IFAU
Institute for Evaluation of Labour
Market and Education Policy

Essays on determinants of individual performance and labor market outcomes

Olof Rosenqvist

DISSERTATION SERIES 2016:2
Presented at the Department of Economics, Uppsala University

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P O Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

This doctoral dissertation was defended for the degree of Doctor in Philosophy at the Department of Economics, Uppsala University, September 9, 2016. Essay 2 has been published by IFAU as Working paper 2016:14. Essay 3 has been published by IFAU as Working paper 2016:6. Essay 4 has been published by IFAU as Working paper 2015:28.

ISSN 1651-4149

Dissertation presented at Uppsala University to be publicly examined in Ekonomikum, Hörsal 2, Kyrkogårdsgatan 10, Uppsala, Friday, 9 September 2016 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Ghazala Azmat (Queen Mary University of London).

Abstract

Rosenqvist, O. 2016. Essays on Determinants of Individual Performance and Labor Market Outcomes. *Economic studies* 161. 151 pp. Uppsala: Department of Economics, Uppsala University. ISBN 978-91-85519-68-2.

Essay 1 (with Oskar Nordström Skans): This paper provides field evidence on the causal impact of past successes on future performances. Since persistence in success or failure is likely to be linked through, potentially time-varying, ability it is intrinsically difficult to identify the causal effect of succeeding on the probability of performing well in the future. We therefore employ a regression discontinuity design on data from professional golf tournaments exploiting that almost equally skilled players are separated into successes and failures half-way into the tournaments (the "cut"). We show that players who (marginally) succeeded in making the cut substantially increased their performance in subsequent tournaments relative to players who (marginally) failed to make the cut. This success-effect is substantially larger when the subsequent (outcome) tournament involves more prize money. The results therefore suggest that past successes provide an important prerequisite when performing high-stakes tasks.

Essay 2: Recent experimental evidence suggests that women in general are more discouraged than men by failures which potentially can explain why women, on average, are less likely than men to reach top-positions in firms. This paper provides the first quasi-experimental evidence from the field on this issue using data from all-female and all-male professional golf tournaments to see if this result can be replicated among competitive men and women. These top-performing men and women are active in an environment with multiple rounds of competition and the institutional set-up of the tournaments makes it possible to causally estimate the effect of the result in one tournament on the performance in the next. The results show that both male and female golfers respond negatively to a failure and that their responses are virtually identical. This finding suggests that women's difficulties in reaching top-positions in firms are caused by external rather than internal barriers.

Essay 3: Voting is a fundamental human right. Yet, individuals that are younger than 18 do typically not have this right since they are considered uninformed. However, recent evidence tentatively suggests that the political knowledge of youths is endogenous to the voting age. I test for the existence of such dynamic adjustments utilizing voting age discontinuities caused by Swedish laws. I employ a regression discontinuity strategy on Swedish register data to estimate the causal effect of early age voting right on political knowledge around age 18. The results do not support the existence of positive causal effects of early age voting right on political knowledge. Thus, we should not expect that 16-year-olds respond by acquiring more political knowledge if they are given the right to vote. This finding weakens the case for a lowering of the voting age from 18 to 16.

Essay 4 (with Lena Hensvik): We postulate that firms' production losses from absence depend on the employees' internal substitutability, incentivizing firms to keep absence low in positions with few substitutes. Using Swedish employer-employee data we show that absence is substantially lower in such positions even conditional on establishment and occupation fixed effects. The result reflects sorting on both entry and exit margins, with stronger separations responses when it was difficult to predict the absence of the employees beforehand. These findings highlight that internal substitution insures firms against production disruptions caused by absence and that absence costs are important aspects of firms' hiring and separations decisions.

Keywords: Confidence, success, failure, performance, regression discontinuity design, golf, gender differences, glass ceiling, voting age, political knowledge, civic interest, dynamic effects, sickness absence, production disruption, coworker substitutes, hiring decisions, separations

Olof Rosenqvist, Department of Economics, Box 513, Uppsala University, SE-75120 Uppsala, Sweden.

Acknowledgements

Writing a Ph.D. thesis is very much like playing a round of golf. You start off on the first tee full of confidence and new ideas thinking that this will be the day when everything comes together. But soon you realize that the wind is strong and that your new swing ideas weren't that solid after all. Every hole becomes a struggle and you are forced to hit shots that you didn't know existed. When you finally reach the 18th green you are tired but also stimulated and your thoughts circle around one thing: how can I improve?

Holing out on the 18th green without exceeding the stipulated number of strokes requires that you are surrounded by good people that stop you from going out-of-bounds or into the water. My coach and supervisor Oskar Nordström Skans has done just that. Throughout my doctoral studies he has given me invaluable support through his enthusiastic attitude, his strategic thinking and his honest way of delivering critique. I would like to thank you for making my time as a Ph.D. student so stimulating and exciting.

Many thanks also go to my assistant supervisor Lena Hensvik for keeping my eyes on the ball and giving practical and constructive comments on my ideas and manuscripts. It has been a pleasure doing research with you and you have taught me valuable things about writing and working with data along the way. My gratitude further goes out to my licentiate-opponent Erik Lindqvist and my final seminar discussant Anna Sandberg for giving clear and constructive feedback on the papers included in this thesis. I would also like to thank Uta Schönberg for making it possible for me to stay a semester at University College London. This visit gave interesting insights into the international academic world and provided excellent opportunities for personal development.

I started my time as a Ph.D. student at the Department of Economics at Uppsala University but have spent the bulk of the time at the Institute for Evaluation of Labour Market and Education Policy (IFAU) in Uppsala. I would like to thank the administrative staff at both these organizations for very professional and effective support during the writing of my Ph.D. thesis. Furthermore I would like to thank my first-year companions Daniel, Evelina, George, Gunnar, Jacob, Josefine, Justice, Kristin and Sebastian for all the support and kindness you have given me. At IFAU I am grateful to Olof Åslund for giving me the chance to spend time in an environment with so many experienced researchers, and to all the people there who have been very welcoming. Special thanks go to Björn Ö, Kalle and Martin L who have provided valuable comments on my work and to Ali for our intense table tennis sessions.

I also wish to express my gratitude to friends and family. Thanks to Fredrik, my childhood friend, for always challenging me to go further and to Johan for making my time in Uppsala so enjoyable. I am lucky to have a loving family who has always supported and encouraged my academic ambitions. I would like to thank my parents, Kurre and Liane, for their unreserved love and my two older brothers, Björn and Eric, for paving the way for me and learning me about life.

My thoughts especially go to my beloved fiancée Linda. Ever since we met you have kept my heart warm through your love and my mind sharp through your delicious and nutritious vegan food. Without you as my caddie I wouldn't have made the cut. Thanks also to Signe and Zacke for lightning up my life.

Finally I would like to express my gratitude to the game of golf which over the years has stimulated my private as well as my professional life (you don't say). This great game has made major contributions to this thesis which I wish to repay by spreading the virtues of the game as described in the poem by Paul Bertholy:

What is Happiness?
Golf is happiness for
Happiness is achievement.
The father of achievement is motivation
The mother is encouragement.
The fine golf swing is truly achievement
Man may lie, cheat, and steal for gain.
But, these will never gain the golf swing
To gain the golf swing man must work.
Yet it is work without toil
It is exercise without the boredom.
It is intoxication without the hangover
It is stimulation without the pills.
It is failure yet its successes shine even more brightly
It is frustration yet it nourishes patience.
It irritates yet its soothing is far greater
It is futility yet it nurtures hope.
It is defeating yet it generates courage
It is humbling yet it ennobles the human spirit.
It is dignity yet it rejects arrogance
Its price is high yet its rewards are richer
Some say it's a boy's pastime yet it builds men
It is a buffer for the stresses of today's living.
It cleanses the mind and rejuvenates the body
It is these things and many more.
For those of us who know it and love it
Golf is truly happiness.

Uppsala, June 2016
Olof Rosenqvist

Contents

Introduction

This thesis consists of four self-contained essays. The general purpose of the thesis has been to investigate the extent to which factors beyond ability (intelligence) and education affect individuals' performance and shape their labor market outcomes. Understanding and paying attention to the subtleties that determine our ability to perform complex tasks and that influence our career paths is essential for forming welfare-enhancing policies and for explaining patterns on the labor market. I study three broad and potentially important determinants of performance and labor market outcomes: confidence, motivation and health.

The research on confidence is based on data from professional golf tournaments while the other topics are studied through empirical analyses of Swedish register data with rich information on, among other things, individuals' labor market outcomes.

The first two essays specifically study the causal effect of a previous success, which is assumed to build *confidence*, on current performance among individuals participating in professional golf tournaments. The third essay investigates whether having the right to vote at an early age, which is assumed to increase the *motivation* to acquire political knowledge, improves the high school grade in Social Studies. The fourth essay considers how *health* and especially sickness absence shape labor market outcomes and career opportunities. I discuss these three themes below in turn.

Confidence

It has long been known that human performance is sensitive to emotions and stereotypes (see, e.g., Ellis et al. [1997] and Steele and Aronson [1995]). At the same time, humans are known to exaggerate previous successes and forget previous failures, so that they think of themselves more highly than others do (see, e.g., Guthrie [2001] and Weinstein [1980]). In an interesting theoretical contribution, Compte and Postlewaite (2004) suggest that the latter behavior is an adaption to the former sensitivity. Their argument builds on the notion that confidence enhances performance and that the perceived empirical frequency of success affects confidence levels. Thus, according to Compte and Postlewaite (2004), it can actually be rational and welfare-enhancing to have positively biased recollections of previous performances.

Of course, this reasoning relies on a positive causal relationship going from confidence to performance. Although intuitively reasonable, this relationship has proven hard to show empirically. If confidence is defined by previous successes and failures, there will typically be a positive relationship between confidence and genuine ability since more able individuals more often succeed. Thus, regressing current performance on past outcomes leads to biased estimates, since the underlying ability is correlated with both current performance and past outcomes. Estimating the causal effect of confidence on performance instead requires a situation where ability can be held constant.

Situations like that are rare in naturally occurring settings which, up to this point, has led researchers to investigate the question using laboratory experiments where assignment to success and failure can be manipulated at the will of the researchers. These studies typically find that successful outcomes improve subsequent performance, while failures worsen future performance (see Gill and Prowse [2014] and Bélanger et al. [2013]). But do the results hold true in a real-world situation where the stakes are high?

This is the question that Oskar Nordström Skans and I try to answer in Essay 1 using data from male professional golf tournaments which offer an opportunity to study almost equally skilled players that experience successes and failures. Golf tournaments are played over four days (typically Thursday–Sunday) and 18 holes are played each day. Roughly 140 players participate in the typical tournament. Only players that are tied for 65th place or better after two days are qualified for the remaining two days and all of them will receive at least some prize money. Players who do not qualify must leave the tournament and receive no prize money at all, making participation a financial loss. Players close to the qualification threshold have performed almost equally well (i.e. they have almost the same ability) but will arguably remember the tournament differently in terms of success and failure.

We use a regression discontinuity design to compare the performance of marginally successful and unsuccessful players in the subsequent tournament, which is played the next week. We can thereby remove the confounding influence of underlying ability and estimate the causal effect of success, which is assumed to build confidence, on performance. We find that marginally successful players substantially outperform marginally unsuccessful players in the next tournament. Marginally successful players are about three percentage points more likely to pass the qualification threshold in the next tournament and use 0.25 fewer shots after two rounds. This finding shows that confidence in one's own ability is also crucial for human performance in high stakes real-world situations, which confirms the relevance of the hypothesis that it is rational and welfare-enhancing to have positively biased recollections of previous performances.

In Essay 2, I complement the analysis with more data from female professional golf tournaments. It has been suggested in previous research that women become more discouraged than men by experiencing failures and that this might be one explanation for why women are underrepresented in top-positions in firms and other organizations (see Gill and Prowse [2014]). In competitive environments, individuals are bound to occasionally experience failures and if those failures worsen future performance more for women than for men, then initially unsuccessful women might face problems making future career advancements. Some women might also avoid competition intensive careers altogether if they anticipate these dynamics.

The hypothesis of gender differences in the performance response to previous failure has, however, not been tested using data from real-world situations with repeated competition. In Essay 2, I do just that using data from all-male and all-female professional golf tournaments. Using the same empirical strategy as in Essay 1, I find that the difference in current performance between previously marginally successful and marginally unsuccessful players is virtually the same for men and women. This result, although from a rather particular empirical context, suggests that men and women are actually equally sensitive to previous competitive outcomes, which indicates that gender differences in the ability to cope with failure is an unlikely explanation for the underrepresentation of women in top-positions on the labor market.

Motivation

As economists, we think it is important for policy-makers to consider potential dynamic effects of changes in policies and laws. Individuals might, e.g., adjust their labor supply in response to a change in the income tax, making fiscal calculations based on a static world misleading. Similarly, it is also important to take motivation or incentives into account in areas outside pure economics. In Essay 3, I therefore step into the world of political science to investigate whether a lower voting age can have dynamic effects with respect to the level of political knowledge among young people, i.e. is the level of political knowledge at age 16 greater if an individual can vote at age 16 rather than at age 18?

This piece of research is motivated in part by the recent popular debate in many western countries, including Sweden, about the appropriate voting age, and also by two studies that have examined the level of political knowledge for different age groups in the UK and Austria. The voting age in the UK is 18, and that has been the case for several decades, whereas the voting age in Austria was lowered to 16 in 2007. The study carried out in the UK (Chan and Clayton [2006]) used data from 1991 and 2001 and investigated the level of political knowledge among individuals who were at least 16 years of age. They found a clear age gradient in political knowledge, with 16-year-olds exhibiting the worst results. Based on this result, the researchers suggested that the voting age should not be lowered to 16 in the UK. Note that this conclusion was reached under a voting age regime of 18. The study carried out in Austria (Wagner, Johann and Kritzinger [2012]) used data from 2009 and performed a similar analysis. Contrary to the first study, they found that 16-year-olds possessed virtually the same level of political knowledge as their somewhat older fellow citizens and based on that finding, they concluded that a voting age of 16 was reasonable. Wagner, Johann and Kritzinger (2012) also suggested that their study had a substantial advantage compared to Chan and Clayton (2006), since their conclusion was reached under a voting age regime of 16. They argue that this is important because 16-year-olds might become more politically knowledgeable simply because they are given the right to vote.

But the diverging results in Chan and Clayton (2006) and Wagner, Johann and Kritzinger (2012) could also potentially be explained by cross-country differences other than the voting age between the UK and Austria, such as the school system. In Essay 3, I therefore investigate the hypothesis that having the right to vote increases the motivation to become politically knowledgeable in a more controlled framework. Conceptually, I try to answer the following question: Consider a pair of twins, where one of the twins can vote at 16 and the other at 18. Will the twin who can vote at 16 exhibit higher levels of political knowledge at age 16 than the other twin, since he or she has stronger incentives to learn about the political system and the different political alternatives? Empirically, I answer this question by comparing the high school grade in Social Studies of individuals who turn 18 just before and just after a major election in Sweden. This allows me to estimate the causal effect of early age voting right on levels of political knowledge around age 18. I find that individuals turning 18 just before and just after a major election in Sweden exhibit virtually identical levels of political knowledge around or shortly after age 18. Even though I study 18-year-olds rather than 16-year-olds, this result generally indicates that having the right to vote does not make young individuals more motivated to learn about politics and society.

Thus, while lowering the voting age to 16 might be reasonable based on general human rights considerations, we should not expect 16-year-olds to become more engaged in and more knowledgeable about society just because they are given the right to vote.

Health

Sickness absence rates are rising in Sweden and in many other European countries. While high absence rates and associated sickness insurance expenditures are generally acknowledged as a major national fiscal problem that needs to be tackled, less focus has been given to the individual consequences of sickness absence. By definition, sickness absence is of course associated with different degrees of suffering for the sick individual, but can it also lead to adverse consequences on the labor market, such as decreased access to certain types of jobs?

This is the question that Lena Hensvik and I study in Essay 4. Our starting point is the idea that there are complementarities between job characteristics and worker characteristics that can lead to match-specific gains in productivity (see Sattinger [1975] and Tinbergen [1956] for the origin of this idea). A substantial part of the research in labor economics is devoted to how firms and workers search for their optimal partners and how this affects wages and job separation patterns. While the focus of this research has been on complementarities between workers' skills and the skill requirement of different jobs (see, e.g., Abowd et al. [2007], Andersson et. al [2009] and Lazear [2009]), little is known about how other worker characteristics, such as health, go along with different types of jobs.

In Essay 4, we test the novel hypothesis that there is a positive association between individuals' sickness absence probabilities and their degree of substitutability on the workplace. The fundamental idea is that firms are reluctant to hire individuals with high sickness absence probability for jobs that require skills that few other employees on the workplace possess. Absence from those kinds of jobs can lead to major production disruptions for a firm, since nobody on the workplace can fill the position and carry out the associated tasks. Conceptually, think of a workplace in which all employees are dependent on IT services, but that can only afford to employ one IT technician. If the IT technician is absent, the rest of the employees may face problems performing their work.

We measure sickness absence by a dummy for receiving sickness pay (i.e. being absent more than 14 days) and substitutability by the number of other employees in the same combination of workplace and occupation (i.e. potential *substitutes*).

We reach four major conclusions. First, in the cross-section, it is clear that having few substitutes is associated with low sickness absence. Second, *new hires* in jobs with relatively few substitutes had significantly lower *pre-hire* sickness absence than new hires in jobs with a relatively high number of substitutes. Third, by following individuals over time, we can see that the same individual has significantly lower sickness absence when he or she works in a job with few substitutes than when he or she works in a job with a relatively high number of substitutes. Fourth, exhibiting sickness absence in a job with few substitutes is more strongly associated with job separation than exhibiting sickness absence in a job with a relatively high number of substitutes.

The results are consistent with our hypothesis that there should be a positive association between individuals' sickness absence probabilities and their degree of substitutability in the workplace. The results also show that firms use three different strategies to achieve an allocation where low-absence workers work in low substitutability jobs and vice versa: pre-hire screening, on-the-job pressure and post-hire adjustments. Thus, despite the fact that an individual's probability of being sick is arguably much harder to observe for an employer than, for example, formal credentials (such as education level) it is somehow observed, or at least approximated, and it blocks individuals from entering certain jobs. However, more research is needed to find out exactly how employers are able to assess the sickness absence probabilities of job seekers.

