was larger due to market income-related health changes than considering total income-related health changes. In other words: the redistributive effect of transfers declined as a result of cuts in social transfers due to the austerity measures, especially for pensions.

The consequences of the Greek austerity measures are less obvious when looking at the *mobility* results between 2010 and 2013, though they are visible in the 2010-2011 comparison when the *transfer mobility* term is large and positive.[27] The immediate impact of the Greek austerity measures was a worsening of the incomes of the elderly relative to the young, as the drop in pensions between 2010 and 2011 was larger than the drop in income from work. This worsened the relative income position of older groups, and increased IRHI. However, the *transfer mobility* term switches sign to become negative between 2011 and 2012. This is due to the sudden nature of the cut in transfer incomes, compared to the more gradual decline in market incomes. While incomes were already falling in Greece between 2010 and 2011, it is in the subsequent two years that the largest falls occur (see

[26]OECD (2013) provides a list of pension reforms that occurred as a result of these austerity measures.

[27]See Appendix Figure A.5.4.

Figure 5.2). Between 2011 and 2013, incomes from work in Greece shrunk sufficiently to outweigh the initially IRHI increasing effects of the austerity measures.

5.5 Conclusion

We believe we make a number of contributions to the literature on health inequalities. First, for a range of European countries, we show for the first time how IRHI have evolved between 2004 and 2013, a time period that covers the largest global economic contraction in the post-war era. We document distinct time and geographic trends in IRHI. Before the crisis, southern countries, and to a lesser extent continental countries, saw IRHI rising. After 2008, IRHI started falling in the southern countries that were most affected by the crisis. These European trends confirm the largely pro-cyclical pattern of IRHI documented for 2 countries (China and Spain) in earlier work (Baeten et al., 2013; Coveney et al., 2016).

Secondly, using a novel decomposition, we provide important new empirical regularities concerning IRHI growth. We find that market incomes tend to increase inequalities in health, while the relation between social transfers and IRHI reveals a more varied pattern, in some cases decreasing and in other cases increasing IRHI. This mixed pattern occurs because social transfers – most importantly pensions – are targeted at the elderly and other poor groups who are otherwise excluded from gains from income growth, but also because in some countries, the young tend to live longer in their parental household and therefore benefit from their pension benefits upon retirement. The decomposition also explains the – perhaps initially puzzling – finding that IRHI falls during crises. This occurs as the logical consequence of the stickiness of pensions relative to income from work, and the age-based income re-ranking effect that it generates.

Finally, we examine how government policies relate to IRHI change. We look at the heterogeneity in the IRHI decreasing effect of transfers across countries and time, and find that the most "successful" pensions payments (in terms of reducing IRHI) are those that do not leave the very elderly (75+) groups behind in times of income growth. Our results also demonstrate that the large reduction in pension payments that occurred between 2010 and 2013 due to the austerity measures in Greece initially coincided with an increase in IRHI, and is likely to have dampened the IRHI reducing effects of transfers in later years.

Based on these empirical findings, our results suggest that government transfer policies can and do appear to have a large effect on IRHI. Especially in times of crisis, pensions help to reduce IRHI by improving the relative income position of the elderly. In that sense, they can be argued to add a "silver lining" to the generally dark future prospects characterizing recessions. In periods of economic growth, however, transfers may not provide adequate protection for these groups. Key is the income protection afforded to the elderly, and in particular the very elderly, a group whose population share is likely to keep growing in the near future. Our results demonstrate that in situations where this group is excluded from gains, the net effect of transfers may no longer be IRHI reducing. But the finding also points at potential policy levers. Governments that are concerned with rising levels of IRHI should develop policies that improve the relative incomes of the very elderly. While our descriptive decomposition method can not causally assess the IRHI reducing effectiveness of alternative policies, our findings do suggest that pension generosity alone does not guarantee lower levels of IRHI. Other pension related policy options that favor the eldest groups, such as progressive indexation and appropriate discretionary increases, have greater potential to successfully reduce IRHI. Finally, it is worth noting that, while in general IRHI is pro-cyclical, the Greek experience shows that austerity measures can kill much of the IRHI reducing effect of pensions during crises, thereby removing most of the silver from the lining.

5.A Appendix

Table A.5.1: EU-SILC Data Available per Country

	Observations per Year (Unbalanced)											Missing Rotation Groups	Missing Health Info	Selected Countries
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total			
Austria	4,674	7,113	10,458	13,393	10,955	11,057	11,493	8,158	5,434	2,532	85,267		0%	✓
Belgium	2,571	6,122	9,063	11,803	11,487	11,382	11,539	7,803	4,928	2,501	79,199		1%	✓
Bulgaria	0	0	5,125	7,161	9,796	12,336	13,776	11,321	6,759	2,784	69,058	1,2	0%	
Cyprus	0	4,441	6,589	8,468	8,087	7,557	9,106	6,895	5,101	3,232	59,476	1	0%	
Czech Republic	0	8,628	14,856	19,293	22,644	19,713	18,209	12,890	9,033	4,565	129,831	1	15%	
Denmark	3,015	5,352	7,376	8,243	7,610	7,326	7,155	5,215	3,418	1,682	56,392	7	50%	
Estonia	2,232	3,415	6,982	10,005	10,849	11,308	11,213	7,844	5,068	2,646	71,562	7	15%	
Greece	3,526	6,405	9,831	12,345	14,119	15,043	14,784	9,429	5,475	2,573	93,530		0%	✓
Spain	7,875	15,895	21,989	28,809	30,081	30,217	30,229	20,643	13,021	5,856	204,615		1%	✓
Finland	4,018	7,528	10,602	13,578	13,222	12,569	15,551	11,428	7,878	4,861	101,235		53%	
France	2,108	5,268	8,392	11,392	12,697	12,666	12,854	9,209	6,051	2,788	83,425		0%	✓
Croatia	0	0	0	0	0	0	8,511	14,628	9,336	2,609	35,084	1,2,3,4	42%	
Hungary	0	8,351	13,127	18,452	18,623	20,334	20,410	14,414	9,626	3,961	127,298	1	1%	
Ireland	1,500	3,667	6,020	6,111	4,531	1,870	2,668	2,306	1,498	812	30,983	4,5,6	0%	
Iceland	1,561	2,980	4,140	5,398	5,427	5,409	5,658	3,998	2,480	1,185	38,236		5%	
Italy	13,335	24,769	35,329	44,619	44,273	43,067	40,305	26,562	15,656	6,508	294,423		1%	✓
Lithuania	0	4,910	7,969	10,913	10,472	11,211	11,603	8,225	5,636	2,874	73,813	1,2,5,6	11%	
Luxembourg	7,602	7,522	7,726	7,745	7,513	6,136	8,409	7,276	6,542	1,680	68,151	7	0%	
Latvia	0	5,408	7,399	9,269	9,020	12,202	12,992	9,505	6,055	3,007	74,857	1	1%	
Malta	0	0	2,710	4,979	6,454	8,482	8,716	6,501	4,601	2,299	44,742	1,2	0%	
Netherlands	0	13,604	15,310	19,623	19,519	18,254	19,134	12,847	7,719	3,845	129,855	1	48%	
Norway	1,538	1,576	1,542	1,503	1,355	2,499	3,619	3,238	2,284	1,411	20,565	7	49%	
Poland	0	18,705	27,739	34,675	33,694	31,671	30,803	21,878	14,029	6,907	220,101	1	6%	
Portugal	2,946	5,456	7,180	7,852	6,692	6,184	6,471	5,915	4,956	3,005	56,657		0%	✓
Romania	0	0	0	12,760	16,525	16,282	12,001	7,825	3,923	0	69,316	1,2,3,7	0%	
Sweden	3,606	6,859	9,157	12,319	12,441	12,431	12,223	7,993	4,866	2,200	84,095		50%	
Slovenia	0	15,702	20,616	24,546	24,908	25,379	25,239	16,505	9,666	4,348	166,909	1	64%	
Slovakia	0	6,424	9,377	12,027	12,524	13,472	13,524	9,877	6,584	3,137	86,946	1	0%	
United Kingdom	0	11,799	14,933	17,458	16,804	15,615	15,111	9,396	5,523	2,509	109,148	1	5%	

Notes:

1. Table shows the number of observations available per country in the 2004-2013 longitudinal release of EU-SILC, and the criteria used to select countries for analysis. Data are presented with minimal cleaning, so numbers do not necessarily represent usable final sample sizes. In addition to observations per year, the numbers of the missing rotation groups between 2004 and 2013 are also listed, as well as the proportion of missing self-assessed health – our two main criteria for determining country selection. Of countries without missing information between 2004 and 2013, 3 countries are excluded due to high % of missing health data: Iceland, Sweden and Finland.

Table A.5.2: Percentage of Market and Transfer Income Due to Pensions and Wages per Rotation Group

		Rotation Group						
		1	2	3	4	5	6	7
Austria	**Pension %**	40.77	42.03	44.11	41.73	39.64	39.85	38.77
	Wage %	70.91	69.38	66.15	68.49	66.62	67.11	67.4
Belgium	**Pension %**	32.69	33.2	35.32	32.95	32.8	30.65	31.06
	Wage %	68.35	66.96	65.1	66.95	67.94	66.87	66.85
Greece	**Pension %**	63.69	62.03	62.04	61.36	61.19	63.57	62.74
	Wage %	57.21	53.09	55.71	57.36	58.19	56.83	56.04
Spain	**Pension %**	54.85	60.21	54.49	52.7	48.63	37.18	33.12
	Wage %	77.57	78.53	80.88	80.33	79.95	73.59	71.92
France	**Pension %**	39.39	41.04	40.65	41.37	44.62	45.63	46.14
	Wage %	69.56	70.77	69.44	67.35	65.07	65.2	64.94
Italy	**Pension %**	57.59	55.91	55	52.47	54.16	55.34	54.87
	Wage %	60.48	62.73	61.23	62.63	62.38	60.88	61.76
Portugal	**Pension %**	41.73	45.42	45.69	43.12	46	47.51	45.91
	Wage %	75.13	76.66	77.94	78.79	77.75	79.21	78.18

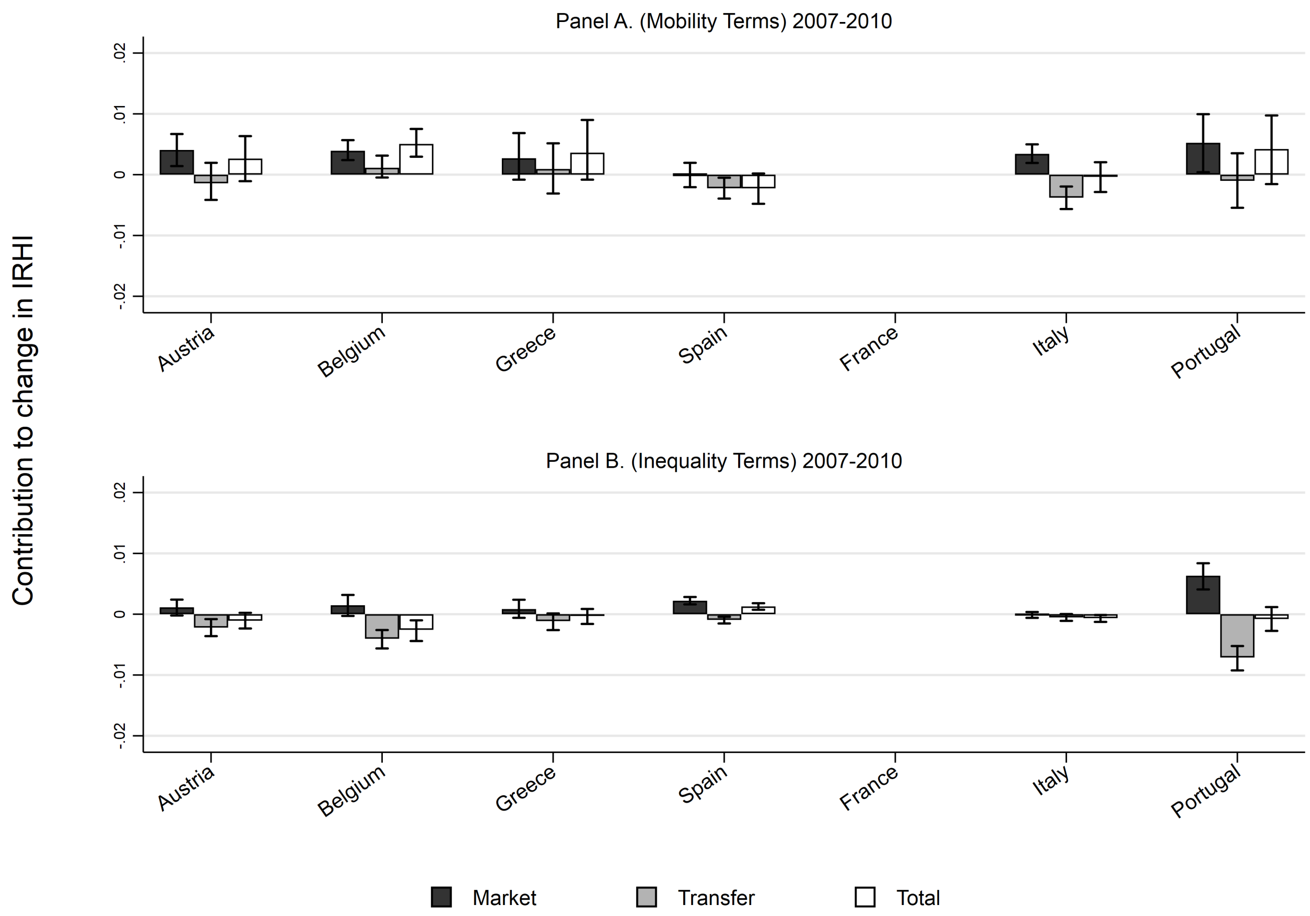
Notes:
1. Table shows for each rotation group and country the percentage of transfer income from pensions (first row per country) and the percentage of (gross) market income from wages (second row per country), computed on the sample for which component information is available. Our data indicate that on average wages and pensions are the most important sources of income within market and transfer income, respectively. Our definition of pensions includes public pension payments, care allowances, disability cash benefits, lump sum payments at the time of retirement and other cash benefits. It does not include any payments from private pension plans, which enter the market income definition. Wages include all employee cash or near cash income. See Table 5.2 for further information on the components of transfer and market income.

Table A.5.3: Interval Regression Results

Variable	Austria	Belgium	Greece	**Country** Spain	France	Italy	Portugal
Eqinc	0.0321***	0.0388***	0.0268***	0.0193***	0.0220***	0.0133***	0.0730***
Eqinc^2	-0.0027***	-0.0039***	-0.0031***	-0.0018***	-0.0017***	-0.0001***	-0.0106***
F 16-25	-0.0017	-0.0054*	0.00178	-0.0011	-0.0084***	-0.0014	-0.0062
M 26-35	-0.0138***	-0.0179***	-0.0118***	-0.0132***	-0.0179***	-0.0137***	-0.0242***
F 26-35	-0.0119***	-0.0226***	-0.0082***	-0.0153***	-0.0245***	-0.0152***	-0.0263***
M 36-45	-0.0295***	-0.0310***	-0.0219***	-0.0267***	-0.0324***	-0.0276***	-0.0400***
F 36-45	-0.0311***	-0.0391***	-0.0245***	-0.0291***	-0.0386***	-0.0312***	-0.0532***
M 46-55	-0.0541***	-0.0534***	-0.0422***	-0.0424***	-0.0537***	-0.0446***	-0.0699***
F 46-55	-0.0568***	-0.0530***	-0.0490***	-0.0499***	-0.0595***	-0.0510***	-0.0969***
M 56-65	-0.0834***	-0.0585***	-0.0785***	-0.0684***	-0.0670***	-0.0713***	-0.1160***
F 56-65	-0.0679***	-0.0589***	-0.0866***	-0.0813***	-0.0674***	-0.0806***	-0.1520***
M 66-75	-0.0827***	-0.0587***	-0.1190***	-0.0864***	-0.0906***	-0.1040***	-0.1510***
F 66-75	-0.0870***	-0.0747***	-0.1440***	-0.1140***	-0.0929***	-0.1250***	-0.1940***
M 76-85	-0.1220***	-0.0818***	-0.1860***	-0.1220***	-0.1330***	-0.1510***	-0.1950***
F 76+	-0.1480***	-0.0973***	-0.2160***	-0.1580***	-0.1290***	-0.1800***	-0.2480***
Observations	57,028	50,848	66,192	139,372	61,932	195,552	59,084

Notes:
1. Table shows the interval regression results, used to generate the ratio-scaled health variable and the coefficients of which are used in the decomposition. *Eqinc* refers to equivalized household income. The constant has been suppressed, as well as the regional dummies because they are small and not important to the decomposition. Full regression results are available from the authors upon request. Robust standard errors used, clustered at the individual level.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