Concluding comments

Overall, the findings in this thesis support the idea that factors other than ability and education substantially affect individuals' performance and labor market outcomes. The results suggest that confidence is crucially important when performing complex high stakes tasks, while a good and stable health appears to be of major importance for getting access to positions in firms where absence potentially can lead to major production disruptions.

The positive causal relation between confidence and performance that was established in Essay 1 and Essay 2 is of particular interest from a social inequality perspective since initial confidence has been found to be positively correlated with socioeconomic background (see Twenge and Cambell [2002]). Confidence might therefore be a mechanism that widens the performance and income gap between individuals from advantaged respectively disadvantaged socioeconomic backgrounds and that consequently contributes to intergenerational rigidity in socioeconomic mobility.

The fact that the effects of earlier successes (which are assumed to build confidence) on current performance appear very similar for men and women further indicates that career differences between men and women, especially with respect to representation in top positions in the labor market, are not caused by gender differences in the psychological sensitivity to previous competitive outcomes which Gill and Prowse (2014) speculated.

The thesis also present novel results on sickness absence which should be considered a central labor market phenomenon. And, quite surprisingly, economists have paid little attention to how sickness absence affect firms' production processes and to what extent varying costs of production disruptions caused by absence affect hiring, firing and absence rates. The results clearly indicate that the interaction between production disruption costs and sickness absence are of crucial importance for understanding the relationship between health and labor market outcomes. Notably, it is well established in the literature that sickness absence is decreasing in socioeconomic status (see, e.g., Fuhrer et al. [2002]). Given the adverse consequences of exhibiting sickness absence for labor market outcomes that were documented in Essay 4 it is thus also troubling, from an equality perspective, that individuals from poor backgrounds are particularly susceptible to spells of sickness absence and, as the results indicate, may thus also be excluded from the parts of the labor market where absence is more costly.

All the hypotheses that are tested in this thesis are, however, not supported by the data. The analysis of the interaction between voting right and political knowledge in Essay 3 suggests that young people do not rise to the occasion and become more politically knowledgeable just because they are given the right to vote. Thus, while motivation is a key driver for human behavior in many aspects of life, having the right to vote does clearly not generate sufficient motivation to learn more about politics. This is an important piece of evidence when considering the potential effects of lowering the voting age.

References

Abowd, J. M., J. Haltiwanger, J. Lane, K. L. McKinsey, and K. Sandusky. 2007. "Technology and the demand for skill: an analysis of within and between firm differences". NBER Working Paper 13043.

Andersson, F., M. Freedman, J. Haltiwanger, J. Lane and K. Shaw. 2009. "Reaching for the stars: who pays for talent in innovative industries?". *The Economic Journal*, 119(538), F308–F332.

Bélanger, J. J., M. K. Lafrenière., R. J. Vallerand, and A. W. Kruglanski. 2013. "Driven by fear: the effect of success and failure information on passionate individuals' performance". *Journal of Personality and Social Psychology*, 104(1), 180–195.

Chan, T. W. and M. Clayton. 2006. "Should the voting age be lowered to sixteen? Normative and empirical considerations". *Political Studies*, 54(3), 533–558.

Compte, O. and A. Postlewaite. 2004. "Confidence-enhanced performance". *American Economic Review*, 94(5), 1536–1557.

Ellis, H. C., S. A. Ottaway, L. J. Varner, A. S. Becker, and B. A. Moore. 1997. "Emotion, motivation, and text comprehension: the detection of contradictions in passages". *Journal of Experimental Psychology*, 126(2), 131–46.

Fuhrer, R., M. J. Shipley, J. F. Chastang, A. Schmaus, I. Niedhammer, S. A. Stansfeld, M. Goldberg, and M. G. Marmot. 2002. "Socioeconomic position, health, and possible explanations: a tale of two cohorts". *American Journal of Public Health*, 92(8), 1290–1294.

Gill, D. and V. Prowse. 2014. "Gender differences and dynamics in competition: the role of luck". *Quantitative Economics*, 5(2), 351–376.

Guthrie, C., J. Rachlinski, and A. Wistrich. 2001. "Inside the judicial mind: heuristics and biases". *Cornell Law Review*, 86(4), 777–830.

Lazear, E. (2009). "Firm-specific human capital: a skill-weights approach". *Journal of Political Economy*, 117(5), 914–940.

Sattinger, M. (1975). "Comparative advantage and the distribution of earnings and abilities". *Econometrica*, 43, 455–468.

Steele, C. M. and J. Aronson. 1995. “Stereotype threat and the intellectual test performance of african-americans”. *Journal of Personality and Social Psychology*, 69(5), 797–811.

Tinbergen, J. (1956). “On the theory of income distribution”. *Weltwirtschaftliches Archiv*, 77, 156–173.

Twenge, J. M. and W. K. Cambell. 2002. “Self-esteem and socioeconomic status: a meta-analytic review”. *Personality and Social Psychology Review*, 6(1), 59–71.

Wagner, M., D. Johann, and S. Kritzinger. 2012. “Voting at 16: turnout and the quality of vote choice”. *Electoral studies*, 31(2), 372–383.

Weinstein, N. D. 1980. “Unrealistic optimism about future life events”. *Journal of Personality and Social Psychology*, 39(5), 806–20.

Essay 1: Confidence enhanced performance? – The causal effect of success on future performance in professional golf tournaments[1]

(With Oskar Nordström Skans)

[1] This essay is published in *Journal of Economic Behavior & Organization*, Volume 117, September 2015, Pages 281–295.

1 Introduction

The determinants of performance in different aspects of life constitute a key element within economics. As a consequence, economists have devoted considerable effort to quantifying the extent to which factors such as education, experience, effort, cognitive and non-cognitive skills, beauty and height affect performance in various contexts such as schools and labor markets. One possible mechanism through which many of these factors may affect the ability to perform demanding or complex tasks is by altering the individual's own perception of the ability to perform these tasks, i.e. through building confidence. One specific version of this mechanism is the idea of "hot-hand" effects in the world of competitive sports. In this paper we provide quasi-experimental field evidence on the effects of past successes on future performance.

A starting point for our analysis is provided by Compte and Postlewaite (2004) who postulated that there may be causal link from past successes on future performance through "confidence", defined as the perception of previous performances. Conceptually, consider two identically skilled surgeons performing identical surgeries where one patient dies and the other one survives due to random chance. Arguably, the surgeon whose patient died will think of the event as a failure, whereas the other will think of it as a success. The question is whether this difference in perception will have a causal impact on their performances in the future. There is so far very little credible field evidence on the empirical relevance of this fundamental idea. A likely reason for the scarce set of previous evidence on the issue is that it is inherently difficult to analyze since those that perform well today also may perform well tomorrow, and vice versa, simply because more able individuals are likely to persistently perform better than less able individuals.

In this paper we use data from professional golf tournaments, relying on a special feature in these tournaments enabling us to perform a regression discontinuity (RD) analysis to identify the causal impact of succeeding in one tournament on the performance in the next.[2] Midway through most of the professional golf tournaments on the European PGA Tour there is a qualification threshold (the cut) that a player must pass (make) in order to complete the tournament and earn prize money. Players around the threshold have performed almost equally well but, arguably, when the players look back on their performance, those who barely passed will have a perception of success whereas those who barely failed will have a perception of failure. By study-

[2] In using data from the world of sports to study fundamental economic processes we follow in a long line of previous papers. Ehrenberg and Bognanno (1990a and 1990b), Orszag (1994) and Melton and Zorn (2000) all use data from golf tournaments to study the predictions of tournament theories and Pope and Schweitzer (2011) use data from golf tournaments to measure loss aversion.

ing the results in the subsequent tournament we are able to isolate the effect of present success on future performances.[3]

Empirically, we are perhaps closest in spirit to the study of "hot-hand" effects within sports economics, i.e. the notion that athletes' performances during certain (short) periods are significantly better than otherwise; see e.g. Livingston (2012) and Clark III (2005) for two examples using data from golf tournaments, Abrevaya (2002) for an example using data from bowling tournaments, Crust and Nesti (2006) for a wider review of the phenomenon and Rabin and Vayanos (2010) for a theoretical discussion. We do, however, provide two novelties relative to the existing hot-hand literature. Firstly, we look at a longer time interval (a week) between subsequent performances and thus we are able to study effects that go beyond the immediate state of mind. Secondly, instead of only following the performance of an individual player over time we can control for the counterfactual development (i.e. the player in a failure state) and thereby reduce the confounding impact of exogenous time-varying but persistent factors that may affect the players. Thus, we believe that we are more likely to be able to detect a causal "hot-hand" relationship than previous studies.[4]

The previous literature also contains a number of studies where the assignment to success and failure is manipulated in a controlled laboratory setting. Most of these studies find that (perceived) successes tends to lead to improved subsequent performance while (perceived) failures deteriorate future performance, see, e.g. Gill and Prowse (2012, 2014) and Bélanger et al. (2013) for a recent review. A number of relatively recent studies have also examined the importance of *relative* feedback information on subsequent performance, finding mixed results. Azmat and Iriberri (2010) and Tran and Zeckhauser (2012) both report that when individuals in a group receive information about their relative performance ranking in the group, the group as a whole performs better in the future. Since they find improvements in subsequent performance over the whole distribution, the results suggest that relative feedback information spurs a general will to compete. Eriksson et al. (2009), on the other hand, find no general effect from relative performance feedback but instead document a reduction in the quality of low performers' work. Along the same lines Murphy and Weinhardt (2013) find that students in English primary schools are sensitive to their local rank. They compare equally able students (same score on a standardized national test at the end of primary school) that differ with respect to their local rank. Students in the top of the performance distribution at their school outperform

---

[3] Making the cut improves the ranking of a player compared to failing to make the cut; but this ranking has, unlike in some other sports (e.g. tennis), no direct benefits in the upcoming tournament.

[4] See Wardrop (1999), Frame et al (2003) and Bar-Eli et al. (2006) for discussions regarding difficulties when attempting to estimate hot-hand effects.

students with poor local rank in secondary school even though they had the same test score.[5]

A key difference between these studies and our set-up is that we isolate the impact of success in a setting where the subjects are fully informed regarding the underlying process, once success or failure has been determined. Players know how many strokes they used, what their final rank became, and whether they passed the cut or not. Thus, scoring one stroke less to pass the cut is not more informative regarding the own ability than scoring one stroke less to get closer to (or further from) the cut.[6]

To preview our results, we show that players just above and below the cut indeed are comparable in terms of predetermined characteristics allowing us to infer the causal impact of the initial success on future outcomes. We further show that making the cut in a tournament has a large and statistically significant positive causal effect on the outcomes in the subsequent tournament. The number of strokes after two rounds falls by a quarter of a stroke and the probability of making the cut in the following tournament increases by 3 percentage points from a baseline of 50 percent. This is a sizable effect in relation to the importance of other variables in the data such as years of professional experience and the average score per round the previous year.

An important caveat in terms of the interpretation is that players who pass the cut earn prize money and the opportunity to continue playing competitive golf during two more days in the initial ("treatment") tournament. In order to tentatively investigate if these factors contribute to (or mitigate) the effects we find, we have analyzed if the effects vary depending on the stakes (prize money) in the treatment tournament and subsequent ("outcome") tournament played during the week after. The results show that the effects are independent of the prize money involved in the treatment tournament, suggesting that (at least the magnitudes of) financial rewards are unlikely to be the key determinant of our main results. On the other hand, we do find that the benefits of past successes are confined to the cases when the stakes in the outcome tournament are high, which we interpret as consistent with the notion of a psychological transmission mechanism.[7] We do, however, acknowledge that we are unable to, with certainty, rule out the possibility that either earned prize money or the two days of additional competitive experience to some

---

[5] A related literature studies the importance of "stereotype effects", where subjects' performance on tests are found to be affected by information regarding the average performance among people with similar characteristics as themselves, see e.g. Cadinu et al. (2003), Cadinu et al. (2005) and Aronson et al. (1999).

[6] Notably, Compte and Postlewaite (2004) show that it can indeed be rational to have biased recollections of previous performances as long as performance is directly affected by (the perception of) previous performances.

[7] Making the cut can also give a player additional positive publicity and a feeling of being on a lucky streak; we interpret any positive effects this entails as being part of the overall impact of confidence.

extent contribute to, or attenuate, the estimated overall effect of past successes on future performance within professional golf tournaments.

The rest of the paper is structured as follows. In section 2 we present a stylized model and the key elements of our empirical analysis. Section 3 explains the empirical setting and describes the data. Section 4 deals with the validity of the empirical strategy, shows the baseline results and analyses the role of the involved stakes. Section 5 concludes.

2 Empirical strategy

In this section we outline a stylized model of performance and confidence. The purpose of the model is to highlight the logic of our identification strategy and to provide a useful framework for discussing threats to our key identifying assumptions.

2.1 A stylized model

We borrow our basic set-up from Compte and Postlewaite (2004) who put the notion that confidence can affect performance into a formal model. However, since our objective is different from theirs, we reformulate their model to make it richer in details that are relevant for our setting, but instead ignore other aspects. Compte and Postlewaite build their model around two broad ideas. The first is that individuals find it easier to recollect good memories (successes) than bad memories (failures) and that individuals tend to attribute successes to their own performance and failures to various exogenous circumstances. The second notion is that emotions and stereotypes affect performance. Building on these ideas, they construct a model where performance is a function of confidence and confidence in turn is a function of the "perceived empirical frequency of success" and show that it is welfare enhancing to have a positively biased perception of previous performances if confidence affects performance.

Following Compte and Postlewaite (2004) we assume that higher confidence (C) will increase performance. For practical reasons we formulate increased performance as a reduction in the number of mistakes (Z). In addition, we assume that individuals differ in their abilities (A) and that the frequency of mistakes has a random orthogonal component (ε) with mean zero. Hence, consider the following two-period model of mistakes:

$$Z_{it} = \alpha_t + \beta A_i + \gamma C_{it} + \varepsilon_{it} \tag{1}$$

where t = 1, 2. Since Z is the number of mistakes, we expect both β and γ to be negative. It should be evident that the model includes two distinct factors (ability and confidence), each of which can generate an inter-temporal correlation in the number of mistakes. Since both of these factors are unobserved, it is intrinsically difficult to isolate their empirical relevance. However, empirical identification is possible in settings where performance is continuous, but confidence is built as a result of discrete successful events. Conceptually, such settings (could) include cases such as when a player makes (or misses) the golf cut, a surgeon's patient survives (or dies) during surgery or a student passes (or fails) an exam. To mimic these settings, we assume that performance leads to success ($S = 1$) when the number of mistakes underscores a threshold (T), but leads to failure ($S = 0$) otherwise. Thus:

$$S_{it} = I(Z_{it} \leq T_t) = I(\alpha_t + \beta A_i + \gamma C_{it} + \varepsilon_{it} \leq T_t) \quad (2)$$

where $I(.)$ is an indicator function taking the value 1 if the argument is true and zero otherwise. Compte and Postlewaite (2004) assume that positive recollections increase the agents' level of confidence in the next period. In our setting, we assume that these effects arise from past successes, i.e. we expect μ in equation (3) to be positive and that confidence evolves according to:[8]

$$C_{i2} = C_{i1} + \mu S_{i1} - \mu/2 \quad (3)$$

Hence, performance in the second period is a function of ability, initial confidence and an indicator for succeeding in the first period:

$$Z_{i2} = \alpha_2 + \beta A_i + \gamma C_{i1} + \gamma\mu I(\alpha_1 + \beta A_i + \gamma C_{i1} + \varepsilon_{i1} \leq T_1) - \gamma\mu/2 + \varepsilon_{i2} \quad (4)$$

Three things are worth noting from equation (4). First, it implies that initial confidence (C_{i1}) affects the performance in the second period. Second, since the success function (2) propagates the impact of ability, the unconditional returns to ability are higher in the second period than in the first period due to the confidence effect. Finally, note that transitory lucky circumstances in the first period (i.e. ε_{it}) improve performance in the second period.

2.2 Empirical specification

For empirical purposes it is convenient to rewrite the model as:

$$Z_{i2} = Z_{i1} + (\alpha_2 - \alpha_1) + \gamma\mu I(Z_{i1} \leq T_1) - \gamma\mu/2 + (\varepsilon_{i2} - \varepsilon_{i1}) \quad (5)$$

[8] Assuming that the probability of success is one half, as in our empirical application, this formulation ensures that the average confidence level is stable.

With this structure it is evident that we can think of the first period as the *treatment* period, and the second period as the *outcome* period, where the (potential) success in the treatment period is allowed to have a causal impact on the performance in the outcome period corresponding to $\gamma\mu$. Conditional on previous mistakes, the impact of making the threshold captures the causal effect of succeeding. This argument highlights our general strategy for separating the impact of ability from the impact of success. However, instead of relying on the linear functional form for identification, we focus the identification on agents closest to the threshold using a *Regression Discontinuity* (RD) approach (see e.g. Lee and Lemieux, 2010 for a detailed discussion).[9]

Conceptually, the RD-design serves as a comparison of outcomes between players that are placed infinitely close to a threshold on the two sides. In practice, all RD-designs use data from a data window around the threshold to generate (and compare) separate predictions from each side of the threshold. The differences between these predictions provide estimates of the causal impact of passing the threshold regardless of the nature and functional forms of other confounding factors, as long as these are smoothly distributed. Hence, the RD model is identified even if (e.g.) A and C are not linearly separable (see e.g. Hahn et al, 2001). In our empirical analysis, we follow common practice and identify the effect of success on performance by running a pooled regression with separate linear terms on each side of the threshold within a specific data window (which is to be varied across specifications). Formally, we estimate:

$$Z_{i2} = \beta_0 + \beta_1 I[Z_{i1} \leq T_1] + \beta_2 (Z_{i1} - T_1) + \beta_3 I[Z_{i1} \leq T_1](Z_{i1} - T_1) + u_{i2} \tag{6}$$

where β_1 denotes the impact of past successes. Note that estimates from equation (6) can be given a causal interpretation under any continuous functional form. The key assumption is continuity in the underlying confounders, which requires that there is no perfect sorting around this threshold (see e.g. Lee and Lemieux, 2010). In the empirical section below, we show a number of tests of this assumption.

3 Institutions and data

3.1 European professional golf tournaments

Our paper relies on data from professional golf tournaments. These tournaments are well-suited for the RD approach outlined above since they use a

[9] See also Thistlethwaite and Campbell (1960) for the original application of a regression discontinuity design.

cut midway through which effectively sorts players into elimination or qualification based on their number of strokes. Players close to the threshold thus performed almost equally well, but some failed and others succeeded.