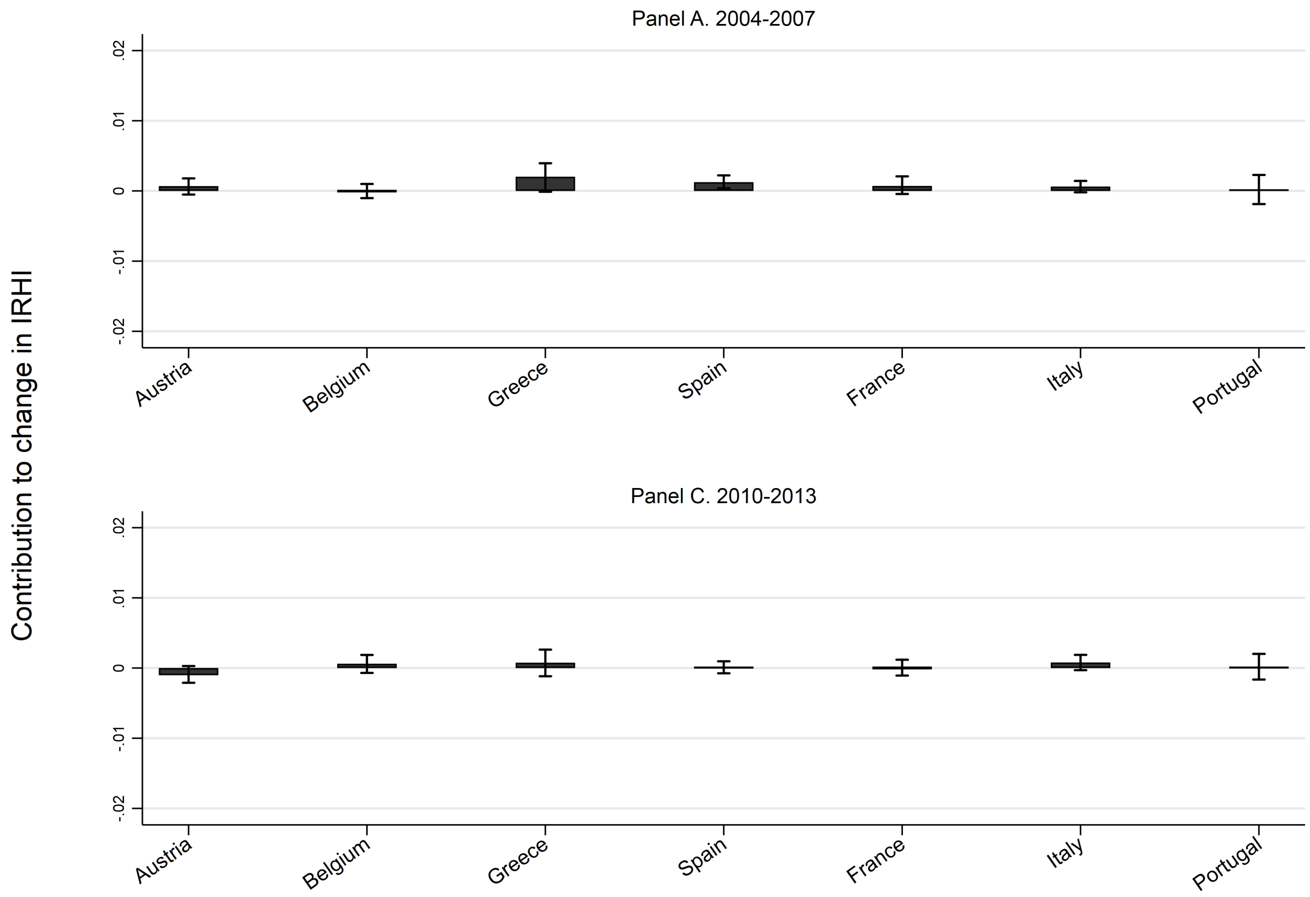
Figure A.5.1: Decomposition of 2007-2010 Results



Notes:

1. Figure shows the results of the inequality and mobility decomposition terms for the 2007-2010 rotation group. Notice that France is excluded due to data issues.

Figure A.5.2: Non-Income Factors



Notes:

1. Figure shows the results of the ageing and migration decomposition term (term 5 in equation 5) for the 2004-2007 and the 2010-2013 rotation groups.

Table A.5.4: Sub-Mobility Terms for Austria

	Pre-crisis (2004-2007)						Post-crisis (2010-2013)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	4.87	71.37***	-8.802	2882.9*	-429.8	5.16	37.83	-23.89	1018.1	-912.3
F 16-25	-0.00172	4.94	12.65	-22.37	56.78	91.47	3.1	27.09	-29.42	-498	-988.5
M 26-35	-0.0138***	7.12	-8.664	-18.94	-1855.7	64.65	7.79	0.567	-35.89**	-1260.8	-294
F 26-35	-0.0119***	6.3	51.41*	-23.03	2128	-1042.6	6.89	11	-13.58	-1627.5	146.1
M 36-45	-0.0295***	10.47	36.14*	-17.28*	1237.3	-472.7	7.42	37.25*	1.276	529.9	357.6
F 36-45	-0.0311***	10.6	10.39	-23.77**	-559.8	-207.2	8.81	21.4	-15.11	-737.3	-37.78
M 46-55	-0.0541***	9.37	35.45*	-15.97	1151.8	-324.9	10.33	24.95*	-4.515	37.28	-188.3
F 46-55	-0.0568***	9.53	0.00472	-12.21	-368.7	580.1	10.42	-1.454	-10.09	-1187.9	225.2
M 56-65	-0.0834***	7.22	-77.99***	71.50***	-5070.7***	4621.3***	8.1	-53.67**	36.37**	-3966.7***	2758.0***
F 56-65	-0.0679***	9.15	-80.39***	58.81***	-5330.1***	4365.0***	8.94	-59.19***	46.78***	-4752.6***	3481.0***
M 66-75	-0.0827***	6.2	12.6	-15.89	-908.4	899	6.89	-5.316	22.68	-1347.7	1384.9*
F 66-75	-0.0870***	6.28	-3.961	-10.06	-1068.1	357.8	7.39	-8.168	4.638	-1528.7*	841.7
M 76-85	-0.122***	3	-55.14	18.39	-4198.6*	793.9	3.81	-12.57	-20.96	-1960.7	-4.252
F 76-85	-0.148***	4.96	-16.5	37.10*	-1013	2650.3***	4.94	8.501	8.034	-284	282.5

Notes:

1. Table shows components of the market and transfer mobility terms for Austria in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.

2. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table A.5.5: Sub-Mobility Terms for Belgium

				Pre-crisis (2004-2007)					Post-crisis (2010-2013)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	5.15	23.37	-17.31	2301.2*	-752.3	4.41	26.41	-9.864	1581.2	-418.3
F 16-25	-0.00535*	4.96	35.22	20.25	2269.1	773.9	4.12	26.17	-31.28	1004.7	-537
M 26-35	-0.0180***	6.55	24.89	13.32	2023.3	734.3	6.53	16.92	19.37	1292.9	1057.0*
F 26-35	-0.0226***	6.63	56.92**	7.056	3219.5**	413.4	6.83	45.38**	29.90*	2750.3**	773
M 36-45	-0.0308***	9.51	19.63	-2.297	1703.9*	-15.73	9.68	13.3	4.89	1151.3	666.5
F 36-45	-0.0393***	10.36	1.277	-10.44	855.4	109.6	9.21	11.75	10.32	1305.8	692.9
M 46-55	-0.0528***	11.51	-10.69	-11.94	131.5	181.9	9.05	-8.712	-7.544	-584	214.8
F 46-55	-0.0528***	9.68	-27.39	3.794	-872.5	1172.2*	9.98	-3.926	-11.81	26.7	224.6
M 56-65	-0.0581***	7.06	-74.17**	42.79**	-2624.6*	3626.2***	7.44	-29.29	5.4	-2356.0**	1885.0***
F 56-65	-0.0582***	6.89	-43.04*	35.60*	-1676.9	2736.7***	8.05	-58.41***	28.88*	-4399.6***	3286.4***
M 66-75	-0.0568***	5.61	6.525	-11.55	-209.1	616.2	5.91	-0.961	-13.87	-1658.2*	1009.6*
F 66-75	-0.0732***	6.58	-0.13	-20.41	-1.682	307.7	6.91	-15.82	-10.21	-2345.9**	1344.7**
M 76-85	-0.0805***	3.74	26.43	-50.41**	241.2	-253.7	4.5	-10.79	-4.537	-1052.1*	808.9
F 76-85	-0.0958***	5.78	6.657	-12.84	566.1	-1.225	7.38	11.88	-29.06*	-355.1	124.6

Notes:
1. Table shows components of the market and transfer mobility terms for Belgium in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5.6: Sub-Mobility Terms for Greece

	Pre-crisis (2004-2007)						Post-crisis (2010-2013)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	3.72	34.27	-13.09	1860.6	507.1	3.69	-27.64	-55.38**	-7245.6***	-114.3
F 16-25	0.00178	4.27	9.961	6.878	1737.7	1042.4*	3.74	21.2	-7.963	-5665.0***	431.1
M 26-35	-0.0118***	8.05	32.33	-0.111	1376.2	869.9*	7.8	-2.049	10.89	-4715.2***	-425.8
F 26-35	-0.00824***	9.03	-4.421	18.27	10.24	1358.7**	8	21.4	-26.5	-4770.1***	-71.14
M 36-45	-0.0219***	9.38	31.50*	-9.901	1451.7**	362.7	9.91	13.82	-42.87***	-5466.9***	-1011.4*
F 36-45	-0.0245***	9.55	6.752	-15.18**	690	372.6*	9.55	21.14	-30.62**	-4849.5***	-417.9
M 46-55	-0.0422***	8.43	-11.23	-11.85	-487.7	559.0**	8.34	2.382	-16.29	-5343.5***	-219
F 46-55	-0.0490***	8.47	-20.83	24.94*	-540.9	1680.9***	8.26	-25.08	0.41	-5890.9***	-160.6
M 56-65	-0.0785***	6.91	-38.01	58.17**	-769.7	2796.7***	7.1	-5.091	29.35	-4681.2***	-465.3
F 56-65	-0.0866***	7.63	-20.34	20.25	-841.5	1750.1***	7.8	-19.12	15.17	-3769.6***	-1462.7*
M 66-75	-0.119***	6.17	-2.534	-2.548	-301.9	1298.3***	6.5	-68.47**	75.60***	-3895.3***	-1784.4***
F 66-75	-0.144***	8.06	-15.53	-15.76	-557.8	727.3**	7.22	-9.718	31.61*	-2340.9***	-2265.2***
M 76-85	-0.186***	4.51	-5.529	-28.14**	-328.8	243.7	5.25	47.16**	2.728	-1003.8**	-2582.6***
F 76-85	-0.216***	5.81	18.04	-52.43***	83.33	-202.7	6.85	24.65	22.04	-1134.3**	-2079.0***