The two major male professional golf tours in the world are the U.S. PGA Tour and the European PGA Tour. In this paper we employ data from the male European PGA Tour. [10] Every year about 50 tournaments are played on this tour; that is, virtually every week a tournament is being played. Despite the name of the Tour, a fair number of tournaments are played outside of Europe; destinations include South Africa, New Zealand, Australia and China. The typical tournament is played during four days (from Thursday to Sunday) with 18 holes each day. Based on the players' score after two days (36 holes) a qualification score called the "cut" is determined. Among the (typical) set of 140 players who start the tournament, the cut allows those who reach the 65th place (70th place before 2006) or better to continue playing over the weekend and earn prize money (the amount depend on their final position) whereas the other players are eliminated from the tournament without financial rewards.

We find it reasonable to assume that making the cut can be described as a success, and failing to make the cut as a failure. There will only be a single-stroke difference between players just making the 36-hole cut and players just failing to make it; hence only one stroke separates a success from a failure. The cut thus provides us with a situation where individuals who performed almost equally well are split into successes and failures.

3.2 Data

Our data are drawn from tournaments between the start of the 2000 season and the 1st of April 2012. Using these data we define empirical counterparts to the theoretical model outlined in section 2. The *number of strokes* after two days of playing serves as the empirical measure of the number of mistakes (Z). The *cut,* separating players after two days, corresponds to the threshold T. Thus, success S, for a player, is defined as the case when the number of strokes is at least as low as the cut in the relevant tournament.
To mimic the two-period structure of the model, we use data on the number of strokes (after two days) in a tournament as our measure of period-1 performance. We use data from the tournament played the week after as our measure of the outcome in period 2. An observation in our data thus pertains to a *tournament pair* consisting of the number of strokes in a *treatment* (period 1) tournament and the number of strokes in an *outcome* (period 2) tournament.

For each treatment tournament we have collected data on players within a six stroke window (on either side) of the cut. Throughout, we use the nor-

[10] www.europeantour.com

malized scores (corresponding to $Z_{i\tau} - T_\tau$) ranging from -5 to 6 at most. In addition to the number of strokes/mistakes (Z), we use the probability of passing the cut in the outcome tournament as an alternative measure of performance.

Two potential concerns arise because of the institutional set-up. The first is that we need to handle the fact that thresholds are tournament specific. In order to address this, we enrich equation (6) by including tournament specific fixed effects (δ_τ) for each pair (τ) of a treatment and an outcome tournament. Thus, using Y to denote outcomes (number of strokes, or a dummy for passing the cut, in the outcome tournament), S to denote success ($I[Z_{i1} \leq T_1]$, i.e. a dummy for making the cut) and X to denote the normalized running variable ($Z_{i\tau} - T_\tau$) our final empirical model can be written as:[11]

$$Y_{i\tau} = \beta_0 + \beta_1 S_{i\tau} + \beta_2 X_{i\tau} + \beta_3 S_{i\tau} X_{i\tau} + \delta_\tau + u_{i\tau} \quad (7)$$

The second possible concern is that not all players from the treatment tournament participate in the outcome tournament. To minimize our exposure to selective dropouts, we focus the analysis on tournament pairs where the participation rate in the outcome tournament is at least 60 %.[12] We further address the issue of potential systematic selection into outcome tournaments in a set of robustness checks presented below.

3.3 Descriptive statistics

Table 1 provides descriptive statistics for the widest used sample of 16,515 observations.[13] Note that players with a result equal to the cut also pass the cut, which explain why the normalized strokes on average have a one stroke lower absolute value amongst those passing the cut. A crude comparison between successful (made the cut) and unsuccessful (did not make the cut) players reveals that successful players have fewer strokes (and make the cut to a greater extent) also in the outcome tournaments. This is natural even without a confidence effect since we expect the successful players to be inherently better than their unsuccessful counterparts, i.e. good players persistently perform well and bad players persistently perform badly (A in the model). The difference in ability between the two groups is also reflected in the fact that successful players had a lower stroke average in the previous year.

[11] We also present results from models where we allow tournament specific slopes above and below the thresholds.
[12] Our data are collected manually from the tour web page. A detailed protocol for the tournament selection process is available in Appendix A. All the micro data and a complete list of tournaments are available on demand.
[13] This includes the full set of players within a 6 stroke window of the cut except for those (around 5,000) players who did not participate in the adjacent outcome tournament.

Table 1. Descriptive statistics for the used sample

	All	Made the cut	Did not make the cut
Treatment tournaments			
Average cut (T)	144.0	144.0	144.0
Normalized strokes ($Z - T$)	0.249	-2.053	3.023
… standard deviation	(3.015)	(1.638)	(1.651)
Made the cut ($I[Z \leq T]$)	0.546	1	0
Outcome tournaments			
Average cut	143.6	143.6	143.6
Strokes relative to the cut	0.323	-0.189	0.941
… standard deviation	(4.519)	(4.445)	(4.527)
Made the cut	0.549	0.594	0.495
Player characteristics			
Years as pro	11.2	11.2	11.2
…standard deviation	6.44	6.43	6.67
…nonmissing	0.954	0.962	0.944
Stroke average in previous year	72.0	71.9	72.1
…standard deviation	(1.14)	(1.09)	(1.19)
…nonmissing	0.978	0.982	0.973
Made the cut in previous tournament	0.554	0.578	0.525
…nonmissing	0.876	0.881	0.870
Number of tournaments	189	189	189
Number of clusters (tournament/strokes)	2,257	1,132	1,125
Number of observations	16,515	9,025	7,490

Note. Sample includes players within 6 strokes of the cut, i.e. with strokes relative to the cut in the [-5,6] interval. The cut is defined by the maximum number of strokes allowed for players to proceed in the tournament. This implies that values of strokes relative to the cut will be one stroke higher on average (in absolute values) among those not passing the cut if the stroke distribution is symmetric around the cut.

4 Validity and results

In this section we first assess the robustness of our empirical strategy through a number of validity checks and then turn to our main results and robustness checks of these. Finally, we present a few extensions.

4.1 Validity of the empirical strategy

We are primarily concerned with two issues which potentially could confound the causal interpretation of our final estimates. Firstly, since it is important for players to make the cut, players could potentially push themselves to exactly match the threshold in order to end up on the right side of the cut. If that happens, players' abilities may not be smoothly distributed across the threshold which would invalidate our identifying assumptions.

Secondly, since not all players participate in the relevant outcome tournament, we also need to worry about biased attrition. The key assumption in this dimension is that the ability of those that leave the sample is uncorrelated with whether they made the cut or not, conditional on the statistical model. It could be noted that the two concerns are related in the sense that they both refer to the possibility that the ability of players within the *used sample* jumps discontinuously at the threshold (the cut).

Figure 1 addresses the first concern. The left-side panel shows the distribution of players around the cut in the treatment tournaments. Naturally, there are a lot of observations close to the threshold since this is defined by the median score. This implies that many players just barely made the cut. As shown in the right-side panel, however, the relative difference in the number of observations between scores 1 and 0 fall well within the normal range.[14] The figures thus imply that there are roughly as many players that just miss the cut as there are players that just make the cut which is reassuringly consistent with the assumption of smoothness across the threshold.

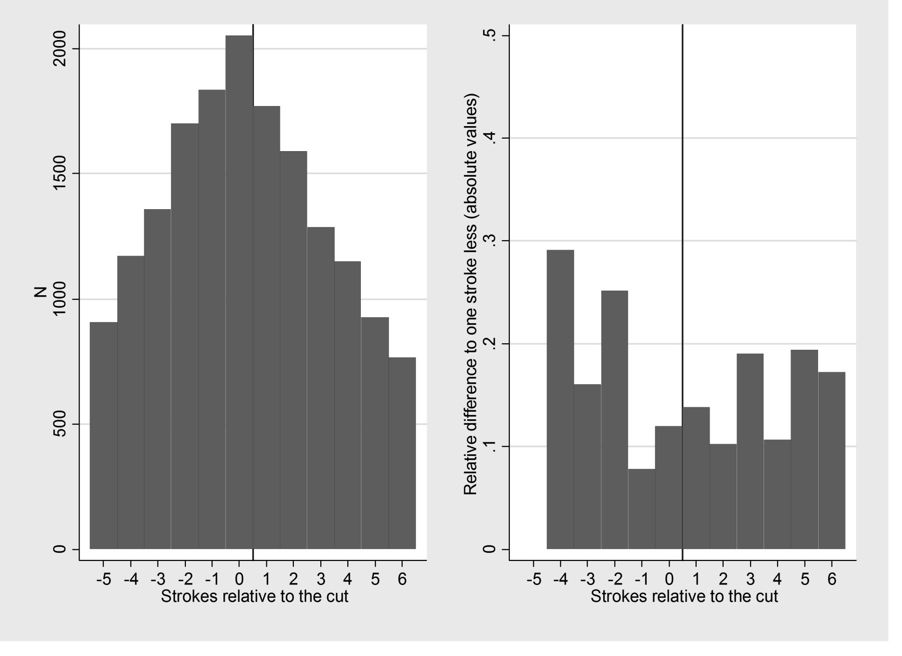


Figure 1. Score density on either side of the threshold, and the distribution of relative differences relative to one stroke less.

Figure 2 addresses the second concern. The figure clearly shows that there is systematic selection into the outcome tournament amongst the poor perform-

[14] This impression also holds if we focus on absolute rather than relative differences, despite the obvious bias towards larger numbers in the center of the distribution in that case.

ing players, but this selection is only related to the average score, and not related to whether or not the player succeeded in making the treatment tournament cut or not. The gradually lower participation rate among relatively poorly performing players is most likely due to the fact that some tournaments invite a few local players to attend through special invitations. On average, these players are likely to perform worse and they do not always have the right to attend the outcome tournament (we return to this issue in a robustness check in section 4.1). Leaving the general trend aside, we conclude that the participation rate seems to evolve smoothly over the cutoff.[15] Thus, there is no clear evidence that attrition at the threshold is a problem within our overall sample. Nevertheless, we perform various robustness checks in the empirical section to verify that the results are robust if we zoom in on players with a higher-than-average propensity to participate in the outcome tournaments.

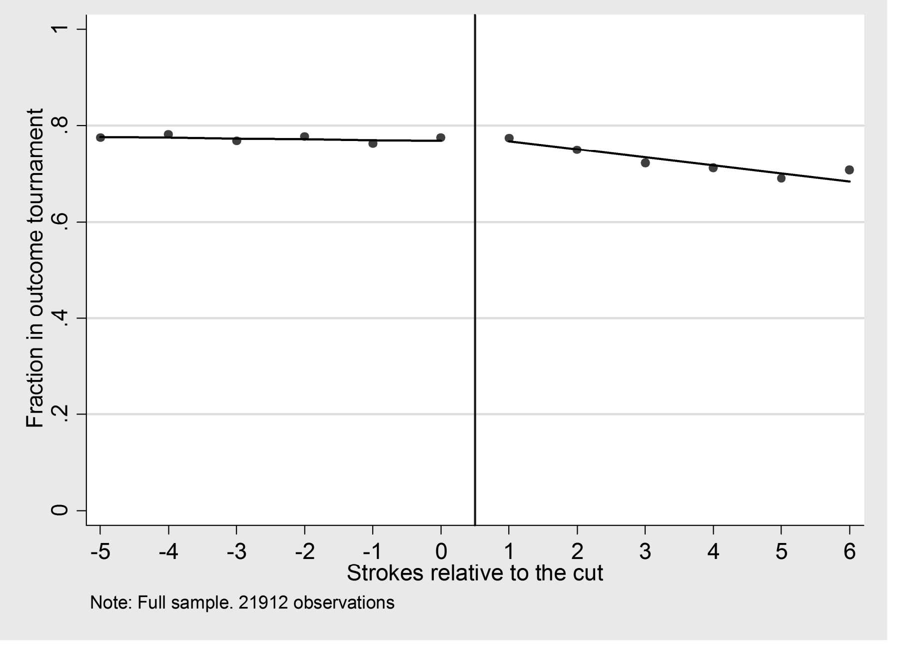


Figure 2. Fraction playing in the outcome tournament.

As a further specification test, we expect the distribution of predetermined individual characteristics to evolve smoothly across the threshold. Discontinuous jumps in characteristics between treated and untreated players at the cut would constitute a violation of our main identifying assumption of a smooth evolution of abilities across the threshold and therefore also a threat

[15] As should be evident from the figure, the jump at the threshold is not statistically significant.

to the causal interpretation of our main estimates. In Figure 3, we investigate the concern using data on the average scores-per-round during the preceding year and the number of years as a pro (experience) within the sample of players actually participating in an outcome tournament (i.e. within our used sample). The left-hand side panel of Figure 3 shows that poorly performing players had a higher stroke average in the preceding year. Reassuringly, however, there is no sign of a discontinuous jump at the threshold.[16] The right-hand side panel of Figure 3 shows that a corresponding description for experience, measured by the number of years as a professional player. The results imply that experience evolves smoothly across the threshold. It should also be noted that the relationship we see in the figure is relatively weak. The standard deviation in years as pro is 6.5 years (see Table 1 above) and the high- and low-points of the figure are only separated by 0.4 years.

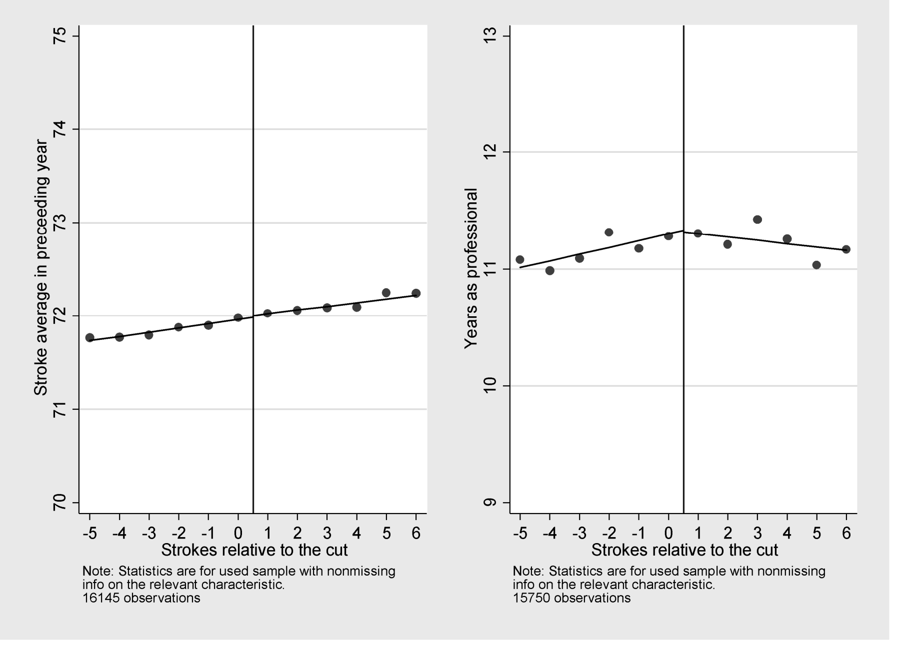


Figure 3. Stroke average during the preceding year by strokes relative to the cut (left panel) and years as a professional player by strokes relative to the cut (right panel).

In Table 2 we present statistical balancing tests where we assert that the pre-determined covariates are uncorrelated with making the cut, conditional on the statistical model (i.e. that they evolve smoothly around the cut). The table shows the results for the widest and for the most narrow sample win-

[16] Interpreting the stroke average during the preceding year as an indicator of ability, the picture is very much in line with our stylized model which presumes that ability matters linearly.

dows (*bandwidths* in standard RD-terminology) for which we can estimate the model (i.e. 2 to 6 strokes on either side of the cut). The table shows the balancing of the stroke average during the preceding year, the number of years as pro and a dummy for making the cut in the last previous tournament the player participated in. The insignificant results imply that we cannot reject the hypothesis that they evolve smoothly across the threshold. In addition, we show that the joint performance prediction (predicted probability of passing the cut) based on these variables does not jump at the threshold. Finally, we display the p-values from F-tests showing that the variables are jointly insignificant in predicting the threshold.

Table 2. Validity tests – do baseline covariates jump at the threshold?

Outcome:	Stroke average last year	Years as pro	Made the cut in previous tournament	Joint ability score: The predicted probability of passing the cut based on Years as pro, Stroke average and Making the cut in the previous tournament
Panel A (full sample)	[-5,6]	[-5,6]	[-5,6]	[-5,6]
Jump at threshold	-0.0234	-0.0149	-0.0132	0.000939
	(0.0322)	(0.192)	(0.0150)	(0.00160)
N	16,145	15,750	14,466	16,515
	p-value for test of joint significance: 0.79			
Panel B (small sample)	[-1,2]	[-1,2]	[-1,2]	[-1,2]
Jump at threshold	-0.0315	-0.0507	-0.0195	0.00165
	(0.0614)	(0.356)	(0.0272)	(0.00302)
N	7,079	6,914	6,350	7,246
	p-value for test of joint significance: 0.72			
Controls for assignment variable (strokes from cut):				
Linear	Yes	Yes	Yes	Yes
By treatment status	Yes	Yes	Yes	Yes

Note: Panel A includes the full data window with strokes in the range [-5,6]. Panel B only uses players within two strokes of the treatment tournament cut, i.e. within the range [-1,2]. Data include the sample with non-missing information on the outcomes. Tournament specific controls for the assignment variable are interacted with treatment status by tournament. Joint ability score in the last column is the predicted ability to pass the cut from the three variables in the three preceding columns and corresponding dummies for missing values on these. P-value for joint significance is for the test that the three variables and their corresponding dummies for missing information jointly predict the threshold. Standard errors are clustered at the strokes times tournament level. */**/*** significant at the 10 /5/1 percent level. Insignificant results imply that we cannot reject that covariates are balanced across the threshold conditional on the RD-model.

In summary, these validity checks are fully consistent with the notion of a smooth distribution of confounding factors around the threshold. Notably, these tests are run on the used data which implies that they should detect systematic sorting around the threshold in terms of the strokes within the treatment tournaments as well as biased attrition before the outcome tournaments.