Notes:
1. Table shows components of the market and transfer mobility terms for Greece in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5.7: Sub-Mobility Terms for Spain

				Pre-crisis (2004-2007)					Post-crisis (2010-2013)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	4.53	27.52	0.956	1570.5**	317.4	4.28	19.7	-26.74*	-2102.3**	183.9
F 16-25	-0.00107	4.14	55.99**	2.05	2065.8*	309.2	4.03	23.54	6.371	-2059.4***	864.1**
M 26-35	-0.0132***	7.77	25.61	-11.61	2006.1**	122	7.15	8.356	-19.59	-3087.5***	319.5
F 26-35	-0.0153***	8.76	11.91	-23.75**	1375.9	-41.51	7.1	40.75*	-18.90*	-1581.7*	79.76
M 36-45	-0.0267***	10.06	-13.36	-6.835	343	325.3	10.44	30.92**	-35.31***	-1656.6***	-517.7
F 36-45	-0.0291***	10.91	25.30*	-4.839	2003.3***	139.2	10.98	11.25	-28.66***	-2684.3***	-261.7
M 46-55	-0.0424***	8.86	0.297	6.108	307.7	913.7**	10.3	2.699	-13.69	-3077.2***	321.7
F 46-55	-0.0499***	8.89	10.33	8.142	595.3	855.8*	9.98	-5.86	-5.204	-3415.6***	831.0**
M 56-65	-0.0684***	6.75	-64.46***	28.84	-2236.0**	2303.5***	7.19	-47.15**	46.54***	-4617.2***	2700.5***
F 56-65	-0.0813***	7.17	-68.26***	43.51**	-2372.3**	2450.9***	7.48	-66.19***	54.23***	-5573.0***	2601.6***
M 66-75	-0.0864***	5.72	-29.81	6.225	-1383.0*	1475.9**	4.82	-45.17*	44.31**	-3962.5***	1130.4**
F 66-75	-0.114***	6.3	4.318	-14.71	124.8	609.6	6	0.64	9.993	-1625.1**	-109.5
M 76-85	-0.122***	3.88	11.2	-13.52	327.1	548.1*	4.23	8.977	24.52*	-1446.9***	299.6
F 76-85	-0.158***	6.26	20.56	-18.64	963	153.1	6.02	10.86	20.40**	-1511.6***	422.9

Notes:
1. Table shows components of the market and transfer mobility terms for Spain in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5.8: Sub-Mobility Terms for France

	Pre-crisis (2004-2007)						Post-crisis (2010-2013)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	5.5	-10.08	-9.324	-901.9	1704.2	4.77	13.27	-18.3	-1341.3	459.6
F 16-25	-0.00842***	5.49	-25.51	-41.17*	-1087.2	-256.7	3.74	48.06	-4.055	1023.6	-325.8
M 26-35	-0.0179***	8.35	28.27	-6.367	1189.5	562.5	5.55	33.34	7.683	761.1	529.7
F 26-35	-0.0245***	8.86	43.67**	-2.626	2408.9**	117.8	6.44	52.19***	4.266	1720.8*	301.5
M 36-45	-0.0324***	9.8	11.47	-18.63	492.3	-292.6	8.28	36.28**	-11.03	1713.6*	-455
F 36-45	-0.0386***	10.86	29.6	-0.112	1468.9	266.1	8.88	22.61*	-13.62	1440.7	-353.5
M 46-55	-0.0537***	8.77	7.619	-1.254	737.6	945.6	8.31	0.878	-12.98	-1173.4	115.5
F 46-55	-0.0595***	9.42	-26.98	-13.95	-1567.6	467.3	10.35	-21.49	-9.36	-2533.5**	922.7*
M 56-65	-0.0670***	7.09	-76.14**	76.78***	-3491.5*	4429.4***	9.64	-60.14***	46.81**	-4465.9***	4095.9***
F 56-65	-0.0674***	7.8	-54.77*	57.38*	-3215.3*	3443.4***	9.96	-70.55***	40.85**	-4988.4***	3149.1***
M 66-75	-0.0906***	4.69	0.395	-8.21	-928.6	1070.2	6.42	-7.063	0.235	-2034.6*	1623.5**
F 66-75	-0.0929***	6.08	16.65	-33.08*	398.9	-324.2	7.17	13.67	-11.29	-903.1	1322.3**
M 76-85	-0.133***	2.98	40.53**	-18.6	1093.9*	-432.4	4.47	19.5	-22.6	-239.6	337.2
F 76-85	-0.129***	4.3	16.11	-3.895	103.1	706.1	6.02	15.26	-37.92***	-136.4	-202.9

Notes:
1. Table shows components of the market and transfer mobility terms for France in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.
2. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5.9: Sub-Mobility Terms for Italy

			Pre-crisis (2004-2007)						Post-crisis (2010-2013)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	3.8	22.53	1.636	1594.2**	189.7	3.71	17.87	-37.88***	-2115.5***	-432.9
F 16-25	-0.00143	3.84	33.70**	17.88*	2173.1***	214.1	3.84	5.191	-5.838	-3006.8***	1014.6
M 26-35	-0.0137***	8.14	27.29**	-5.054	1897.2***	-52.83	6.71	30.88*	-8.388	-1385.9*	124.3
F 26-35	-0.0152***	7.88	16.02	16.28**	1451.6**	873.1***	5.71	16.04	8.223	-2100.9**	833.4
M 36-45	-0.0276***	9.42	12.08	-13.87**	1410.6***	-317.7	8.24	30.65***	-27.52**	-1505.0**	-496.9
F 36-45	-0.0312***	10.2	8.38	-6.77	1220.8**	-284.8	9.1	29.72***	-29.46***	-1525.0**	-688.7***
M 46-55	-0.0446***	8.18	6.458	-1.24	1395.2***	100.3	9.43	10.16	-28.46***	-2632.2***	-417
F 46-55	-0.0510***	7.95	-5.579	8.953	725.2	475.4	10.62	-5.887	-2.869	-2964.2***	644.7*
M 56-65	-0.0713***	7.67	-36.18***	34.23***	-870.5	1797.9***	8.02	-46.33***	34.16***	-4442.0***	1965.6***
F 56-65	-0.0806***	7.61	-23.22*	19.51**	-760.7	1447.6***	7.78	-43.99***	22.49*	-4229.7***	1543.4***
M 66-75	-0.104***	5.89	-13.98	-12.36	-993.7*	513.3	6.48	-25.89*	15.5	-2513.8***	331.6
F 66-75	-0.125***	7.1	-19.09*	-27.02***	-866.0**	-293.3	7.39	-27.28*	20.84*	-2427.6***	385.4
M 76-85	-0.151***	4.54	-5.046	-20.55**	-317.2	314.9	5.11	-0.0254	29.53**	-1530.4**	670.3*
F 76-85	-0.180***	7.79	-9.769	-9.822	-383.4	333.6	7.85	18.03*	14.25*	-817.2*	150.5

Notes:

1. Table shows components of the market and transfer mobility terms for Italy in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.