4.2 Main results

Turning to the main results, we start by showing the results in terms of figures, and then move on to a more formal analysis. We start by analyzing the effect on the probability of making the cut in the outcome tournament. This effect is displayed in Figure 4. As a contrast, the figure also shows the probability of having made the cut in the tournament before (thus, in effect an additional specification test). If the assignment to treatment is truly exogenous we should not see any effect on the rate of lagged successes whereas if the treatment (success) really matters we are to see an effect on the success rate in the outcome tournament. Clearly, Figure 4 supports both the empirical strategy and the theoretical prediction since players around the threshold are similar with respect to previous performance (left-side panel), but differ with respect to subsequent performance (right-side panel). In essence, Figure 4 tells us that the treatment effect observed in the right-side panel is not due to the fact that some players consistently (marginally) make the cut while some players consistently (marginally) fail to make the cut. Instead, marginally making the cut appears to generate a genuine performance improvement as compared to marginally failing to make the same cut. The right-side panel suggests that the event of making the cut in the treatment tournament increases the probability of making the cut by about 3 percentage points (corresponding regression estimates are found in the table below). Considering a baseline probability of around 50 percent, we interpret this as a fairly substantial effect.

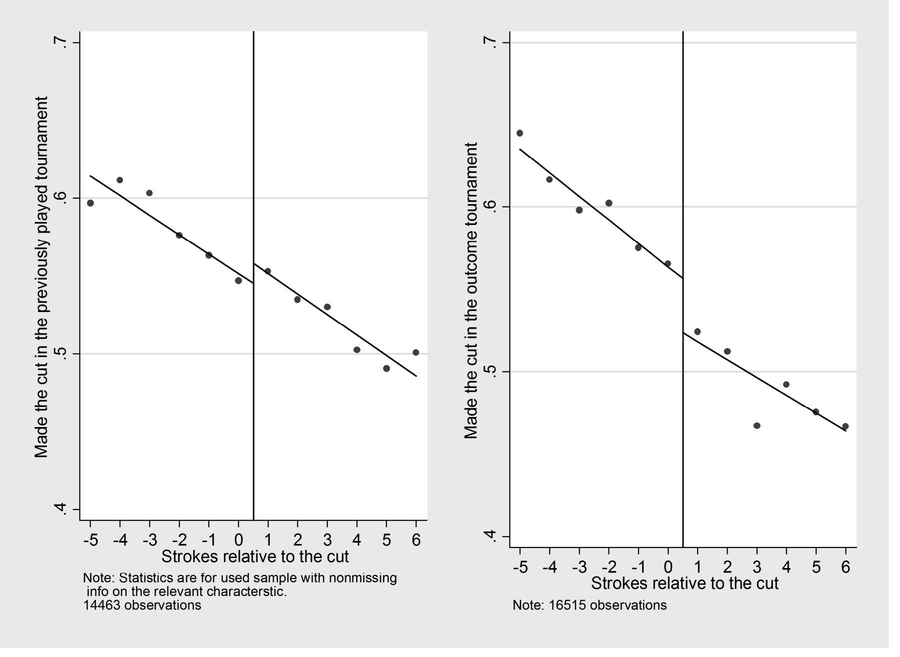


Figure 4. Probability of passing the cut in the previous tournament (specification test, left-side panel) and in the outcome tournament (right-side panel) by strokes relative to the cut in the treatment tournament.

Our main outcome is the number of shots after two rounds in the outcome tournament normalized by the subtraction of the cut in the relevant tournament (i.e. when the cut is set). This continuous performance measure corresponds closely to the index of mistakes (Z) in the theoretical model. The results for this outcome are shown in Figure 5.[17] Clearly, the more shots a player had in the treatment tournament the more shots he had in the outcome tournament (again, roughly consistent with a linear, time constant, impact of ability A). More importantly, however, there is a visible performance gap at the threshold. Those marginally making the cut outperformed those marginally failing by about a quarter of a shot after two rounds in the outcome tournament (after accounting for the linear terms).[18] The magnitude of the effect corresponds to one quarter of a standard deviation (see Table 1) across play-

[17] We have verified that making the cut does not have a positive relationship to the number of strokes in the previous tournament (corresponding to the making the cut analysis in the left side panel of figure 5). However, since these strokes come from different tournaments for different players (we do not want to restrict the sample to those playing in three subsequent tournaments), we have chosen not to include the results in the paper. They are available on request, however.
[18] We have verified that the effect is smooth across the (outcome) stroke distribution. Thus, the effect on strokes is not confined to performances around the cut in the outcome tournament.

ers in yearly stroke averages or 5 percent of a standard deviation of the (noisy) outcome tournament performance.

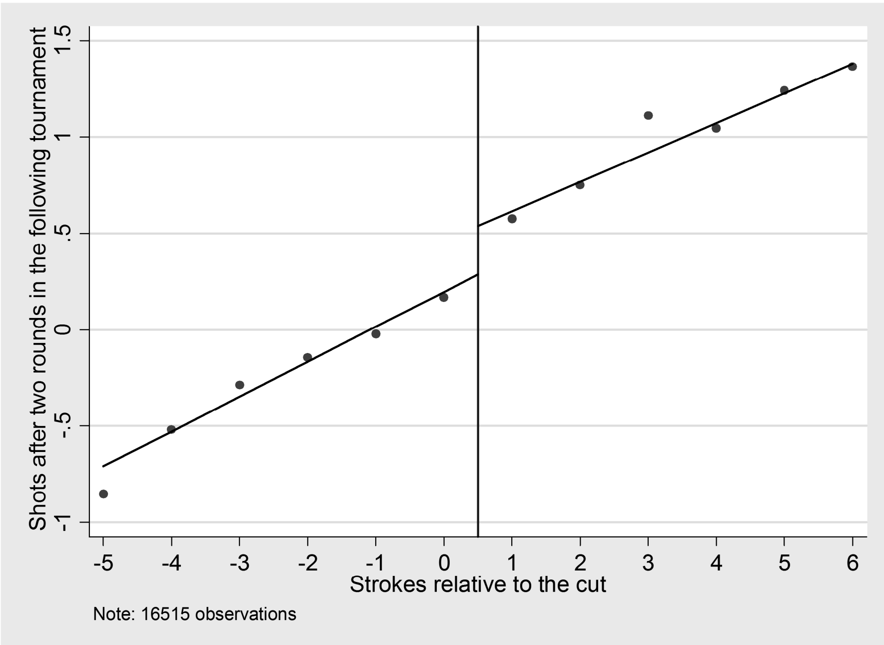


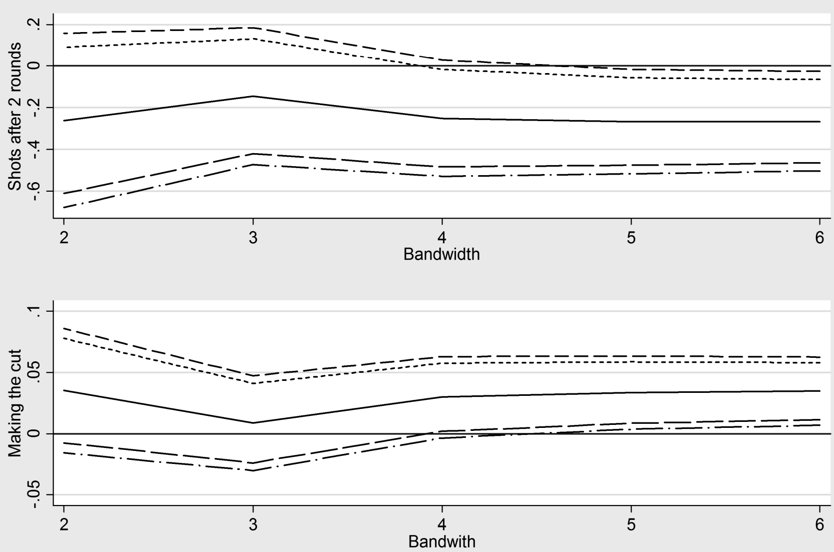
Figure 5. The impact on strokes in the outcome tournament by strokes relative to the cut in the treatment tournament.

It is notable that the slope displayed in Figure 5 implies that a player who plays one more stroke during the initial two rounds of the treatment tournament does on average only need about 0.2 strokes more to complete two rounds in the outcome tournament. This indicates that the ability-based persistence in performance across tournaments is fairly low, or in other words that there is substantial randomness in the tournament-specific performance. This randomness is positive for our identification strategy since it means that the assignment variable has a very large random component which reduces the risk of ability-based sorting exactly at the threshold.

In Figure 6 and Table 3 we show a large set of variations of the formal model. Throughout, all these variations, tournament fixed effects are included and all standard errors are clustered at the strokes-times-tournament level and robust to heteroscedasticity following Lee and Card (2008).[19] In the figure, we vary the number of observations included in the analysis. We let the data window (the bandwidth) move from two (which is the lowest number

[19] The clustering is motivated by a joint specification error for each stroke-group. See also Lee and Lemieux (2010) for a discussion. Standard errors are nearly identical if clustering on tournaments instead.

possible for estimating the model) to six strokes on either side of the cut. As is evident, the point estimates are centered around 0.03 and –0.25 respectively. The estimates are statistically significant at bandwidths above four. Varying the bandwidth means trading off precision (which is lost when going too far towards the cut) against heavy reliance on the linear functional form of the controls for treatment-tournament strokes (when moving away from the threshold). Overall we find it reassuring that the point estimates do not show any clear relationship to the size of the bandwidth (the one caveat being that the estimates at bandwidth 3 are somewhat muted) even though precision, as expected, becomes an issue when slicing the data too thin.



Note: Estimates with 90 and 95 percent confidence intervals for different sample windows around the cut (bandwidths). Number of observations varies between 16,515 and 7,246.

Figure 6. Estimated impact on strokes in the outcome tournament and probability of making the cut by size of sample window (bandwidth).

In Table 3, we show estimates for further variations of the model. Specification (1) imposes the same linear term above and below the threshold. Specification (2) shows the estimates corresponding to Figure 6 (i.e. with separate trends above and below the threshold). The fact that the results in specification (1) and (2) are close to identical reflects the fact that the impact of strokes in the treatment tournament is fairly linear across the whole sample window. Specification (3) adds controls for our baseline covariates (strokes in the previous year, years as pro and making the cut in the previous tournament) and the results remain stable. Specification (4) zooms in on the two closest observations (i.e. one stroke below and one stroke above the cut) without controlling for the assignment variable and as expected (from not controlling for treatment-tournament strokes) the estimates become somewhat larger. Specification (5) shows estimates when controlling a higher order polynomial in the number of strokes (quadratic) instead of the separate terms above and below the threshold. Specification (6) instead allows the linear controls (above and below) the cut to be specific for each tournament. Throughout, the point estimates remain reasonably stable.

In Panel B (specifications 7-12), we repeat these exercises for the "making the cut" outcome, again with very similar results.

In Panel C (specifications 13-18) we introduce a number of alternative variations. Since all of these are quite demanding in terms of data, we rely on the model with the highest statistical power (i.e. with the largest data window and controls for baseline covariates) as the starting point. The first three variations address the potential concern that our results are driven by selective attrition. To investigate this issue we have re-estimated the model for samples of players who are "stable" participants. Specification (13) only uses players who participate in at least 70 percent of outcome tournaments during their career (conditional on being in the preceding treatment tournament). Specification (14) raises the bar to 80 percent. Specification (15) instead focuses on players who participate in at least 25 tournament-outcome pairs during our sample period in order to remove the influence of the local players discussed in section 4.1. Throughout, the estimates appear robust, although as expected with lower precision as we remove about a third of the sample in the most restrictive variation.

Specification (16) instead controls for player fixed effects (alongside the tournament effects and the RD-controls) in order to remove any potential for unobserved ability differences through sorting or selective attrition around the threshold. Specification (17) becomes even more stringent as we let the individual fixed effects be separate for each season while still keeping all the other controls. Again, the results appear reasonably stable. Finally, specification (18) shows estimates when only including tournaments played during the first half of the year. The reason for this exercise is the potential concern that players, during the final tournaments of the season, may behave differently because of concerns related to end-of-season rankings. The results are, however, similar to those of the main model.

Panel D repeats the exercises of Panel C (i.e. "stable players", "fixed effects" and "early tournaments") for the binary outcome "making the cut". The results remain as robust as when we used the continuous outcome, with one exception. The one case where the results do not hold up is when we use season-by-player fixed effects, where we instead find an estimate close to zero. Since the results do hold up nicely when using player fixed effects (i.e. allowing for variation across seasons as well), we interpret the zero estimate as most likely being due to the fact that there is little variation left in the dependent variable when combining (perhaps too) high-dimensional fixed effects with the RD set-up using a binary outcome variable.

Overall, we interpret the results as jointly suggesting that making the cut does have a positive effect on the players' performance in next week's tournament. We interpret the estimated magnitudes as reasonably stable, considering the wide set of variations we expose the model to. Our preferred estimates show a positive effect of around a quarter of a shot in the outcome tournament, or, alternatively, an increase in the probability of passing the cut in the outcome tournament by 3 percentage points.

Table 3. Main results – the impact of success on performance

Panel A	*Strokes*					
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	-0.246**	-0.265**	-0.254**	-0.359***	-0.244**	-0.255**
Strokes	(0.118)	(0.121)	(0.119)	(0.0971)	(0.116)	(0.111)
N	16,515	16,515	16,515	3,823	16,515	16,515
Panel B	*Making the cut*					
Specification:	(7)	(8)	(9)	(10)	(11)	(12)
Outcome:	0.0323**	0.0347**	0.0338**	0.0365***	0.0324**	0.0351***
Making the cut	(0.0139)	(0.0143)	(0.0141)	(0.0124)	(0.0137)	(0.0132)
N	16,515	16,515	16,515	3,823	16,515	16,515
Controls for assignment variable (strokes from cut):						
Sample window	[-5,6]	[-5,6]	[-5,6]	[0,1]	[-5,6]	[-5,6]
Linear	Yes	Yes	Yes	No	Yes	Yes
By treatment status	No	Yes	Yes	No	No	Yes
Quadratic	No	No	No	No	Yes	
By tournament	No	No	No	No	No	Yes
Covariates	No	No	Yes	Yes	Yes	Yes
Panel C	*Strokes*					
Specification:	(13)	(14)	(15)	(16)	(17)	(18)
Outcome:	-0.266**	-0.235	-0.292**	-0.196*	-0.229*	-0.236
Strokes	(0.126)	(0.150)	(0.138)	(0.118)	(0.120)	(0.153)
N	14,337	10,660	12,313	16,515	16,515	10,386
Panel D	*Making the cut*					
Specification:	(19)	(20)	(21)	(22)	(23)	(24)
Outcome:	0.0312**	0.0241	0.0463***	0.0262*	-0.00756	0.0301*
Making the cut	(0.0152)	(0.0175)	(0.0167)	(0.0142)	(0.0145)	(0.0179)
N	14,337	10,660	12,313	16,515	16,515	10,386
Controls for assignment variable (strokes from cut):						
Sample window	[-5,6]	[-5,6]	[-5,6]	[-5,6]	[-5,6]	[-5,6]
Linear	Yes	Yes	Yes	Yes	Yes	Yes
By treatment status	Yes	Yes	Yes	Yes	Yes	Yes
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Variations						
Participation rate	>70	>80	-	-	-	-
# Tournaments	-	-	>25	-	-	-
First half	-	-	-	-	-	Yes
Fixed effects						
Player	-	-	-	Yes	-	-
Player by season	-	-	-	-	Yes	-

Note: Covariates are: Years as pro-player, Stroke average during the previous year and Passing the cut in the previous tournament, as well as a set of indicator variables for missing values on each of these variables. Tournament specific controls for the assignment variable are interacted with treatment status by tournament. Participation rate (in outcome tournament if participating in the treatment tournament) is calculated by player during the full sample period. # Tournaments is the number of tournaments by player during the full sample period. First half indicate that the tournament is played on the first half of the year. Standard errors are clustered at the strokes times tournament level. */**/*** significant at the 10 /5/1 percent level.

4.3 Confidence effects for different types of players

We have also explored the extent to which the effects differ depending on the characteristics of the players. We characterize players in three dimensions using a dummy for experienced players (years as pro above the median), a dummy for stroke average above median in the previous year and a dummy indicating if they managed to pass the cut in the last tournament they played before the treatment tournament. For each of these dummy variables, we display estimates for players having the values one and zero respectively, and also display the difference with standard errors. These estimates are derived from expanded versions of the model shown in equation 7 (as used in Specifications 3 and 9 of Table 3). The expansions build on interactions between the main variables (treatment status S, normalized strokes X and the interaction of these) and the dummy for the relevant characteristic (D_i^{Char}). Formally:

$$Y_{i\tau} = \beta_0 + \beta_1 S_{i\tau} + \beta_2 X_{i\tau} + \beta_3 S_{i\tau} X_{i\tau} + \beta_4 D_i^{Char} S_{i\tau} + \beta_5 D_i^{Char} X_{i\tau} + \beta_6 D_i^{Char} S_{i\tau} X_{i\tau} + \delta_\tau + u_{i\tau} \quad (8)$$

The results are displayed in Table 4 with indicators for the baseline groups of weak players (inexperienced, high stroke average and missed the previous cut) on the top row. For ease of interpretation, the table shows the estimated effect for the baseline group ($\hat{\beta}_1$), the estimated effect for the alternative group ($\hat{\beta}_1 + \hat{\beta}_4$) and the estimated difference between these ($\hat{\beta}_4$). The point estimates suggest that the effects are more pronounced for the relatively stronger players. Precision is, however, clearly an issue for this exercise and only the differences between players with a high and a low stroke average is statistically significant.

Although the lack of precision suggests that we should interpret these estimates with great care, they do seem to imply that the effect of past successes, if anything, is larger when the initial endowment of ability (and confidence) also is large. We have also estimated the model, splitting the sample instead according to the predicted performance from a regression of all the variables on the probability of passing the cut (i.e. the same variable as in the final column of Table 2), and the results (not in the table) suggest that the positive effect is restricted to the most well performing half of the players.[20] If this indeed is the case, then it reinforces the distributional impact of confidence effects that was implied by the stylized theoretical model where endowments and effects were assumed to be additively separable.

[20] The results with shots as the outcome are -.49 with p-value 0.001 for the high ability group and 0.06 with p-value 0.69 for the low ability group. Using the cut-dummy as the outcome we get 0.066 (p-value 0.000) and 0.004 (0.79) respectively. Differences are significant in both cases.