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

Table A.5.10: Sub-Mobility Terms for Portugal

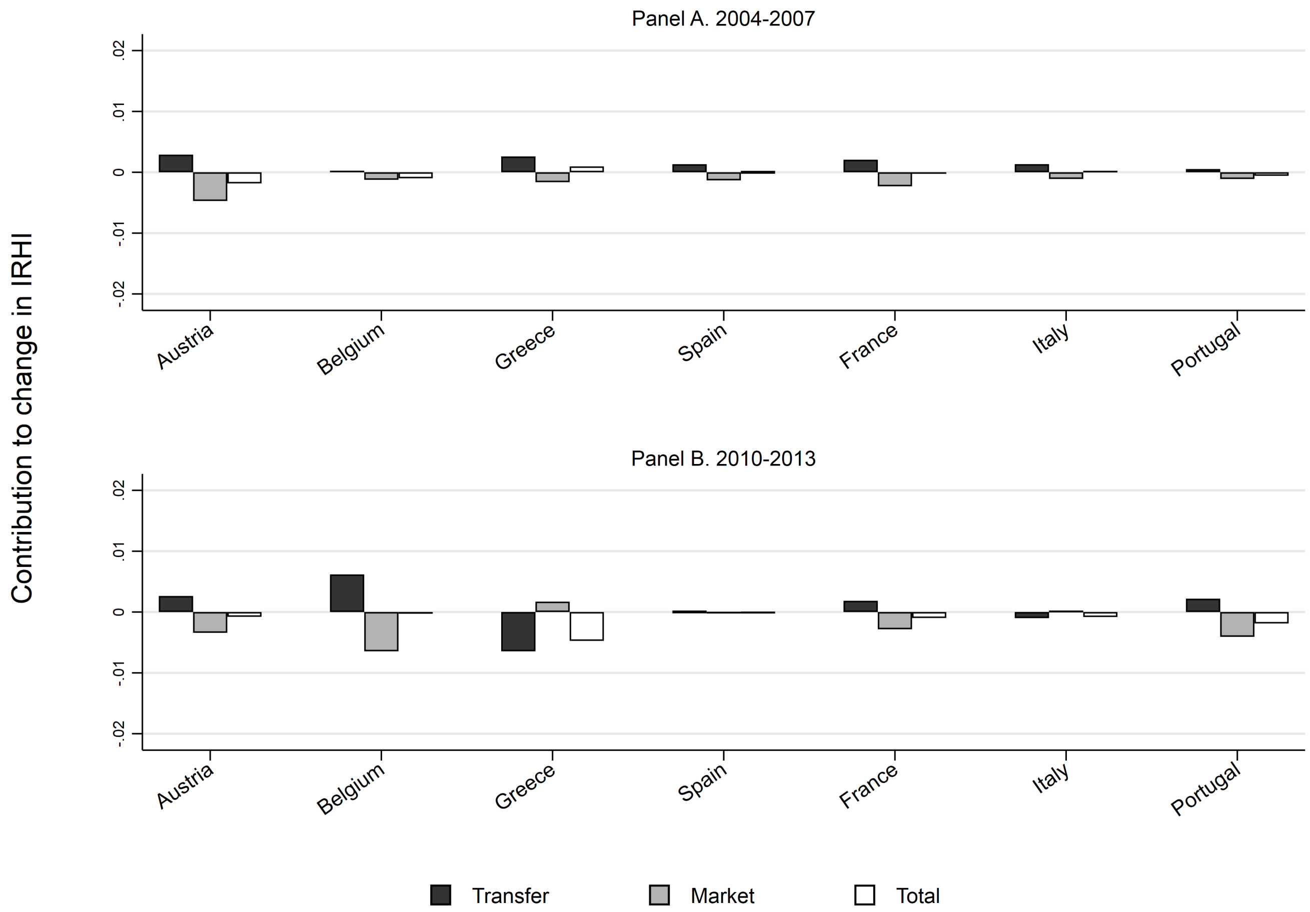
	Pre-crisis (2004-2007)						Post-crisis (2010-2013)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Age-Sex Group	**Coeff.**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**	**Prop.**	**Market Weight**	**Transfer Weight**	**Market Income (€)**	**Transfer Income (€)**
M 16-25	-	5.37	6.78	0.641	-478	555.2*	3.62	40.45*	-13.89	-703	-138.2
F 16-25	-0.00621	4.38	51.55**	-15.07	1205.6*	-59.6	3.8	19.68	-10.75	-1027	195
M 26-35	-0.0242***	7.83	6.407	-13.9	368.1	-82.51	6.04	23.16	-26.78	-901.7	-557.5
F 26-35	-0.0263***	7.53	48.31*	-37.13*	1301.3**	-489.5	6.49	25.21	-23.25	-803.3	25.49
M 36-45	-0.0400***	8.25	38.70**	-9.996	1150.2**	-213	8.74	-7.735	-31.05**	-2638.2***	85.41
F 36-45	-0.0532***	9.09	2.019	-15.25	13.51	-230.6	10.25	-0.255	-26.90***	-2441.4***	43.39
M 46-55	-0.0699***	8.89	0.851	-2.633	-291.4	493.1*	8.34	-6.581	-3.896	-1873.0***	475.8
F 46-55	-0.0969***	9.7	-2.91	7.746	-392.7	589.4**	9.49	-0.275	4.726	-2092.6***	536.8**
M 56-65	-0.116***	6.34	-59.86**	35.16	-1547.2**	1574.9***	7.82	-27.77	39.38**	-2781.4***	1272.1***
F 56-65	-0.152***	7.75	-48.50*	14.73	-1276.1**	1085.7***	8.85	-39.20*	34.38*	-2684.6***	1050.9***
M 66-75	-0.151***	6.02	-32.61	31.79*	-1024.1*	1159.5***	6.96	-31.33*	46.79**	-2158.3***	907.7*
F 66-75	-0.194***	8.18	-3.944	18.91	-452.4	829.8***	7.5	3.48	16.25	-1248.7***	215.1
M 76-85	-0.195***	3.68	13.28	7.988	-186.6	120.9	5	29.69*	-26.12	-326.1	-447.6
F 76-85	-0.248***	6.99	-8.314	-13.2	-587.6	179.3	7.1	29.60*	1.347	-537.6	467.8

Notes:

1. Table shows components of the market and transfer mobility terms for Portugal in the 2004-2007 and 2010-2013 decompositions. Column 1 shows the age/sex coefficient in the health regression, column 2 shows the proportion of individuals in each age/sex group in 2007, columns 3, 4, 5 & 6 show the change between 2004 and 2007 for each age/sex groups in market income weights, transfer income weights, actual market incomes and transfer market incomes, respectively. Columns 7-11 show the identical information for the post-crisis 2010-2013 period. These changes are summarized by a no-constant regression where the change in the income weight/amount is regressed on a set of age/sex dummies which refer to the last wave in the rotation group. All currency amounts in 2013 euros.

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

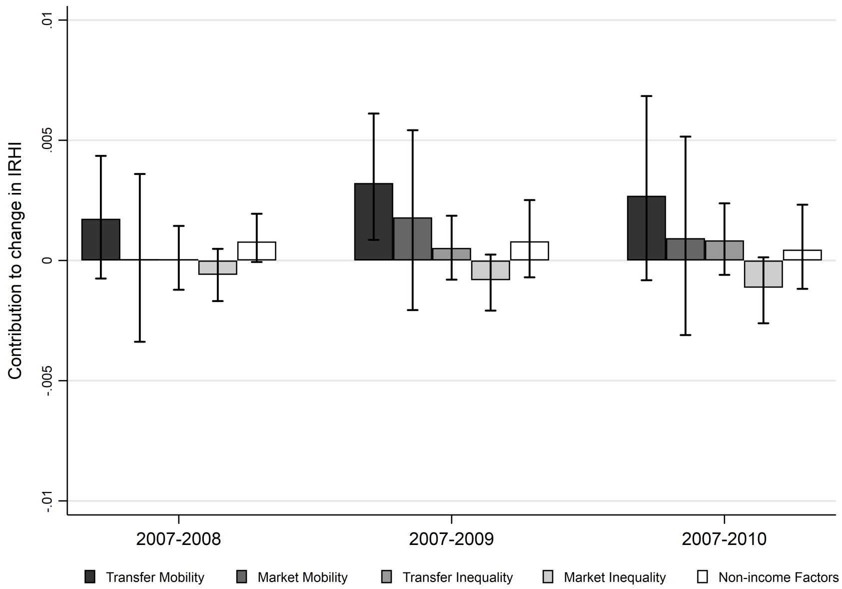
Figure A.5.3: Inequality Terms with Transfer Income Consisting Only of Pensions



Notes:

1. Figure shows the inequality term results of a decomposition where transfer income consists of pensions only, and where all other transfer components have been added to market income.

Figure A.5.4: Year-by-Year Decomposition for Greece for 2010-2013 Rotation Group



Notes:

1. Figure shows the full decomposition results for Greece for each year of the 2010-2013 rotation group.

Chapter 6

Conclusion

The designing and implementation of effective, efficient, and ethical policies in the spheres of education and health requires evidence. In this spirit, the chapters of this thesis aim to provide scientific evidence to aid policy makers, as well as uncover promising directions for further research.

Despite the voluminous existing research, the goal of implementing predictable policies exploiting peer effects remains elusive. A 2014 review of the literature in this area concluded, somewhat pessimistically, that "*despite potential temptation, we have not reached the point at which we can reliably use knowledge of peer effects to implement policies that improve outcomes for students and other human subjects*" (Sacerdote, 2014, p. 254). The results presented in Chapter 2 provide the first step towards reaching this goal. Examining the mechanisms through which peer effects operate is vital to providing workable policy guidance, as well as improving our theoretical understanding of such effects. Though a variety of mechanisms have been suggested by the literature our results provide strong support for only one: peer effects arising through social interaction between students. More suggestive evidence points towards these interactions involving group studying outside of the classroom.

The message for future interventions in this area is clear. It is not enough to focus only on achieving an "optimal" mix of students in terms of ability within a classroom. It is also necessary for these students to actually interact with each other in order for the desired peer effects to occur. Encouragingly, the results from this paper suggest that, at least within our experimental context, the institutional manipulation of friendships is possible in the short term. A small intervention in the form of several informal meetings managed to create meaningful bonds among some students (close peers), while not among others (distant peers). A repeat of the experiment by Carrell et al. (2013) with the addition of some encouragement of social interaction between assigned peers would be the logical next step from these conclusions.

In Chapter 3, we turned our attention to one aspect of educational policy that has attracted considerable debate: the large increase in the admissions of international students. Faced with claims that these students take university places and other resources away from locals, advocates for the so-called internationalization of education point out that interaction with foreigners holds many advantages for native students. For instance, a recent letter to parliament by the education minister claimed that the admission of international students helped natives "*learn to deal with diversity*", and that by "*acquiring inter-cultural skills, students learn to cope in different environments*", and gain a "*worldwide view and a vision*" (van Engelshoven, 2018, p. 2). Given that interactions are necessary for diversity to yield benefits (Camargo et al., 2010), a crucial question is then to what extent foreign and native students actually do interact on campus.