Table 4. Heterogeneous effects – individual characteristics

Years as pro		Stroke average last year		Previous tournament cut	
Outcome: Strokes					
Inexperienced	-0.150	High	0.0725	Missed	-0.220
	(0.154)		(0.145)		(0.201)
Experienced	-0.264*	Low	-0.594***	Made	-0.371**
	(0.147)		(0.146)		(0.176)
Difference	0.114	Difference	0.667***	Difference	0.151
	(0.176)		(0.159)		(0.279)
N	15,751		16,145		14,466
Outcome: Made the cut					
Inexperienced	0.026	High	0.015	Missed	0.021
	(0.018)		(0.017)		(0.023)
Experienced	0.037**	Low	0.056***	Made	0.057***
	(0.017)		(0.017)		(0.021)
Difference	-0.011	Difference	-0.041**	Difference	-0.037
	(0.020)		(0.018)		(0.032)
N	15,751		16,145		14,466
Controls for assignment variable (strokes from cut):					
Linear	Yes		Yes		Yes
By treatment status	Yes		Yes		Yes
By characteristic	Yes		Yes		Yes
Covariates	Yes		Yes		Yes

Note: Model is as Table 2, Column (3) except for interaction terms. Assignment (separately above and below the threshold) and treatment are interacted with the displayed characteristics. Stroke average during the year before and years of pro are split by the median within the sample. Covariates are: Years as pro-player, Stroke average during the previous year and Passing the cut in the previous tournament, as well as a set of indicator variables for missing values on each of these variables. Tournament specific controls for the assignment variable are interacted with treatment status by tournament. Standard errors are clustered at the strokes times tournament level. */**/*** significant at the 10 /5/1 percent level.

4.4 The role of tournament stakes

A potential concern regarding the interpretation of our main estimate is that passing the cut also entails prize money, which could possibly have an independent effect on the outcome. This effect could either be positive, if money is an input in the performance technology,[21] or negative, if the returns to financial rewards are diminishing and the receipt of money therefore reduces

[21] A player with more money might be able to afford better travel arrangements, better accommodation at the next tournament and more trainer-led practice sessions affecting the preconditions to perform well.

ambition. In addition, passing the cut allows the player to continue playing for two additional days which, potentially, could have an additional effect on the performance in the subsequent tournament. This effect could also, conceptually, be either positive (through training) or negative (through fatigue). Although it is not clear whether these mechanisms should entail a positive or a negative bias to the estimates, they remain as a source of uncertainty regarding the exact interpretation of the estimates. In order to tentatively investigate their potential importance, we use data on the prize sums in the studied tournaments.

We separate between high- and low-stakes tournaments by defining a *high stakes* (HS) tournament as a tournament in which the prize money is larger than the median prize money for that season. We repeat this categorization separately for treatment and outcome tournaments. Our first conjecture is that we should see larger effects for high stakes treatment tournaments if prize money had an independent positive effect on the outcomes.

A complication is that the stakes in the treatment and outcome tournaments are correlated. Thus, we need to analyze the impact of treatments within a joint framework. To this end, we proceed in two steps. We first estimate a separate regression (corresponding to equation 7) for each tournament pair. In a second step, we use the 189 tournament specific estimates as outcome variables in regressions where we explain the estimated effects with two indicator variables, one capturing high treatment-stakes and one capturing high outcome-stakes.[22] The results are shown in Table 5. The results first show that the stakes in the treatment tournament is unrelated to the effect of interest, which implies that the magnitudes of the prize sums are unrelated to the magnitude of the effects, which, in our view, makes financial rewards a less likely candidate as a main driver of the overall effects. (In terms of equation 3, the results imply that μ is independent of the financial rewards in the treatment tournament).

However, the effects are found to be substantially larger for *outcome* tournaments with high stakes. This suggests that past successes are particularly important in high-pressure situations. (In terms of equation 1, γ is an increasing function of the stakes) We also find this result reassuring regarding our preferred interpretation on the main effect (although other interpretations remain possible), since we conjecture that the impact of additional practice during the continued treatment tournament should be relatively independent of the stakes in the outcome tournament, whereas it seems more likely that the role of confidence is more pronounced when stakes are high.

[22] Notably, the average of the underlying estimates of this analysis are closely related to column 6 of Table 3 since we use tournament-specific assignment-variable effects in both cases.

Table 5. Confidence effects and tournament stakes

	(1)	(2)	(3)	(4)
	Strokes		*Making the cut*	
HS Treatment	0.227	0.123	-0.007	0.014
	(0.290)	(0.406)	(0.036)	(0.050)
HS Outcome	-0.802***	-0.920**	0.113***	0.137**
	(0.289)	(0.432)	(0.036)	(0.053)
Interaction		0.213		-0.044
		(0.582)		(0.072)
Mean dep v.	-0.307	-0.307	0.040	0.040
N	189	189	189	189

Note: HS denotes dummies for "high stakes" tournaments, i.e. tournaments with above-median prize sums. Covariates are: Years as pro-player, Stroke average during the previous year and Passing the cut in the previous tournament, as well as a set of indicator variables for missing values on each of these variables. Tournament specific controls for the assignment variable are interacted with treatment status by tournament. Standard errors are clustered at the strokes times tournament level. */**/*** significant at the 10 /5/1 percent level.

5 Conclusions

In this paper we have used data from professional golf tournaments to analyze the joint conjecture that success breeds confidence and that confidence in turn influences future performance. Using an RD design, we are able to isolate the causal impact of present success on future performance, and our results indicate that these effects are indeed substantial. Passing the cut in one tournament is found to causally increase the probability of doing so in the next tournament as well by 3 percentage points from a baseline of about 50 percent. A special feature of the setting we are analyzing is that players who succeed in making the cut also gain price money and the opportunity to play competitive golf during two additional days. Although we are unable to present a definite test that would completely exclude the possibility that these factors are important for the causal link we observe we do, however, show that previous success matters the most when a player is competing in a high stakes outcome tournament, suggesting that the success effect primarily works through a confidence mechanism. In addition, we show that the effects are of a similar magnitude if the success in the initial tournament brings a relatively large or small financial reward, implying that at least the size of the prize sum is unrelated to the magnitude of the effect.

As we show in our very stylized model, a performance-propagating role of confidence may result in far-reaching consequences for the relationship between ability and performance, since both initial confidence and early luck will affect future performance. It is noteworthy that the literature suggests that initial confidence tends to be positively correlated with social background (see e.g. Twenge and Cambell, 2002, for a review), which indicates that confidence may contribute to observed intergenerational rigidities in social mobility.[23] Furthermore, confidence provides a mechanism through which the importance of ability may grow over time, thus providing a novel rationale for growing returns to ability over the life course.[24] Given these potentially far-reaching consequences of confidence effects as modelled in this paper, we believe that our results from the world of professional golf tournaments call for more research on the nature and determinants of a causal relationship between past successes and future performance in a variety of settings.

References

Abrevaya, J. 2002. "Ladder tournaments and underdogs: lessons from professional bowling". *Journal of Economic Behavior & Organization*, 47(1), 87–101.

Altonji, J. G. and C. R. Pierret. 2001. "Employer learning and statistical discrimination". *Quarterly Journal of Economics*, 116(1), 313–350.

Aronson, J., M. J. Lustina, C. Good, and K. Keough. 1999. "When white men can't do math: necessary and sufficient factors in stereotype threat". *Journal of Experimental Social Psychology*, 35(1), 29–46.

Azmat, G. and N. Iriberri. 2010. "The importance of relative performance feedback information: evidence from a natural experiment using high school students". *Journal of Public Economics*, 94(7–8), 435–452.

Bar-Eli, M., S. Avugos, and M. Raab. 2006. "Twenty years of "hot hand" research: review and critique". *Psychology of Sport and Exercise,* 7(6), 525–553.

Bélanger, J. J., M. K. Lafrenière., R. J. Vallerand, and A. W. Kruglanski. 2013. "Driven by fear: the effect of success and failure information on passionate individuals' performance". *Journal of Personality and Social Psychology*, 104(1), 180–195.

Cadinu, M., A. Maass, A. Rosabianca, and J. Kiesner. 2005. "Why do women underperform under stereotype threat?: evidence for the role of negative thinking". *Psychological Science*, 16(7), 572–578.

[23] See, e.g., Solon (1992).

[24] See ,e.g., Altonji and Pierret (2001) regarding how returns to cognitive skills change with labor market experience and Lindqvist and Vestman (2011) for a discussion on the wage returns to cognitive and non-cognitive abilities.

Cadinu, M., A. Maass, S. Frigerio, L. Impagliazzo, and S. Latinotti. 2003. "Stereotype threat: the effect of expectancy on performance". *European Journal of Social Psychology*, 33(2), 267–285.

Clark III, R. 2005. "An examination of the "hot hand" in professional golfers". *Perceptual and Motor Skills*, 101(3), 935–942.

Compte, O. and A. Postlewaite. 2004. "Confidence-enhanced performance". *American Economic Review*, 94(5), 1536–1557.

Crust, L. and M. Nesti. (2006). "A review of psychological momentum in sports: why qualitative research is needed". *Athletic Insight*, 8(1), 1–15.

Ehrenberg, R. G. and M. L. Bognanno. 1990a. "The incentive effects of tournaments revisited: evidence from the European PGA Tour". *Industrial & Labor Relations Review*, 43(3), 74S–88S.

Ehrenberg, R. G. and M. L. Bognanno. 1990b. "Do tournaments have incentive effects?". *Journal of Political Economy*, 98(6), 1307–1324.

Eriksson, T., A. Poulsen, and M. C. Villeval. 2009. "Feedback and incentives: experimental evidence". *Labour Economics*, 16(6), 679–688.

Frame, D., E. Hughson, and J. C. Leach. 2003. "Runs, regimes, and rationality: the hot hand strikes back". Unpublished manuscript, University of Colorado, Leeds School of Business.

Gill, D. and V. Prowse. 2012. "A structural analysis of disappointment aversion in a real effort competition". *American Economic Review*, 102(1), 469–503.

Gill, D. and V. Prowse. 2014. "Gender differences and dynamics in competition: the role of luck". *Quantitative Economics*, 5(2), 351–376.

Hahn, J., P. Todd, and W. Van der Klaauw. 2001. "Identification and estimation of treatment effects with a regression-discontinuity design". *Econometrica*, 69(1), 201–209.

Lee, D. S. and D. Card. 2008. "Regression discontinuity inference with specification error". *Journal of Econometrics*, 142(2), 655–674.

Lee, D. S. and T. Lemieux. 2010. "Regression discontinuity designs in economics". *Journal of Economic Literature*, 48(2), 281–355.

Lindqvist, E. and R. Vestman. 2011. "The labor market returns to cognitive and noncognitive ability: evidence from the swedish enlistment". *American Economic Journal: Applied Economics*, 3(1), 101–128.

Livingston, J. A. 2012. "The hot hand and the cold hand in professional golf". *Journal of Economic Behavior & Organization*, 81(1), 172–184.

Melton, M. and T. S. Zorn. 2000. "An empirical test of tournament theory: the Senior PGA Tour", *Managerial Finance*, 26(7), 16–32.

Murphy, R. and F. Weinhardt. 2013. “The importance of rank position”. Centre for Economic Performance, Discussion Paper No 1241.

Orszag, J. M. 1994. “A new look at incentive effects and golf tournaments”. *Economics Letters*, 46(1), 77–88.

Pope, D. G. and M. E. Schweitzer. 2011. “Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes”. *American Economic Review*, 101(1), 129–157.

Rabin, M. and D. Vayanos. 2010. “The gambler’s and hot-hand fallacies: theory and applications”. *Review of Economic Studies*, 77(2), 730–778.

Solon, G. 1992. “Intergenerational income mobility in the United States”. *American Economic Review*, 82(3), 393–408.

Thistlethwaite, D. L. and D. T. Campbell. 1960. “Regression-discontinuity analysis: an alternative to the ex post facto experiment”. *The Journal of Educational Psychology*, 51(6), 309–317.

Tran, A. and R. Zeckhauser. 2012. “Rank as an inherent incentive: Evidence from a field experiment”. *Journal of Public Economics*, 96(9-10), 645–650.

Twenge, J. M. and W. K. Cambell. 2002. “Self-esteem and socioeconomic status: a meta-analytic review”. *Personality and Social Psychology Review*, 6(1), 59–71.

Wardrop, R. L. 1999. “Statistical tests for the hot-hand in basketball in a controlled setting”. *American Statistician*, 1, 1–20.

Appendix A: Protocol for selection of tournaments

1. We first collect data from a given "treatment" tournament. We then follow the players and record their results in the subsequent, "outcome" tournament. The subsequent tournament is the tournament played during the weekend directly following the treatment tournament.

2. We have collected data from tournaments during the time period between the start of the 2000 season to the 1st of April 2012.

3. All players participating in the treatment tournament will not participate in the outcome tournament. To minimize selection problems, we place 5 restrictions on our used data (labeled A to E below). Data for tournaments excluded through restrictions A to D have not been collected at all.

A) The treatment tournament and the outcome tournament should be played in the same geographical region. The relevant regions are Europe (including Morocco), South Africa, Australia and New Zealand, United Arab Emirates Qatar and Bahrain, India, China and Hong Kong, Thailand Malaysia Indonesia and Singapore, Russia, South Korea.

B) The treatment tournament and the outcome tournament should both be "Normal" tournaments. This excludes Majors, "World tournaments", Match tournaments, Team tournaments, Low status tournaments, tournaments without cuts and Alfred Dunhill Links Championship (played every fall in Scotland). Tournaments that are **Match**, **Team** and **No cut** type of tournaments cannot be used since they do not produce results in the normal sense. Tournaments that are **Major**, **World** and **Low status** type of tournaments typically contain different players than a normal tournament. We have also chosen to exclude the Alfred Dunhill Links Championship since it has a cut after three rounds and contains a Pro-Am element. By choosing two subsequent normal tournaments the chance of obtaining two similar entry lists in the tournaments is reasonably high.

C) On some occasions two tournaments are played in the same week. These data are excluded.

D) The outcome tournament should be played in the week directly following the treatment tournament.

E) We exclude tournaments where less than 60 percent of players appear in the outcome tournament.

Essay 2: Is there a gender difference in the ability to deal with failures? Evidence from professional golf tournaments

1 Introduction

Women are underrepresented in top positions in firms and other organizations (see, e.g., Bertrand and Hallock [2001], Wolfers [2006] and Bertrand [2009]). One important explanation for this observation is that on average, women choose less competition-intensive careers than men. For example, women are typically underrepresented in the private sector (see, e.g., Lanfranchi and Narcy [2015]). Experiments have also shown that to a greater extent than men, women choose piece rate schemes over winner-takes-it-all schemes (see Niederle and Vesterlund [2007] and Dohmen and Falk [2011]).

More puzzling, however, is the observation that there also seems to be a glass ceiling for the women who actually enter highly competitive work environments, i.e. even competitive women struggle to reach top positions in firms (see, e.g, Albrecht, Björklund and Vroman [2003] and Arulampalam, Booth and Bryan [2007] for evidence of the glass ceiling effect). Discrimination against women is probably one explanation for this phenomenon, but it could potentially also be driven by remaining gender differences in competitiveness. Individuals that want to make career progress in competitive environments typically participate in multiple rounds of competitions in which they repeatedly compete for new positions and promotions. Most individuals are bound to experience multiple failures in the initial stages of their careers because they are competing against more experienced competitors and such negative outcomes might be detrimental for their confidence and subsequent performance (see Rosenqvist and Skans [2015] for the importance of previous competitive outcomes on current performance). Having a firm belief in one's ability is arguably important for not becoming too discouraged by failures and since previous evidence indicates that men have higher levels of confidence than women (see, e.g., Lundeberg, Fox and Punćcohaŕ [1994], Barber and Odean [2001] and Niederle and Vesterlund [2011]), women are potentially more vulnerable to failures with respect to the ability to bounce back. Consistent with this hypothesis, recent experimental evidence from a study on university students in the UK suggests that women, on average, respond more negatively than men to failures with respect to subsequent performance, which might explain why women, on average, are less likely than men to make substantial career advancements (see Gill and Prowse [2014] for the experimental study).

While the finding in Gill and Prowse (2014) is highly interesting for the understanding it provides of the general behavior of men and women, its relevance for the glass-ceiling phenomenon hinges on whether it can be replicated off the lab in situations where stakes are high, and in particular among men and women who have chosen to pursue careers in competition-intensive work environments. However, identifying causal effects of successes/failures on subsequent performance for competitive men and women on the regular labor market is difficult due to the general scarcity of relevant data and because of systematic ability differences between individuals that fail and individuals that succeed. The situation is, however, more favorable when focusing on the world of sports. While performance in an athletic setting relates to very particular tasks, it is a setting in which highly competitive men and women are active and in which performance data is often readily available. As such, sports competitions constitute a useful testing ground for theories about the behavior of competitive men and women.[1]

Wozniak (2012) and Jetter and Walker (2015) both use data from all-male and all-female professional tennis tournaments to study how the probability of winning the current game is affected by previous results. Using selection-on-observables strategies to identify the causal effect of previous results on current performance, they both find that men and women are more likely to win the current game if they have experienced recent successes and that these effects are very similar in magnitude across the genders. Similarly, Banko, Leeds and Leeds (2016) study whether female tennis players are more likely than men to lose in straight sets (the hypothesis being that women find it harder to come back after losing the first set), but do not find any gender differences. Overall, these findings would suggest that the result in Gill and Prowse (2014) about women being particularly sensitive to failures does not hold among competitive men and women, who are instead equally sensitive to previous competitive outcomes with respect to current performance. A fundamental problem with these observational tennis studies is, however, that they cannot control for the counterfactual development, which makes it hard to determine whether persistent successes (and failures) are due to causal success/failure effects (i.e. the first success causing the next one) or just time-varying ability. In addition, the result of a tennis game is affected by the performance of the opponent, making it even harder to cleanly estimate causal success/failure effects on subsequent performance.

Regarding this issue, using data from professional golf tournaments on the male European Tour, Rosenqvist and Skans (2015) made a key contribution by providing quasi-experimental evidence from professional golf tournaments in which same-ability players randomly end up in success or failure states. In these tournaments, players are separated into success and failure halfway through the tournaments by the so-called cut (a qualification threshold). Players close to the cut have performed almost equally well, but will arguably perceive their performances differently in terms of success or failure. By comparing the performance of marginally successful players and their marginally unsuccessful "copies" in the next tournament, the confound-

[1] Data from golf tournaments have, e.g., been used to study predictions of tournament theories (Ehrenberg and Bognanno [1990a and 1990b], Orszag [1994] and Melton and Zorn [2000]), peer effects (Guryan, Kroft and Notowidigdo [2009]) and loss aversion (Pope and Schweitzer [2011]).

ing impact of ability can be purged from the analysis and the causal effect of experiencing a success (relative to a failure) can be identified (i.e. a regression discontinuity [RD] strategy is used for identification). Rosenqvist and Skans (2015) found that male golfers substantially enhance their performance after a success, but they did not analyze the corresponding behavior of female golfers. In this paper, I add data from the PGA Tour (main tour for men in the US) and the LPGA Tour (main tour for women in the US), making it possible to use the same RD strategy to examine potential gender differences in the causal impact of a previous success/failure on current performance.

Thus, in this paper I provide the first quasi-experimental evidence from the field on potential gender differences in the productivity response to previous competitive outcomes among competitive men and women. These top-performing male and female athletes are active in an environment with multiple rounds of competition, which resembles the situation for men and women in the corporate sector trying to make career progress.

The results show that the current performance of both male and female golfers is negatively affected by a previous failure and that the effects are virtually identical in magnitude. The results suggest that the confidence of top-performing competitive men and women is affected by previous competitive experiences and that this effect has a substantial impact on subsequent performance. However, women show no tendencies toward being more sensitive than men to previous outcomes. Thus, if the behavior of these professional male and female athletes is similar to the behavior of competitive men and women in the rest of the society, it seems unlikely that women are unable to reach top positions in firms because they are worse than men at dealing with failures. Instead, it seems likely that women's difficulties reaching top positions in firms and other organizations are caused by external barriers, which calls for more research on the structure of these barriers and how to penetrate them.

The remainder of this paper is structured as follows. Section 2 gives a detailed description of the data and Section 3 explains and tests the validity of the identification strategy. In Section 4, I present the main results. Section 5 concludes.

2 Data

2.1 General description

In this paper, I use data from professional golf tournaments. The data come from the European Tour (males)[2], the PGA Tour (males) and the LPGA Tour

[2] This data was used in Rosenqvist and Skans (2015).

(females).[3] The typical tournament in these tours is played over four days (normally Thursday–Sunday) and the players play 18 holes each day. The goal is to use as few strokes as possible to complete the holes. All entrants in a tournament play the first two days and based on the results after two completed days of play, a line is drawn in the list of results that separates the 70 best players from the rest of the field. This line is called the *cutline* or the *cut*, since players outside the top 70 are eliminated from the tournament at this stage.[4] The cut thus specifies the maximum number of strokes that a player is allowed to have to be qualified for the rest of the tournament. If a player satisfies that criterion, he or she *makes the cut*. Note that the cut is decided after two days of play which makes it hard to predict for the players while they are playing. It is, however, highly predictable at the late stages of the second round. Players that make the cut continue the tournament during the weekend and the final result is based on the total number of strokes used after 72 holes. All players that make the cut and finish the tournament receive prize money; the exact amount depends on the player's final position. Players that fail to make the cut must leave the tournament empty-handed after two days of play. Since making the cut brings money, prestige and ranking points, it seems reasonable to assume that players that fail to make the cut experience a sense of failure relative to the players that make the cut. Importantly, at the cut there is only a one-stroke difference between success and failure, making this setting ideal for identifying potential success/failure effects through an RD strategy.

The data is structured in pairs of tournaments played during two consecutive weeks where the first tournament plays the role of a "treatment" tournament and the second the role of an "outcome" tournament. In the first tournament, I have data on the value of the cut and the total number of strokes of each player after 36 holes.[5] It is therefore possible to determine if a player made the cut or not, i.e. if he or she is treated. I have access to the same kind of information for the second tournament, which is used to measure potential performance effects of making the cut in the first tournament (conditional on ability). The number of strokes after 36 holes is used because

[3] The data from the European Tour has been collected manually from the European Tour website (www.europeantour.com). The data from the PGA Tour has been collected manually from the PGA Tour website (www.pgatour.com) and from https://sports.yahoo.com/. The data from the LPGA Tour has been collected manually from the LPGA Tour website (www.lpga.com) and from https://sports.yahoo.com/. Small parts of the data on the female golfers also come from the following sites: www.golfdata.se, http://www.foxsports.com/ and golfweek.com.

[4] The exact rule varies between the different tours and it has also varied within tours over time. However, the most common use of the cut is that players that are tied for the 70th position or better make the cut.

[5] For a large share of the data, I only have access to results for players within a six-stroke difference from the cut. To create consistency across tournaments, this restriction is used throughout the paper, though some tournaments contain data on more players in the list of results.

all players participate up to that point. It should be noted that not all players in the first tournament participate in the second tournament, which means that the outcome is missing for some of the players that participated in the first tournaments. To reduce the potential problem of selective participation in the second tournament I only use results from tournament pairs where the participation rate is at least 60 % in the second tournament.[6]

The dataset also contains information on total prize money for all tournaments and individual player characteristics in the form of experience measures and ability measures.

There are some institutional differences between the male and female golf tours. First, while men always play over four days (unless the weather forces the competition to be shortened) some tournaments on the LPGA Tour only have three days of play. The cut is still after 36, holes but those who make the cut only play 18 additional holes instead of 36. Since this institutional difference has nothing to do with the cut rule or the importance of making the cut, it seems unlikely that it should matter for the success/failure effects. Secondly, the average prize money in men's tournaments is much higher than the average prize money in women's tournaments. The average prize money for men in my data is roughly $4,000,000 while the corresponding amount for women is $1,400,000. Essentially, this means that making (or failing to make) the cut has much larger financial consequences for male golfers than for female golfers. Even though the gender difference in prize money is large, there is still substantial prize money involved in women's tournaments as well, suggesting that perceptions of success or failure following a made or missed cut are likely to emerge both for male and female golfers. To make sure that difference in prize money does not interfere with the analysis, I do robustness checks in section 4.2 on samples that are comparable in terms of prize money.

2.2 Descriptive statistics

The most widely used sample in the paper contains 189 tournament pairs from the European Tour, 251 from the PGA Tour and 202 from the LPGA Tour. The total number of observations pertaining to the European Tour is 21,912 (16,515 participate in the outcome tournaments). The corresponding numbers for the PGA Tour and the LPGA Tour are 28,988 (19,604 participate in the outcome tournaments) and 21,682 (17,014 participate in the outcome tournaments). The number of unique players in the sample of 16,515 observations with non-missing information on outcomes on the European

[6] Making the cut is generally associated with a higher probability of participating in the outcome tournament. But conditional on the empirical RD model, making the cut has a negative effect on participation (statistically significant for men). This is, however, only a problem if it biases the distribution of the skill of the participating players around the cut. I test this in Section 3.2.

Tour is 1,020. The corresponding numbers for the PGA Tour and the LPGA Tour are 807 and 673.

Table 1 provides summary statistics for the sample that I use to examine treatment effects, i.e. for observations with non-missing data on the outcome variables. The stats are presented separately for men and women. Two general facts related to the empirical strategy should be highlighted. First, players that are successful in the treatment tournament (i.e. Cut=1) have better results than the unsuccessful players (i.e. Cut=0) in the outcome tournament. The successful players have fewer strokes after two rounds and are more likely to make the cut. This difference in future performance between successful and unsuccessful players is particularly pronounced for women. Second, while the above finding is consistent with a positive impact of a success on future performance, a complicating factor is that successful players were already better than their unsuccessful counterparts before the treatment tournament (see stroke average in the previous year). Thus, a simple comparison of mean future outcomes between successful and unsuccessful players is biased by ability differences. This highlights the need for an empirical model that allows us to estimate the impact of making the cut, relative to failing to make it, on future performance conditional on ability. A model that does just that is explained in Section 3.

It should be noted that the female golfers have roughly three years less experience than the males as measured by time as a professional golfer (being a professional golfer means the golfer can compete for money). This implies that the two samples are not completely comparable. On the other hand, as Hensvik (2014) shows, women in the higher ranks of firms are often less experienced than their male peers, making the data in this paper empirically relevant. Nevertheless, in Section 4.2 I do robustness checks on samples that are comparable in terms of experience.

Table 1. Descriptive statistics for the used sample

Column:	(1)	(2)	(3)	(4)	(5)	(6)
		Men			Women	
	All	Cut=1	Cut=0	All	Cut=1	Cut=0
Treatment tourn.						
Average cut	143.01	143.01	143.0	145.45	145.51	145.38
Normalized str.	0.19	-2.06	3.01	0.07	-2.18	3.00
… std. deviation	3.00	1.64	1.64	3.05	1.67	1.63
Made the cut	0.56	1.00	0.00	0.56	1.00	0.00
Outcome tourn.						
Average cut	142.75	142.77	142.72	145.61	145.72	145.47
Normalized str.	0.31	-0.14	0.87	0.05	-0.77	1.12
… std. deviation	4.37	4.31	4.38	4.68	4.49	4.71
Made the cut	0.55	0.59	0.50	0.56	0.63	0.47
Player character.						
Years as pro	11.98	12.02	11.93	8.63	8.51	8.80
… std. deviation	6.38	6.29	6.50	5.79	5.59	6.04
… nonmissing	0.98	0.98	0.97	1.00	1.00	1.00
Str. avg. season t-1	71.67	71.56	71.81	72.96	72.72	73.28
… std. deviation	1.16	1.09	1.23	1.36	1.29	1.39
… nonmissing	0.88	0.90	0.87	0.93	0.95	0.91
No. of tournaments	440	440	440	202	202	202
No. of clusters	5,229	2,627	2,602	2,404	1,208	1,196
No. of observations	36,119	20,100	16,019	17,014	9,597	7,417

Notes: A cluster is a tournament * strokes combination.

3 Empirical strategy

3.1 Empirical model

The fundamental assumption behind the empirical strategy is that players with results close to the cut ended up on the right or wrong side of it by chance. If so, the ability of players close to the cut should be virtually identical, which means I can estimate the effect of making the cut on future performance conditional on ability, i.e. I can estimate the causal effect of experiencing a relative success on the performance in the next tournament. The validity of this assumption is, of course, central for this exercise and it will be studied in detail in Section 3.2.

In the ideal RD setting, the researcher can compare the mean outcome of the treated and the controls that are infinitely close to the threshold, since these individuals have balanced covariates. In practice, however, the number of observations typically goes to zero as we get closer to the threshold, forcing the researcher to adopt a wider bandwidth. With a wider bandwidth comes the problem of unbalanced covariates, which means that the simple comparison of outcomes must be abandoned in favor of a method that approximates the value precisely at the cutoff for treated and controls respectively. When the variable that determines assignment to treatment (hereafter called *running variable*) is discrete, as in this paper, then by construction the bandwidth is too wide for a simple comparison of mean outcomes. Instead, I use the relationship between the running variable and the outcome variable to approximate the outcome for hypothetical individuals that are just marginally on the success or failure side of the threshold (see Lee and Lemieux [2010] for a thorough description of this method). This is done using the regression model specified in Eq. (1):

$$Y_{ic} = \beta_0 + \beta_1 I[Z_{ic} \leq 0] + \beta_2 Z_{ic} + \beta_3 I[Z_{ic} \leq 0] Z_{ic} + \delta_c + u_i \quad (1)$$

The outcome, denoted Y_{ic}, is a measure of the performance in the second tournament (typically the number of strokes), where the subscript c indicates that it is a competition-specific outcome. Z_{ic} is the number of strokes after 36 holes in the treatment tournament normalized by the subtraction of the cut in the tournament. Thus, as Z_{ic} crosses zero from the positive side, the treatment goes from off to on. Since both Z_{ic} and Y_{ic} constitute measures of ability (the higher, the worse) we expect a positive relationship between the two. The terms $\beta_2 Z_{ic}$ and $\beta_3 I[Z_{ic} \leq 0] Z_{ic}$ allow this relationship to be different on the two sides of the threshold. With the help of the estimated relationship between Z_{ic} and Y_{ic} it is possible to predict the values of Y_{ic} as Z_{ic} approaches zero from below and above respectively. The difference between these two values measures what happens with the outcome as the treatment is turned on while the running variable is held constant. Thus, the estimate of β_1 approximates the difference in mean outcome for treated and controls that are infinitely close to the threshold (i.e. that have virtually the same ability). δ_c captures competition fixed effects and u_i is an error term.

If players close to the threshold really have the same ability, the estimate of β_1 corresponds to the causal effect of making the cut, relative to failing to make it, on the performance in the outcome tournament. Importantly, β_1 gives the performance difference between marginal winners and marginal losers, not between marginal winners and completely unaffected players. Thus, if marginal winners outperform marginal losers, this potential difference can be driven both by marginal winners improving their performance (relative to their hypothetical unaffected control state) and by marginal losers decreasing their performance (relative to their hypothetical unaffected control state). The data do not allow me to disentangle these two potential mechanisms.

As in Rosenqvist and Skans (2015), I cluster the standard errors at the strokes by tournament level because of potential joint specification errors for each stroke-group (see Lee and Card [2008] for a discussion of standard errors when performing RD analyses with a discrete running variable).

The main purpose of the paper is to investigate whether the value of β_1 is different for men and women and such tests can be done by interacting all variables in Eq. (1) with an indicator for being a woman. By doing so, success/failure effects for both men and women can be estimated in a joint regression framework and potential gender differences can be directly examined. Formally, I use the statistical model specified in Eq. (2):

$$Y_{ic} = \beta_0 + \beta_1 I[Z_{ic} \leq 0] + \beta_2 Z_{ic} + \beta_3 I[Z_{ic} \leq 0]Z_{ic} + \beta_4 Female_i \\ +\beta_5 Female_i I[Z_{ic} \leq 0] + \beta_6 Female_i Z_{ic} + \beta_7 Female_i I[Z_{ic} \leq 0]Z_{ic} + \\ \delta_c + u_i \tag{2}$$

In this model, β_1 corresponds to the causal effect of making the cut, relative to failing to make it, on the performance in the outcome tournament for men, while the sum of β_1 and β_5 gives the corresponding effect for women. The difference between the genders is thus given by β_5 which constitutes the main parameter of interest.

3.2 Validity of the empirical strategy

Two conditions need to be fulfilled in order for the empirical strategy to be valid. First, the ability of players must be continuous in the running variable across the threshold. Second, it is required that the incomplete participation in the outcome tournament does not bias the ability balance at the threshold. Thus, it is not enough that players distribute themselves randomly around the cutoff in the treatment tournament; instead, the test of randomization must be done *conditional on participation in the second tournament*.

To test whether players close to the cutline that actually participated in the subsequent outcome tournament ended up on the success or failure side of the threshold by random chance rather than due to deterministic reasons driven by ability differences, I investigate if the number of observations evolves smoothly over the cutoff and if predetermined measures of experience and ability are continuous around the cutoff, conditional on the empirical model (i.e. Eq. [1]).

Figure 1 describes the distribution of the running variable for males (top panel) and females (bottom panel) respectively. Positive numbers on the running variable indicate that the players failed to make the cut with 1 stroke, 2 strokes and so on. For both men and women the number of observations reaches its maximum at 0, meaning that the most common result after 36 holes is to have the same number of strokes as the cut stipulates. However, this is not an unnatural mass point; rather, it is the distribution one would expect to see even if there was no cut and all players were allowed to complete the tournament. Since by definition the cut lies in the middle of the field and since ability is arguably a normally distributed variable, it is not surprising to see that most players end up exactly on the cut. A more interesting exercise is to instead see how the relative difference in the number of observations between 0 (made the cut) and 1 (did not make the cut) looks in comparison with other relative differences. This is shown on the right-hand side of Figure 1. The value at e.g. -4 represents the absolute difference in the number of observations between -5 and -4 divided by the number of observations at -5. Thus, the relevant bar for our purposes is the one at 1. As can be seen, the relative difference between 0 and 1 does not in any way stand out in the distribution of relative differences, which suggests that neither men nor women can "force" themselves into just barely making the cut.

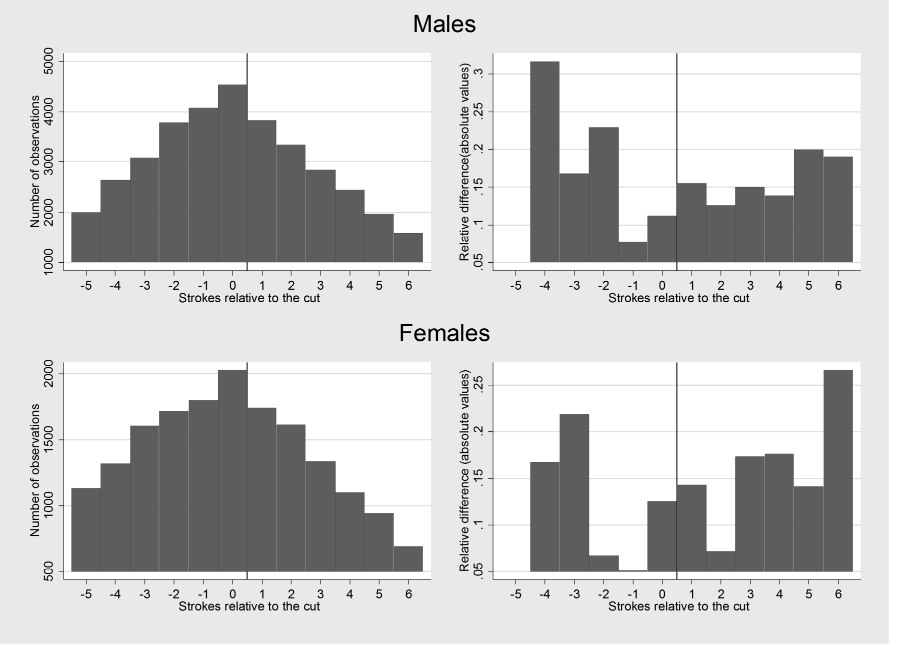


Figure 1. Distribution of running variable for males and females

Notes: The histograms for males are based on 36,199 observations. The histograms for females are based on 17,014 observations.

Even if no signs of manipulation at the cutoff can be found by looking at the distribution of the running variable, we still cannot rule out the possibility that marginal winners and marginal losers are different from each other in a systematic way. In Figure 2, I therefore examine how the predetermined ability of the players evolves over the threshold for women and men. The predetermined ability is measured with the average number of strokes used per round in the season preceding the treatment tournament. This statistic is generally considered a precise measure of a player's underlying ability and it tends to be stable over years. The last point is clear in Figure 2 since we see that female and male golfers that performed poorly in the treatment tournament (e.g. strokes relative to the cut equal to 5 or 6) also displayed a high stroke average in the previous season. The fact that the data show a strong positive association between the running variable and the stroke average in the previous season suggests that the stroke average the previous season is an accurate measure of players' abilities going into the treatment tournament.