Using a novel method to reveal the friendship ties between students we find clear sorting patterns between these students, implying that universities may be missing out on fully realizing the potential benefits of native-foreign interactions. Moreover, forced exposure between natives and foreigners does not appear to be a quick-fix for increasing interaction. While native-foreign student pairs forced to share a classroom are subsequently more likely to be friends, this effect does not hold for natives at risk of exhibiting xenophobic political views, nor for foreigners from countries with radically different cultures from the host country.

These findings have important implications for the debate surrounding the internationalization of education. They suggest that interactions between local and international students represent a potentially beneficial but currently untapped resource, though that exploiting this resources is not straightforward. Future work should investigate the reasons for observed segregation, if more intensive and deliberate interventions could reduce these patterns, and indeed whether the proposed benefits of an international campus do exist.

At least since the Black Report in 1980, which noted the inability of the NHS to halt the widening of socio-economic health disparities in England and Wales since the 1930s (Black et al., 1980), IRHIs have been a public health concern. Indeed, to this day, their reduction remains a primary policy goal of the EU (European Commission, 2009). Despite this, and while their existence has by now been widely confirmed by more than a decade of research, much about the determinants and evolution of IRHIs remains unclear.

Chapter 4 seeks to resolve the important but open question of how IRHIs respond to financial crises, specifically the 2008 Great Recession in Spain. Did the fears of deepening inequalities in health come to pass? Our conclusion is that IRHI actually reduced as a consequence of the crisis. Further investigation shows that this trend is the logical consequence of large health differences between the young and the old, and the differential impact that the crisis had on the incomes of these two groups. The incomes of the elderly stayed relatively flat, protected from the effects of the crisis by the inherent stickiness of pensions, while the incomes of the youngest were disproportionately reduced, especially as a result of the collapse of employment in the construction sector. As a result, the relative income positions of the elderly compared to the young improved during the crisis.

The findings, although based on a narrow context, highlight a perhaps unappreciated role of pensions; their stickiness helps to protect the incomes of those in the worst health (the elderly) during economic crises. However, our focus on a single country prevents us from making more concrete claims about the relationship between IRHI and crises. It is not clear to what degree the institutional settings, demographics, and other country-specific factors may play a role in determining these trends.

It is these limitations that motivated our broadened investigation of IRHI in Chapter 5. We compute IRHI trends between 2004 and 2013 for Spain, Portugal, Italy, France, Belgium, Austria and Greece. This leads to further insights into how IRHI evolved during the Great Recession, and the range of countries permits comparison between those which were heavily and lightly affected. We find that the conclusions in the previous chapter for Spain hold for many of the so-called crisis countries in Europe; the recession tended to decrease IRHI.

The findings from Chapters 4 and 5 may provide some solace to policy makers. While devastating for incomes, the crisis does not appear (at least in the short-run) to have exacerbated existing IRHIs. Instead, disparities in health by income have reduced. However, our analysis also reveals some important qualifiers, as well as potential policy levers. The very elderly - those older than 75 - are extremely vulnerable to being left behind in terms of both income and health. The methods for determining pension benefits for this group are therefore a crucial determinant of the evolution of IRHI. Household structures in some countries, such as Italy, lead to increases in IRHI due to the intergenerational sharing of government transfers. Finally - and importantly - we find that austerity policies like those enacted in Greece decrease the IRHI-reducing effects of pensions.

At first glance the various topics within this thesis may appear unrelated. However, they are united by the fact that their overarching themes – health and education policy – are part of perhaps the most important obligation of modern governments to their citizens. It is imperative that careful and precise scientific evidence should form the backbone of policies in these areas. The chapters contained in this thesis go some way to providing such evidence.

Nederlandse Samenvatting (Summary in Dutch)

De gebieden waar het overheidsbeleid zich in het dagelijks leven van individuen het meest duidelijk voordoet, zijn die van het onderwijs en de gezondheidszorg. Grote aantallen wetten en voorschriften van de overheid bepalen waar, wanneer en hoe burgers omgaan met het onderwijs en de zorg. Het beleid is op deze gebieden niet alleen zichtbaar, het beleid heeft ook zeer belangrijke en verstrekkende gevolgen voor het individu. Regelmatig identificeren filosofen een goede gezondheid en goed onderwijs als de basis van een bevredigend en fatsoenlijk leven en behoren statistieken op het gebied van gezondheid en onderwijsprestaties vaak tot de meest essentiële statistieken van het individuele welzijn en de ontwikkeling van een land.

Vanwege het fundamentele belang van gezondheid en onderwijs, is het absoluut noodzakelijk dat beleid wordt gebaseerd op bewijsmateriaal dat afkomstig is van zorgvuldig en gedetailleerd wetenschappelijk onderzoek. De hoofdstukken van dit proefschrift presenteren het werk dat ik heb ondernomen (met mijn coauteurs), waarbij alle hoofdstukken het thema delen van het leveren van dergelijk bewijsmateriaal in een poging om beleid te adviseren.

De focus van dit proefschrift richt zich op twee verschillende sub-onderwerpen die vallen onder het gemeenschappelijke onderwerp van gezondheid en educatie. Het eerste onderwerp, geadresseerd in hoofdstuk 2 en 3, richt zich op een cruciaal onderdeel van een student zijn educatie: de relatie met zijn medestudenten. Het tweede sub-onderwerp, geadresseerd in hoofdstuk 4 en 5, beschrijft het onderwerp van inkomens-gerelateerde-gezondsheidsongelijkheid ("IRHIs"). Deze ongelijkheden tonen het verschil in gezondheid door inkomen, waar in bijna elke context (ook de Europese), rijkere individuen langer leven en gezondere levens hebben ten opzichte van armere individuen.

Hoofdstuk 2 heeft als doelstelling om een stap dichter-bij het doel te komen van het implementeren van een betrouwbaar en voorspelbaar peer-effecten-beleid door de mechanismen die dergelijke effecten aansturen te onderzoeken. We analyseren de peer-effecten en de wijze waarop deze peer-effecten zich kanaliseren bij een grote Europese universiteit, verspreid over 6 cohorten van economiestudenten. We documenteren kleine positieve effecten tussen de vaardigheden van willekeurig toegewezen peers binnen werkgroepen en de cijfers van studenten. Wanneer we gebruik maken van de structuur van de werkgroepen, dan kunnen we aantonen dat deze peer-effecten alleen voorkomen tussen studenten die vrienden zijn. Dit impliceert dat peer-effecten ontstaan door sociale nabijheid in plaats van via effecten op werkgroep niveau. We laten zien dat sociale nabijheid evolueert in de tijd, dit suggereert dat spillover effecten van toegewezen peers aan werkgroepen van korte duur zijn.

In hoofdstuk 3 geven we inzichten ten behoeve van het huidige beleidsdebat rondom de toegenomen toelatingen van internationale studenten aan veel Europese universiteiten. De lokale-buitenlandse interacties worden vaak gepresenteerd als een centraal voordeel van een internationaal diverse universiteitscampus, maar er is weinig bewijs over de mate waarin dergelijke interacties daadwerkelijk bestaan. Dit hoofdstuk is de eerste die kwantitatief bewijs levert over verdelingspatronen van lokale en buitenlandse studenten. Aan de hand van unieke vriendschapsgegevens die worden ontleend uit de keuzes van studenten aan een grote Europese universiteit, documenteren we een aanzienlijke en substantiële segregatie tussen lokale en buitenlandse studenten. Deze patronen blijven bestaan, zelfs na controle voor vermogen, leeftijd en geslachtsverschillen. We onderzoeken een mogelijke oplossing voor deze geobserveerde segregatie op de campus door te onderzoeken in hoeverre vriendschappen tussen lokale en buitenlandse bevolkingsgroepen kunnen worden aangemoedigd door gedwongen en regelmatige blootstelling binnen kleine jaarlange studiegroepen. Gekoppelde lokale en buitenlandse studenten die gedwongen worden dezelfde groep te delen, hebben een aanzienlijk hogere kans om een vriendschap te vormen. Dit “exposure effect” is echter afwezig voor gekoppelde lokale en buitenlandse studenten waarvoor interacties mogelijk het gunstigst zijn; de koppels waarbij de lokale student het risico loopt op het vertonen van xenofobische politieke opvattingen, en buitenlandse studenten uit landen die cultureel het minst lijken op het gastland.

Regeringen en beleidsmakers hebben lang geprobeerd de ongelijkheden op gezondheidsgebied te verminderen. Er is echter weinig bewijs over de mate hoe dergelijke ongelijkheden zich tijdens de recente financiële crisis hebben ontwikkeld. IRHI-trends in deze periode verdienen speciale aandacht gezien de bezorgdheid dat de crisis mogelijk onevenredig de meest kwetsbare groepen in de samenleving heeft getroffen. Hoofdstuk 4 bespreekt de ontwikkeling van IRHI in Spanje voor, tijdens en na de Grote Recessie van 2008. We zien een sterke daling van gezondheidsongelijkheden na 2008. De daling is vooral het gevolg van het feit dat de financiële crisis onevenredig jongere, gezondere groepen heeft getroffen. We concluderen dat de ongelijke verdeling van inkomensbescherming naar leeftijd op korte termijn gezondheidsongelijkheid kan verminderen na een economische recessie.

Hoofdstuk 5 verbreedt de analyse van IRHI en economische omstandigheden met 6 additionele Europese landen. Voortbouwend op de conclusies van het vorige hoofdstuk, ontwikkelen en passen we een meer genuanceerde decompositie toe om expliciet de rol van veranderende regeringen en hun relatie met de evolutie van IRHI in tijden van economische groei en recessie te onderzoeken. We documenteren een procyclisch patroon van IRHI dat grotendeels kan worden verklaard door de relatieve "stickiness" van ouderdomspensioenuitkeringen in vergelijking met het marktinkomen van jongere groepen. Bezuinigingsmaatregelen verzwakken het IRHI-verminderende effect van overheidsoverdrachten. We concluderen dat pensioenbeleid veel ruimte lijkt te hebben om IRHI te verminderen.

Bibliography

Allport, G.W., Clark, K., Pettigrew, T., 1954. The nature of prejudice. Addison-Wesley Reading, MA.

Angrist, J.D., 2014. The perils of peer effects. Labour Economics 30, 98–108.

Aparicio Fenoll, A., 2010. High-school dropouts and transitory labor market shocks: The case of the spanish housing boom .

Arcidiacono, P., Aucejo, E., Hussey, A., Spenner, K., 2013. Racial segregation patterns in selective universities. The Journal of Law and Economics 56, 1039–1060.

Arcidiacono, P., Nicholson, S., 2005. Peer effects in medical school. Journal of public Economics 89, 327–350.

Ásgeirsdóttir, T.L., Ragnarsdóttir, D.Ó., 2013. Determinants of relative and absolute concentration indices: evidence from 26 european countries. International journal for equity in health 12, 53.

Athey, S., Imbens, G.W., 2017. The state of applied econometrics: Causality and policy evaluation. Journal of Economic Perspectives 31, 3–32.

Bacigalupe, A., Escolar-Pujolar, A., 2014. The impact of economic crises on social inequalities in health: what do we know so far? International journal for equity in health 13, 52.

Baeten, S., Van Ourti, T., Van Doorslaer, E., 2013. Rising inequalities in income and health in china: who is left behind? Journal of health economics 32, 1214–1229.

Bentolila, S., Cahuc, P., Dolado, J.J., Le Barbanchon, T., 2012. Two-tier labour markets in the great recession: France versus spain. The Economic Journal 122, F155–F187.

Black, D., Morris, J., Smith, C., Townsend, P., 1980. Report of the working group on inequalities in health. London: Stationery Office .

Boisjoly, J., Duncan, G.J., Kremer, M., Levy, D.M., Eccles, J., 2006. Empathy or antipathy? the impact of diversity. American Economic Review 96, 1890–1905.

Bonhomme, S., Hospido, L., 2012. Earnings inequality in spain: Evidence from social security data. Trabajo presentado en el XV Encuentro de Economía Aplicada, A Coruña , 7–8.

Booij, A.S., Leuven, E., Oosterbeek, H., 2017. Ability peer effects in university: Evidence from a randomized experiment. The Review of Economic Studies 84, 547–578.

Brunello, G., De Paola, M., Scoppa, V., 2010. Peer effects in higher education: Does the field of study matter? Economic Inquiry 48, 621–634.

Burke, M.A., Sass, T.R., 2013. Classroom peer effects and student achievement. Journal of Labor Economics 31, 51–82.

Burns, J., Corno, L., La Ferrara, E., 2013. Interaction, stereotypes and performance. evidence from south africa. Technical Report. Bocconi University (mimeo).

Camargo, B., Stinebrickner, R., Stinebrickner, T., 2010. Interracial friendships in college. Journal of Labor Economics 28, 861–892.

Carrasco, R., Jimeno, J.F., Ortega Masague, A.C., 2011. Accounting for changes in the spanish wage distribution: the role of employment composition effects .

Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does your cohort matter? measuring peer effects in college achievement. Journal of Labor Economics 27, 439–464.

Carrell, S.E., Hoekstra, M., West, J.E., 2019. The impact of college diversity on behavior toward minorities. American Economic Journal: Economic Policy Forthcoming.

Carrell, S.E., Sacerdote, B.I., West, J.E., 2013. From natural variation to optimal policy? the importance of endogenous peer group formation. Econometrica 81, 855–882.

Carrell, S.E., West, J.E., 2010. Does professor quality matter? evidence from random assignment of students to professors. Journal of Political Economy 118, 409–432.

Coveney, M., García-Gómez, P., Van Doorslaer, E., Van Ourti, T., 2016. Health disparities by income in spain before and after the economic crisis. Health economics 25, 141–158.

Currarini, S., Jackson, M.O., Pin, P., 2009. An economic model of friendship: Homophily, minorities, and segregation. Econometrica 77, 1003–1045.

Dahlberg, M., Edmark, K., Lundqvist, H., 2012. Ethnic diversity and preferences for redistribution. Journal of Political Economy 120, 41–76.

De Giorgi, G., Pellizzari, M., Redaelli, S., 2010. Identification of social interactions through partially overlapping peer groups. American Economic Journal: Applied Economics 2, 241–75.

Doorslaer, E.v., Koolman, X., 2004. Explaining the differences in income-related health inequalities across european countries. Health economics 13, 609–628.

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. American Economic Review 101, 1739–74.

DUO, 2014. Examenmonitor VO 2014. Technical Report. Dienst Uitvoering Onderwijs.

van Engelshoven, I., 2018. Kamerbrief over internationalisering mbo en ho URL: `https://www.rijksoverheid.nl/documenten/kamerstukken/2018/06/04/kamerbrief-over-internationalisering-mbo-en-ho`.

Erreygers, G., 2009. Correcting the concentration index. Journal of health economics 28, 504–515.

Erreygers, G., Van Ourti, T., 2011. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. Journal of health economics 30, 685–694.

European Commission, 2009. Solidarity in health: Reducing health inequalities in the eu. Brussels .

European Commission, 2013. Health inequalities in the eu — final report of a consortium. consortium lead: Sir michael marmot. Brussels .

Eurostat, 2018. Aggregate replacement ratios for pensions (excluding other social benefits) URL: `http://appsso.eurostat.ec.europa.eu`.

Feld, J., Salamanca, N., Zölitz, U., 2018. Are professors worth it? the value-added and costs of tutorial instructors. University of Zurich, Department of Economics, Working Paper No. 293 .

Feld, J., Zölitz, U., 2017. Understanding peer effects: on the nature, estimation, and channels of peer effects. Journal of Labor Economics 35, 387–428.

Fernandez, R., Fogli, A., 2009. Culture: An empirical investigation of beliefs, work, and fertility. American Economic Journal: Macroeconomics 1, 146–77.

Foster, G., 2005. Making friends: A nonexperimental analysis of social pair formation. Human Relations 58, 1443–1465.

Foster, G., 2006. It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. Journal of public Economics 90, 1455–1475.

Frieden, T.R., 2013. Cdc health disparities and inequalities report-united states, 2013. foreword. MMWR supplements 62, 1–2.

Fryer Jr, R.G., Levitt, S.D., 2010. An empirical analysis of the gender gap in mathematics. American Economic Journal: Applied Economics 2, 210–40.

García-Gómez, P., Jiménez-Martín, S., Castelló, J.V., 2012. Health, disability, and pathways into retirement in spain, in: Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms. University of Chicago Press, pp. 127–174.

Garlick, R., 2018. Academic peer effects with different group assignment policies: Residential tracking versus random assignment. American Economic Journal: Applied Economics 10, 345–69.

Gonzalez, L., Ortega, F., 2013. Immigration and housing booms: Evidence from spain. Journal of Regional Science 53, 37–59.

Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. American Economic Journal: Applied Economics 1, 34–68.

Hoffmann, F., Oreopoulos, P., 2009. A professor like me the influence of instructor gender on college achievement. Journal of Human Resources 44, 479–494.

Hofstede, G., 2001. Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. Sage publications.

Hong, S.C., Lee, J., 2017. Who is sitting next to you? peer effects inside the classroom. Quantitative Economics 8, 239–275.

Hoxby, C., 2000. Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867 .

Hoxby, C.M., Weingarth, G., 2005. Taking race out of the equation: School reassignment and the structure of peer effects. Mimeo .

Jenkins, S.P., Brandolini, A., Micklewright, J., Nolan, B., 2012. The great recession and the distribution of household income. OUP Oxford.

Kimbrough, E.O., McGee, A.D., Shigeoka, H., 2017. How do peers impact learning? an experimental investigation of peer-to-peer teaching and ability tracking. NBER Working Paper No. 23439 .

Kogut, B., Singh, H., 1988. The effect of national culture on the choice of entry mode. Journal of international business studies 19, 411–432.

Lacuesta, A., Izquierdo, M., 2012. The contribution of changes in employment composition and relative returns to the evolution of wage inequality: the case of spain. Journal of Population economics 25, 511–543.

Lambert, P., 2001. The Distribution and Redistribution of Income. Manchester University Press.

Lauridsen, J., Christiansen, T., Häkkinen, U., 2004. Measuring inequality in self-reported health—discussion of a recently suggested approach using finnish data. Health economics 13, 725–732.