Reassuringly for the empirical strategy, there is no jump in this ability measure at the cutoff, meaning that any potential jumps in the outcome variables at the cutoff are not driven by predetermined ability differences. While the overall picture looks the same for women and men, the relationship between previous and current performance is stronger for women (i.e. the slope is steeper). This means that female golfers exhibit greater consistency in their performances than men. While this result lies outside the focus of this paper, it is an interesting finding that calls for more research on performance consistency among men and women in different contexts.

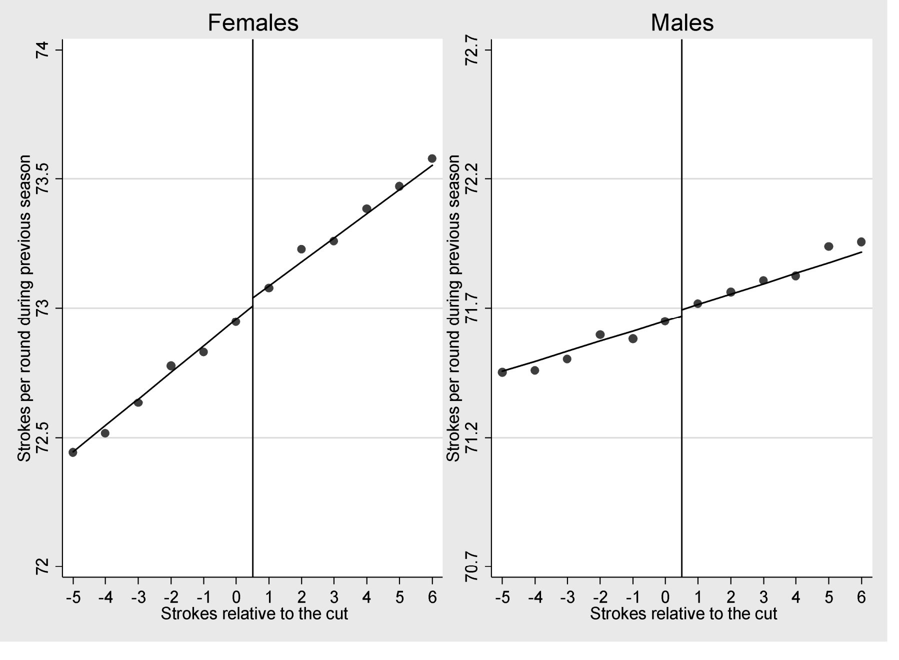


Figure 2. Strokes per round during the previous season

Notes: The figure is based on observations with non-missing data on the relevant characteristic which amounts to 15,878 observations for females and 31,940 observations for males.

Another way to test whether marginal winners and marginal losers are comparable is to examine how the experience of the players evolves over the threshold. If there is a jump at the cutoff such that more experienced players are on the success side to a greater extent, it would indicate that the cut is predictable and that experienced players can better predict the cut and adjust their play so that they just marginally make it. Figure 3 shows, however, that such worries are misplaced, since the measure of experience displays smoothness at the cutoff. The specific measure of experience is number of years as a professional golfer, which captures how long the player has been competing for money. The relationship between the running variable and experience is unclear. For women, relatively inexperienced players have the best results while inexperienced men display large variation in their results. However, the picture around the threshold is similar for the two genders; there is no discontinuity in experience as the running variable crosses the cut.

Overall, these validity checks confirm that the empirical strategy can give a robust identification of success/failure effects on performance.

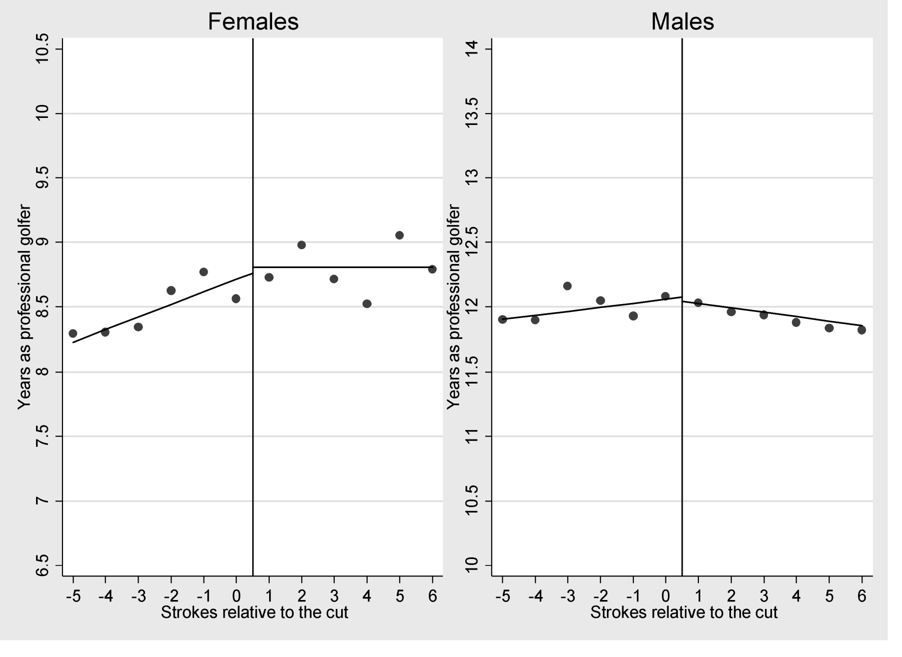


Figure 3. Years as professional golfer

Notes: The figure is based on observations with non-missing data on the relevant characteristic which amounts to 16,994 observations for females and 35,350 observations for males.

4 Results

4.1 Main results

Figure 4, Figure 5 and Table 2 present the main results, i.e. estimates of the effect of a previous (relative) success on current performance conditional on ability.

Figure 4 shows the number of strokes after two days in the outcome tournament (normalized by the cut in the relevant outcome tournament) on the y-axis and the number of strokes relative to the cut in the treatment tournament on the x-axis. Clearly, a good performance in the treatment tournament (i.e. a low x-value) is associated with a good performance in the outcome tournament as well (i.e. a low y-value). Thus, relatively better players consistently perform well whereas relatively worse players consistently perform poorly. The focus of our attention should, however, be the action at the cutoff where we see that the hypothetical marginal winners (just below 0.5) outperform the hypothetical marginal losers (just above 0.5) for both women and men. The difference at the threshold is virtually identical for the two sexes, amounting to roughly 0.16 shots.

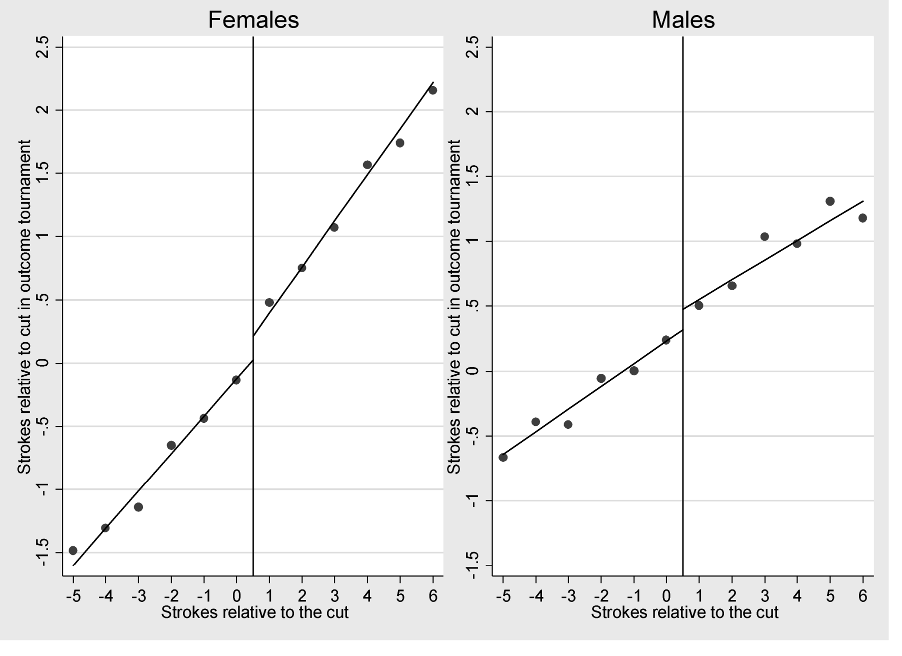


Figure 4. Strokes relative to the cut in the outcome tournament

Notes: The figure is based on observations with non-missing data on the outcome variable which amounts to 17,014 observations for females and 36,119 observations for males.

Figure 5 shows the results when making the cut in the outcome tournament is used as the outcome variable. Again, we see that women and men behave similarly at the cutoff with marginal winners being about 2.5 percentage points more likely to make the cut in the next tournament than marginal losers. Making the cut is of substantial economic importance for the players, so the performance difference at the cutoff that we saw in Figure 4 has quite large real effects.

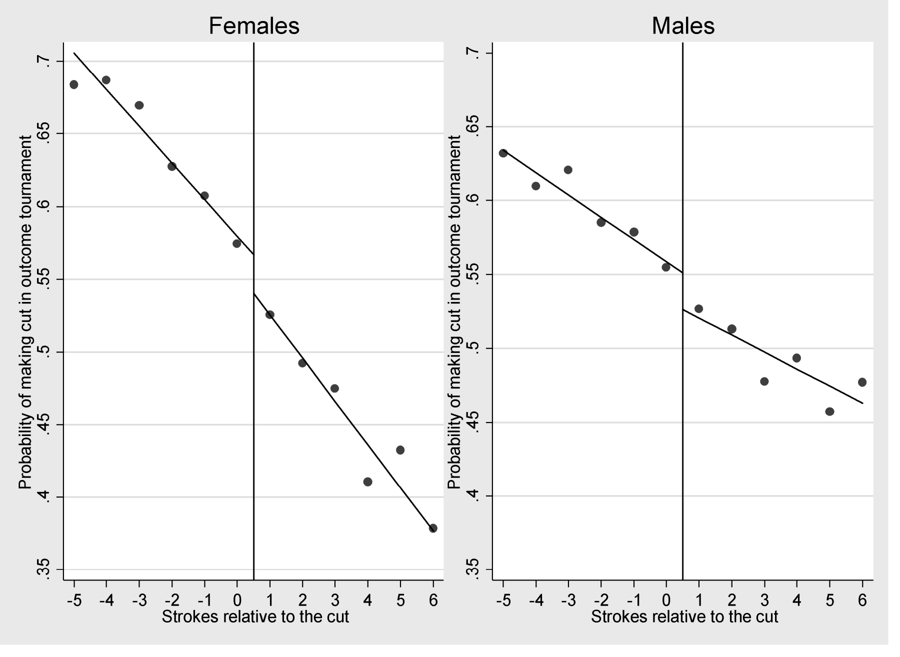


Figure 5. Probability of making the cut in the outcome tournament

Notes: The figure is based on observations with non-missing data on the outcome variable which amounts to 17,014 observations for females and 36,119 observations for males.

Table 2 contains the formal estimates of the results shown so far as well as tests of the differences between men and women. The estimates in column (1) correspond to estimates of β_1 when I estimate Eq. (1) using the full sample. In columns (2) and (3) I repeat the exercise for males and females respectively. The estimates in column (4) correspond to estimates of β_5 when I estimate Eq. (2) using the full sample. The negative effect on the number of strokes in the outcome tournament is significant on the 5 percent level for the full sample (column [1]) and for males (column [2]). As we saw in Figure 4, women exhibit a similar performance difference at the threshold but due to the smaller sample size, the precision is not enough to establish a statistically significant effect (see column [3]). The difference in effect size between men and women is very small (roughly one tenth of the baseline effect in column [1]) and statistically insignificant (see column [4]). This (small) point estimate also has the opposite sign of what should be expected if women are more sensitive than men to previous competitive outcomes.

With regard to performance differences between marginal winners and marginal losers measured by the propensity to make the cut in the outcome tournament, which are presented in panel B, we also find a statistically significant result for women (column [3]). Again, the difference in effect size between men and women is small and statistically insignificant (see column [4]).

Overall, the estimates suggest that performance of both men and women are causally affected by previous successes. Furthermore, the estimates do not lend support for the notion that female performance should respond more than male performance to previous competitive outcomes. Although an important caveat is that the confidence interval of the estimated interaction term allows for substantial differences between men and women (but in either direction). Hence, the second conclusion primarily rests on the fact that the point estimates remain tiny relative to the main effect.

Table 2. Main results – Marginal winners relative to marginal losers

Column:	(1)	(2)	(3)	(4)
Sample:	All	Males	Females	All
Estimate:	Main	Main	Main	Females-Males
Panel A. Strokes after 36 holes in the outcome tournament				
Making the cut	-0.1629**	-0.1706**	-0.1558	0.0148
	(0.0673)	(0.0804)	(0.1200)	(0.1444)
Observations	53,133	36,119	17,014	53,133
Mean	143.8936	143.0593	145.6649	143.8936
Panel B. Making the cut in the outcome tournament				
Making the cut	0.0255***	0.0266***	0.0240*	-0.0025
	(0.0079)	(0.0097)	(0.0133)	(0.0164)
Observations	53,133	36,119	17,014	53,133
Mean	0.5527	0.5487	0.5612	0.5527
Sample window	[-5,6]	[-5,6]	[-5,6]	[-5,6]
Linear RV	Yes	Yes	Yes	Yes
By treatment RV	Yes	Yes	Yes	Yes
Quadratic RV	No	No	No	No
By tournament RV	No	No	No	No
Covariates	No	No	No	No

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level. RV=running variable.

4.2 Robustness checks: model variations, experience and prize money

In Table 3, the results from Table 2 are subjected to a number of robustness checks in the form of model variations. The estimates in all seven columns are from variations of Eq. (2) and they all correspond to estimates of β_5 in that model (i.e. the interaction between making the cut and being a woman).[7] In columns (1–4) the bandwidth is gradually reduced and in column (5) I introduce a quadratic control for the running variable. In column (6) I add covariates (experience and predetermined ability) to the baseline model and specification (7) allows the linear relationship between the outcome and the running variable to be specific for each tournament. Overall, the point estimates of the interaction effect are fairly robust to these model variations, although they display some sensitivity to the very small bandwidths (see especially column [3]). As we can see in Figures 4 and 5, this is mainly driven by the fact that males with a running variable equal to three display quite extreme results relative to the general trend. Given that the other estimates are reasonably similar to the ones presented in Table 2, I interpret the results as suggesting that men and women do indeed respond similarly to previous results with respect to current performance. As Table A1 in the appendix shows, making the cut in the treatment tournament relative to failing to make it decreases the number of strokes after 36 holes in the outcome tournament by roughly 0.15 strokes for both men and women. Similarly, the probability of making the cut in the outcome tournament increases by about two percentage points for both men and women.

[7] Separate results for men and women are presented in Table A1 in the appendix.

Table 3. Robustness checks – model variations

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Strokes after 36 holes in the outcome tournament							
Making the cut * Female	-0.0229	0.0075	-0.2635	-0.1437	-0.0281	0.0382	0.0102
	(0.1548)	(0.1705)	(0.1978)	(0.2622)	(0.1402)	(0.1382)	(0.1364)
Observations	47,725	40,870	32,638	22,964	53,133	53,133	53,133
Mean of dep. var.	143.90	143.90	143.91	143.89	143.89	143.89	143.89
Panel B. Making the cut in the outcome tournament							
Making the cut * Female	0.0019	-0.0119	0.0216	-0.0001	0.0011	-0.0042	0.0003
	(0.0176)	(0.0197)	(0.0229)	(0.0303)	(0.0159)	(0.0161)	(0.0155)
Observations	47,725	40,870	32,638	22,964	53,133	53,133	53,133
Mean of dep. var.	0.5513	0.5505	0.5470	0.5474	0.5527	0.5527	0.5527
Sample window	[-4,5]	[-3,4]	[-2,3]	[-1,2]	[-5,6]	[-5,6]	[-5,6]
Linear RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes
By treatment RV	Yes	Yes	Yes	Yes	No	Yes	Yes
Quadratic RV	No	No	No	No	Yes	No	No
By tournament RV	No	No	No	No	No	No	Yes
Covariates	No	No	No	No	No	Yes	No

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level. RV=running variable.

As pointed out in Section 2, the male and female golfers studied in this paper differ in terms of experience and they are active in environments with different financial conditions. These differences might interfere with the analysis in such a way that I estimate institutional differences instead of gender differences (Rosenqvist and Skans [2015] for example found that the prize money in the outcome tournament affects the results). To examine this potential problem, I reestimate the model in Table 4 using samples that are fairly comparable in terms of experience and prize money. Since I have far more observations for male golfers than for female golfers, I make sample restrictions on the male sample to achieve comparability across the sexes. I drop all men that have more than 18 years of experience and that participated in tournaments with total prize money of more than $3,300,000. These restrictions leave me with a sample of male golfers that, on average, have 9.5 years of experience (compared with 8.6 for women) and that participate in tournaments with an average prize sum of $1,800,000 (compared with $1,400,000 for women). Thus, with these restrictions, the samples are substantially more similar than before, while they still allow me to keep roughly 15,000 observations for male golfers. The results from this exercise, which are presented in Table 4, are similar to the corresponding estimates in Table 2. The success effect for men on the number of strokes goes up somewhat in absolute terms (see column [2] of panel A) while the effect on making the cut instead makes a modest move downwards (see column [2] of panel B). But overall, the results are strikingly robust to these substantial sample restrictions, suggesting that the gender differences in experience and prize money do not interfere with the main conclusion that competitive men and women respond similarly to previous results with respect to current performance.

Table 4. Robustness checks – similar experience and prize money across genders

Column:	(1)	(2)	(3)	(4)
Sample:	All	Males	Females	Females
Estimate:	Main	Main	Main	Females-Males
Panel A. Strokes after 36 holes in the outcome tournament				
Making the cut	-0.1837**	-0.2151*	-0.1558	0.0593
	(0.0868)	(0.1241)	(0.1200)	(0.1726)
Observations	32,164	15,150	17,014	32,164
Mean of dep.	144.61	143.42	145.66	144.61
Panel B. Making the cut in the outcome tournament				
Making the cut	0.0230**	0.0220	0.0240*	0.0021
	(0.0099)	(0.0146)	(0.0133)	(0.0197)
Observations	32,164	15,150	17,014	32,164
Mean of dep.	0.5608	0.5605	0.5612	0.5608
Sample window	[-5,6]	[-5,6]	[-5,6]	[-5,6]
Linear RV	Yes	Yes	Yes	Yes
By treatment RV	Yes	Yes	Yes	Yes
Quadratic RV	No	No	No	No
By tournament RV	No	No	No	No
Covariates	No	No	No	No

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level. RV=running variable.