Lavy, V., Paserman, M.D., Schlosser, A., 2012a. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. The Economic Journal 122, 208–237.

Lavy, V., Schlosser, A., 2011. Mechanisms and impacts of gender peer effects at school. American Economic Journal: Applied Economics 3, 1–33.

Lavy, V., Silva, O., Weinhardt, F., 2012b. The good, the bad, and the average: Evidence on ability peer effects in schools. Journal of Labor Economics 30, 367–414.

Lazear, E.P., 2001. Educational production. The Quarterly Journal of Economics 116, 777–803.

Lecluyse, A., Cleemput, I., 2006. Making health continuous: implications of different methods on the measurement of inequality. Health economics 15, 99–104.

Lu, F., Anderson, M.L., 2014. Peer effects in microenvironments: The benefits of homogeneous classroom groups. Journal of Labor Economics 33, 91–122.

Lyle, D.S., 2009. The effects of peer group heterogeneity on the production of human capital at west point. American Economic Journal: Applied Economics 1, 69–84.

Mackenbach, J.P., Kunst, A.E., Cavelaars, A.E., Groenhof, F., Geurts, J.J., on Socioeconomic Inequalities in Health, E.W.G., et al., 1997. Socioeconomic inequalities in morbidity and mortality in western europe. The lancet 349, 1655–1659.

Mackenbach, J.P., Stirbu, I., Roskam, A.J.R., Schaap, M.M., Menvielle, G., Leinsalu, M., Kunst, A.E., 2008. Socioeconomic inequalities in health in 22 european countries. New England journal of medicine 358, 2468–2481.

Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. The review of economic studies 60, 531–542.

Marie, O., Zölitz, U., 2017. "high" achievers? cannabis access and academic performance. The Review of Economic Studies 84, 1210–1237.

Marmaros, D., Sacerdote, B., 2006. How do friendships form? The Quarterly Journal of Economics 121, 79–119.

Mayer, A., Puller, S.L., 2008. The old boy (and girl) network: Social network formation on university campuses. Journal of public economics 92, 329–347.

Melitz, J., Toubal, F., 2014. Native language, spoken language, translation and trade. Journal of International Economics 93, 351–363.

Nekby, L., Pettersson-Lidbom, P., 2017. Revisiting the relationship between ethnic diversity and preferences for redistribution: Comment. The Scandinavian Journal of Economics 119, 268–287.

OECD, 2005. Pensions at a Glance: Public policies across OECD countries. OECD.

OECD, 2007. Pensions at a Glance: Public policies across OECD countries. OECD.

OECD, 2008. Growing unequal?: Income distribution and poverty in OECD countries. OECD Publishing.

OECD, 2009. Pensions at a Glance 2009: Retirement-income systems in OECD countries. OECD.

OECD, 2013. Pensions at a Glance 2013: OECD and G20 indicators. OECD.

OECD, 2016a. Health at a Glance: Europe 2016: State of Health in the EU Cycle. OECD.

OECD, 2016b. Youth unemployment rate URL: `https://www.oecd-ilibrary.org/content/data/c3634df7-en`.

Oosterbeek, H., Van Ewijk, R., 2014. Gender peer effects in university: Evidence from a randomized experiment. Economics of Education Review 38, 51–63.

Petrie, D., Allanson, P., Gerdtham, U.G., 2011. Accounting for the dead in the longitudinal analysis of income-related health inequalities. Journal of Health Economics 30, 1113–1123.

Pijoan-Mas, J., Sánchez-Marcos, V., 2010. Spain is different: Falling trends of inequality. Review of Economic Dynamics 13, 154–178.

Plotnick, R., 1981. A measure of horizontal inequity. The review of Economics and Statistics , 283–288.

Puente, S., Galán, S., 2014. Analysis of composition effects on wage behaviour. Economic Bulletin .

Regidor, E., Barrio, G., Bravo, M.J., de la Fuente, L., 2014. Has health in spain been declining since the economic crisis? J Epidemiol Community Health 68, 280–282.

De la Rica, S., Rebollo-Sanz, Y.F., 2017. Gender differentials in unemployment ins and outs during the great recession in spain. De Economist 165, 67–99.

Ruhm, C.J., 2000. Are recessions good for your health? The Quarterly Journal of Economics 115, 617–650.

Ruhm, C.J., 2015. Recessions, healthy no more? Journal of health economics 42, 17–28.

Sacerdote, B., 2001. Peer effects with random assignment: Results for dartmouth roommates. The Quarterly journal of economics 116, 681–704.

Sacerdote, B., 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far?, in: Handbook of the Economics of Education. Elsevier. volume 3, pp. 249–277.

Sacerdote, B., 2014. Experimental and quasi-experimental analysis of peer effects: two steps forward? Annu. Rev. Econ. 6, 253–272.

Schindler, D., Westcott, M., 2018. Shocking racial attitudes: Black gis in europe. CESifo Working Paper Series .

Sen, A., 1979. Equality of what? The Tanner lecture on human values 1.

Steinmayr, A., 2016. Contact matters: Exposure to refugees and voting for the far-right .

Stinebrickner, R., Stinebrickner, T.R., 2006. What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. Journal of public Economics 90, 1435–1454.

Stuckler, D., Basu, S., Suhrcke, M., Coutts, A., McKee, M., 2009. The public health effect of economic crises and alternative policy responses in europe: an empirical analysis. The Lancet 374, 315–323.

Thiemann, P., 2017. The persistent effects of short-term peer groups in higher education. IZA Discussion Paper No. 11024 .

Van Doorslaer, E., Jones, A.M., 2003. Inequalities in self-reported health: validation of a new approach to measurement. Journal of health economics 22, 61–87.

Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U.G., Gerfin, M., Geurts, J., Gross, L., Häkkinen, U., Leu, R.E., et al., 1997. Income-related inequalities in health: some international comparisons. Journal of health economics 16, 93–112.

Van Ourti, T., Van Doorslaer, E., Koolman, X., 2009. The effect of income growth and inequality on health inequality: Theory and empirical evidence from the european panel. Journal of Health Economics 28, 525–539.

Whitehouse, E.R., 2009. Pensions, purchasing-power risk, inflation and indexation .

Ye, J., Han, S., Hu, Y., Coskun, B., Liu, M., Qin, H., Skiena, S., 2017. Nationality classification using name embeddings, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ACM. pp. 1897–1906.

Zimmerman, D.J., 2003. Peer effects in academic outcomes: Evidence from a natural experiment. Review of Economics and statistics 85, 9–23.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

697. M. VAN LENT, *Essays on Intrinsic Motivation of Students and Workers*
698. R.W. POLDERMANS, *Accuracy of Method of Moments Based Inference*
699. J.E. LUSTENHOUWER, *Monetary and Fiscal Policy under Bounded Rationality and Heterogeneous Expectations*
700. W. HUANG, *Trading and Clearing in Modern Times*
701. N. DE GROOT, *Evaluating Labor Market Policy in the Netherlands*
702. R.E.F. VAN MAURIK, *The Economics of Pension Reforms*
703. I. AYDOGAN, *Decisions from Experience and from Description: Beliefs and Probability Weighting*
704. T.B. CHILD, *Political Economy of Development, Conflict, and Business Networks*
705. O. HERLEM, *Three Stories on Influence*
706. J.D. ZHENG, *Social Identity and Social Preferences: An Empirical Exploration*
707. B.A. LOERAKKER, *On the Role of Bonding, Emotional Leadership, and Partner Choice in Games of Cooperation and Conflict*
708. L. ZIEGLER, *Social Networks, Marital Sorting and Job Matching. Three Essays in Labor Economics*
709. M.O. HOYER, *Social Preferences and Emotions in Repeated Interactions*
710. N. GHEBRIHIWET, *Multinational Firms, Technology Transfer, and FDI Policy*
711. H.FANG, *Multivariate Density Forecast Evaluation and Nonparametric Granger Causality Testing*
712. Y. KANTOR, *Urban Form and the Labor Market*
713. R.M. TEULINGS, *Untangling Gravity*
714. K.J.VAN WILGENBURG, *Beliefs, Preferences and Health Insurance Behavior*
715. L. SWART, *Less Now or More Later? Essays on the Measurement of Time Preferences in Economic Experiments*
716. D. NIBBERING, *The Gains from Dimensionality*
717. V. HOORNWEG, *A Tradeoff in Econometrics*
718. S. KUCINSKAS, *Essays in Financial Economics*
719. O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*

720. E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*
721. M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*
722. N. CIURILA, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*
723. N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*
724. S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*
725. S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*
726. Y.ZHU, *On the Effects of CEO Compensation*
727. S. XIA, *Essays on Markets for CEOs and Financial Analysts*
728. I. SAKALAUSKAITE, *Essays on Malpractice in Finance*
729. M.M. GARDBERG, *Financial Integration and Global Imbalances*
730. U. THUMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change,*
731. B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*
732. G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*
733. Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*
734. C.M OOSTERVEEN, *Education Design Matters*
735. S.C BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*
736. S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*
737. M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*
738. M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*
739. R.J. DÖTTLING, *Essays in Financial Economics*
740. E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*
741. Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*
742. A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*
743. B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*
744. R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*
745. P.M. GOLEC, *Essays in Financial Economics*
746. M.N. SOUVERIJN, *Incentives at work*