4.3 Additional results: high and low stakes

In Section 4.1 I found that marginal winners, in the treatment tournament, outperform marginal losers with respect to the performance in the outcome tournament. In this section I investigate how sensitive this performance difference is to the magnitude of the initial relative success/failure (i.e. the total prize money in the treatment tournament) and to the stakes in the outcome tournament (i.e. the total prize money in the outcome tournament).[8] The exercises are performed separately for men and women and the sensitivity of the effect sizes to the prize money is then compared. The magnitude of the initial relative success/failure is a binary variable that takes on the value 1 if the total prize money in the treatment tournament was above the median of the prize money in the treatment tournaments within the relevant combina-

[8] Note that I only have data on the total prize money in the tournaments, not the prize money for the individual players. Thus, I use variation in prize money between tournaments and not between players within tournaments.

tion of tour and season and 0 otherwise. The stakes in the outcome tournament are also coded as a binary variable where 1 indicates that the prize money in the outcome tournament was above the median of the prize money in the outcome tournaments within the relevant combination of tour and season.

I also group the players into four categories according to the values of the aforementioned two variables (i.e. zero-zero, zero-one, one-zero and one-one). Doing so, I can in a simple way investigate how the success/failure effect is affected by the stakes in the outcome tournament holding the magnitude of the initial success/failure constant and vice versa.[9]

My findings are presented in Table 5, where for ease of presentation, I only use making the cut in the outcome tournament as the outcome variable. In panel A, I focus on players who participated in a treatment tournament with below-median prize money. Thus, marginal winners and losers in this sample experienced relatively small successes and failures. I then investigate how the performance difference between these players in the outcome tournament is affected by the size of the prize money in the outcome tournament. In panel B, I do the corresponding exercise for players that participated in a treatment tournament with above-median prize money. In panels C and D, I keep the prize money in the outcome tournaments fixed and vary the prize money in the treatment tournaments.

Azmat, Calsamiglia and Iriberri (forthcoming) and Gill and Prowse (2014) have previously conducted similar investigations in other settings. Azmat, Calsamiglia and Iriberri (forthcoming) study potential gender differences in the reaction to changed stakes. They studied Spanish high school students and found that female students tend to choke under pressure in the sense that the gender gap in test results (to the advantage of females) is smaller in high stakes tests than in low stakes tests. Intuitively, positive recollections of previous performances should be particularly important in situations where the probability of choking under pressure is relatively high (i.e. high stakes situations). Thus, the effect of making the previous cut on current performance should generally be higher in high stakes outcome tournaments, and if the results in Azmat, Calsamiglia and Iriberri (forthcoming) are relevant for adult competitive women as well, this pattern should be particularly pronounced for female golfers, since they are suggested to be more likely to choke under pressure. This reasoning implies that the estimates in column (2) of panels A and B should generally be higher than the corresponding estimates in column (1) and that the difference in column (3) should be higher for females than for males. The estimates in columns (1–2) of panels A and B are only statistically significant on one occasion (see

[9] This was also done in Rosenqvist and Skans (2015) using a slightly different empirical approach. They found that the success/failure effect for male golfers on the European Tour was entirely driven by high stakes outcome tournaments.

males in panel B), but the fact that high stakes outcome tournaments always produce greater point estimates than low stakes outcome tournaments strengthens the notion that earlier successes (which are assumed to build confidence) are most valuable in high stakes environments when players are likely to be under pressure. The success/failure effect for women is, however, not more sensitive to the prize money in the outcome tournament than the effect for men; instead, if anything, the estimates suggest that men are more confidence-dependent than women in high stakes situations since they seem to be very sensitive to the outcome of previous performances in exactly those cases. It should, however, be noted that the gender difference in the effect sensitivity to the prize money in the outcome tournaments is statistically insignificant in both panel A and panel B (see the difference-in-differences estimates at the bottom of the respective panels).

Gill and Prowse (2014) studied male and female university students in a laboratory setting. They found that men react to the size of an initial success/failure in such a way that, conditional on losing, their subsequent effort is only negatively affected if the loss was big (i.e. if a lot of money was foregone). Conditional on winning, subsequent effort was not affected by the size of the win. For my setting, this would suggest that the performance difference between marginal winners and marginal losers among men should be highest after a treatment tournament with above-median prize money. This implies that the estimates for males in column (2) of panels C and D should be higher than the corresponding estimates in column (1). The differences between the estimates are, however, very small and go in opposite directions in panels C and D, which suggests that the success/failure effect for male golfers is insensitive to the prize money in the treatment tournament. For women, Gill and Prowse (2014) find that conditional on losing, the subsequent effort is not affected by the size of the loss. Conditional on winning, however, subsequent effort decreases in the prize money. Thus, if the results in Gill and Prowse (2014) hold true in a wider context, the performance difference between marginal winners and marginal losers among female golfers should be at its maximum after a treatment tournament with below-median prize money. This implies that the estimates for females in column (1) of panels C and D should be higher than the corresponding estimates in column (2). Looking at the point estimates, this is true in both cases, although the differences fail to exhibit statistical significance (see column [3]). Still, the rather surprising result from Gill and Prowse (2014) about small previous successes being more beneficial for women's current performance than large ones is tentatively confirmed by my results, which calls for more research on the potential mechanisms behind this peculiar result. Comparing men and women with respect to the effect sensitivity to the prize money in the treatment tournaments, we see that according to the point estimates, women are more sensitive (see the difference-in-differences estimates at the bottom of panels C and D). But since the data are cut so thin in this exercise, the difference-in-differences estimates are not statistically significant.

Overall, the most important findings from this exercise are that women benefit from relatively small rather than large previous successes and that both men and women (especially men) are particularly dependent on positive recollections of previous performances when competing in high stakes situations.

Table 5. The importance of a previous success in situations with high and low stakes

Column:	(1)	(2)	(3)
Panel A. Prize money in treatment tournament low			
Prize money in outcome:	Low	High	Diff. (High-Low)
Males: Making the cut	0.0085	0.0457*	0.0372
	(0.0150)	(0.0255)	(0.0296)
Females: Making the cut	0.0296	0.0474	0.0177
	(0.0202)	(0.0313)	(0.0372)
Difference-in-differences (Females-Males): -0.0195 (0.0476)			
Panel B. Prize money in treatment tournament high			
Prize money in outcome:	Low	High	Diff. (High-Low)
Males: Making the cut	0.0008	0.0526***	0.0518*
	(0.0242)	(0.0178)	(0.0301)
Females: Making the cut	0.0054	0.0100	0.0047
	(0.0320)	(0.0278)	(0.0424)
Difference-in-differences (Females-Males): -0.0471 (0.0520)			
Panel C. Prize money in outcome tournament low			
Prize money in treatment:	Low	High	Diff. (High-Low)
Males: Making the cut	0.0085	0.0008	-0.0077
	(0.0150)	(0.0242)	(0.0285)
Females: Making the cut	0.0296	0.0054	-0.0243
	(0.0202)	(0.0320)	(0.0379)
Difference-in-differences (Females-Males): -0.0166 (0.0474)			
Panel D. Prize money in outcome tournament high			
Prize money in treatment:	Low	High	Diff. (High-Low)
Males: Making the cut	0.0457*	0.0526***	0.0069
	(0.0255)	(0.0178)	(0.0311)
Females: Making the cut	0.0474	0.0100	-0.0373
	(0.0313)	(0.0278)	(0.0419)
Difference-in-differences (Females-Males): -0.0442 (0.0522)			

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level.

4.4 Additional results: good and bad days

For about 80% of the observations I have the results on both round 1 and round 2 in the outcome tournament while the remaining observations only have the aggregate score over the two rounds.[10] This makes it possible to investigate whether a previous success is more beneficial on relatively good or bad days, or if the effect is constant. For each observation with non-missing data, I calculate the best day and the worst day and then examine how a previous success affects the results on the respective days conditional on the RD model.[11] Table 6 shows the results. The pattern of the results suggests that a previous success is particularly important on a relatively bad day when the players are struggling on the course. The making-the-cut effect is roughly twice as big on a bad day compared with a good day and the effect is only statistically significant on bad days (see column [1]). The same general pattern is apparent for both women and men (columns [2–3]). Arguably, these results strengthen the notion that confidence is the main factor behind the positive success, effect since players are more likely to start doubting their ability on relatively bad days. But with a fresh memory of success, these negative thoughts might be easier to keep at bay, making the recently successful players less likely to post very bad results. Effectively, this suggests that confidence reduces variance in performance.

[10] The detailed information exists for 100% of the tournaments on the PGA Tour, about 85% of the tournaments on the LPGA Tour and about 50% of the tournaments on the European Tour.

[11] About 10% of the observations have identical results on the two rounds and consequently, their worst day score is identical to their best day score.

Table 6. The effect of a previous success on good and bad days

Column:	(1)	(2)	(3)	(4)
Sample:	All	Males	Females	Females
Estimate:	Main	Main	Main	Females-Males
Panel A. Number of strokes on a good day				
Making the cut	-0.0425	-0.0541	-0.0197	0.0344
	(0.0403)	(0.0495)	(0.0685)	(0.0845)
Observations	42,160	27,520	14,640	42,160
Mean of dep.	70.22	69.84	70.95	70.22
Panel B. Number of strokes on a bad day				
Making the cut	-0.0931**	-0.0983*	-0.0833	0.0150
	(0.0442)	(0.0549)	(0.0730)	(0.0914)
Observations	42,160	27,520	14,640	42,160
Mean of dep.	73.33	72.96	74.02	73.33
Sample window	[-5,6]	[-5,6]	[-5,6]	[-5,6]
Linear RV	Yes	Yes	Yes	Yes
By treatment RV	Yes	Yes	Yes	Yes
Quadratic RV	No	No	No	No
By tournament RV	No	No	No	No
Covariates	No	No	No	No

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level. RV=running variable.

5 Conclusion

In experiments, women have been found to decrease their performance following a setback, while men appear to be unaffected (see Gill and Prowse [2014]). It has also been suggested that this gender difference in dealing with failures might partly explain the presence of a glass ceiling for women on the labor market, based on the logic that early career failures leave deeper scars on women than on men. However, this result has not been replicated among men and women that have actually chosen to enter competition-intensive work environments (see Jetter and Walker [2015] and Banko, Leeds and Leeds [2016], who have investigated the behavior of professional male and female tennis players). Instead, these studies have found that competitive men and women are equally sensitive to previous results. But in all field studies on this issue so far, identification has relied on selection-on-observables strategies, which leaves uncertainty regarding the robustness of the results.

In this paper, I contribute to the literature by providing quasi-experimental evidence from a field setting very well suited for identifying causal effects of a previous success/failure on current performance. I use data from about 200 all-female and 450 all-male golf tournaments. These tournaments involve a stringent qualification rule that can be used to study the effect of a previous success, relative to a failure, on current performance holding ability constant (this empirical strategy was previously used in Rosenqvist and Skans [2015] but only for male golfers). Halfway through professional golf tournaments, the worst-performing half of the players is eliminated from the tournament. The other players continue the tournament and earn at least some prize money in the end. Players that are just barely eliminated (marginal losers) and players that are just barely allowed to complete the tournament and earn prize money (marginal winners) performed almost equally well, but will arguably experience the performance differently in terms of success or failure. Using an RD design, I estimate the performance difference in the tournament the following week between marginal winners and marginal losers.

The analysis reveals two main findings. First, marginal winners generally outperform marginal losers in the subsequent tournament. Marginal winners have roughly 0.16 fewer shots (significant on the 5 percent level) after 36 holes in the outcome tournament and they are 2.5 percentage points more likely to make the cut (significant on the 1 percent level). This result shows that the finding in Rosenqvist and Skans (2015) about the existence of substantial causal success effects also holds true in a wider setting where women are included. Second, men and women exhibit virtually identical results, suggesting that top-performing women can tackle failures just as well as top-performing men. Thus, if the behavior of these professional male and female athletes is similar to the behavior of competitive men and women in the rest of the society, it seems unlikely that women are unable to reach top-positions in firms because they are worse than men at dealing with failures. Instead, it seems likely that women's difficulties in reaching top-positions in firms and other organizations are caused by external barriers, which calls for more research on the structure of these barriers and how to penetrate them.

The analysis has also produced four additional results. First, the data suggest that female golfers are more consistent performers than male golfers, i.e. women's results are very stable over time. To the best of my knowledge, this is the first time that this has been shown and an interesting avenue for future research is to explore how general this apparent gender difference in performance consistency is. This could potentially be due to male golfers choosing riskier strategies than female golfers, which produces greater variation in the number of strokes. Second, female golfers benefit more from relatively small previous successes than large ones, which replicates the finding in Gill and Prowse (2014). Third, both men and women (especially men) are particularly dependent on having had a recent success when competing in high stakes environments. This result strongly suggests that confidence is the main factor behind the success effect for both women and men, since confidence is arguably crucial when players are under intense pressure. Fourth, for both men and women, higher confidence (from a previous success) tends to help the players by improving their lowest ability level rather than their highest, effectively reducing between-day variance in performance.

References

Albrecht, J., A. Björklund, and S. Vroman. 2003. "Is there a glass ceiling in Sweden?". *Journal of Labor Economics*, 21(1), 145–177

Arulampalam, W., A. L. Booth, and M. L. Bryan. 2007. "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution". *Industrial & Labor Relations Review*, 60(2), 163–186.

Azmat, G., C. Calsamiglia, and N. Iriberri. Forthcoming. "Gender differences in response to big stakes". *Journal of European Economic Association*.

Banko, L., E. M. Leeds and, M. A. Leeds. 2016. "Gender differences in response to setbacks: evidence from professional tennis". *Social Science Quarterly*, 97(2), 161–176.

Barber, B. M. and T. Odean. 2001. "Boys will be boys: gender, overconfidence, and common stock investment". *Quarterly Journal of Economics*, 116(1), 261–292.

Bertrand, M. 2009. "CEOs", *Annual Review of Economics*, 1(1), 121–150.

Bertrand, M. and K. F. Hallock. 2001. "The gender gap in top corporate jobs". *Industrial & Labor Relations Review*, 55(1), 3–21.

Dohmen, T. and A. Falk. 2011. "Performance pay and multidimensional sorting: productivity, preferences, and gender". *American Economic Review*, 101(2), 556–590.

Ehrenberg, R. G. and M. L. Bognanno. 1990a. "The incentive effects of tournaments revisited: evidence from the European PGA Tour". *Industrial and Labor Relations Review*, 43(3), 74S–88S.

Ehrenberg, R. G. and M. L. Bognanno. 1990b. "Do tournaments have incentive effects?". *Journal of Political Economy*, 98(6), 1307–1324.

Gill, D. and V. L. Prowse. 2014. "Gender differences and dynamics in competition: the role of luck". *Quantitative Economics*, 5 (2), 351–376.

Guryan, J., K. Kroft, and M. J. Notowidigdo. 2009. "Peer effects in the workplace: evidence from random groupings in professional golf tournaments". *American Economic Journal: Applied Economics*, 1(4), 34–68.

Hensvik, L. 2014. "Manager impartiality: worker-firm matching and the gender wage gap". *Industrial & Labor Relations Review*, 67(2), 395–421.

Jetter, M. and J. K. Walker. 2015. "Game, set, and match: do women and men perform differently in competitive situations?". *Journal of Economic Behavior & Organization*, 119, 96–108.

Lanfranchi, J. and M. Narcy. 2015. "Female overrepresentation in public and nonprofit sector jobs: evidence from a French national survey". *Nonprofit and Voluntary Sector Quarterly*, 44(1), 47–74.

Lee, D. and D. Card. 2008. "Regression discontinuity inference with specification error". *Journal of Econometrics*, 142(2), 655–674.

Lee, D. and T. Lemieux. 2010. "Regression discontinuity designs in economics". *Journal of Economic Literature*, 48(2), 281–355.

Lundeberg, M. A., P. W. Fox, and J. Punćcohaŕ. 1994. "Highly confident but wrong: gender differences and similarities in confidence judgments". *Journal of Educational Psychology*, 86(1), 114–121.

Melton, M. and T. S. Zorn. 2000. "An empirical test of tournament theory: the Senior PGA Tour". *Managerial Finance*, 26(7), 16–32.

Niederle, M. and L. Vesterlund. 2011. "Gender and competition". *Annual Review of Economics*, 3(1), 601–630.

Niederle, M. and L. Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?". *Quarterly Journal of Economics*, 122(3), 1067–1101.

Orszag, J. 1994. "A new look at incentive effects and golf tournaments". *Economics Letters*, 46(1), 77–88.

Pope, D. G. and M. E. Schweitzer. 2011. "Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes". *American Economic Review*, 101(1), 129–157.

Rosenqvist, O. and O. N. Skans. 2015). "Confidence enhanced performance?–The causal effects of success on future performance in professional golf tournaments". *Journal of Economic Behavior & Organization*, 117, 281–295.

Wolfers, J. 2006. "Diagnosing discrimination: stock returns and CEO gender". *Journal of the European Economic Association*, 4(2−3), 531–541.

Wozniak, D. (2012), "Gender differences in a market with relative performance feedback: professional tennis players", *Journal of Economic Behavior & Organization*, 83(1), 158–171.

Appendix A: Additional results

Table A1. Robustness checks

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Strokes after 36 holes in the outcome tournament - Males							
Making the cut	-0.1087	-0.1051	-0.0230	-0.1624	-0.1604**	-0.1513*	-0.1647**
	(0.0855)	(0.0940)	(0.1101)	(0.1433)	(0.0775)	(0.0791)	(0.0749)
Observations	32,532	27,938	22,410	15,783	36,119	36,119	36,119
Mean of dep.	143.0738	143.0718	143.0944	143.0561	143.0593	143.0593	143.0593
Panel B. Strokes after 36 holes in the outcome tournament - Females							
Making the cut	-0.1315	-0.0976	-0.2866*	-0.3061	-0.1885	-0.1131	-0.1544
	(0.1291)	(0.1423)	(0.1643)	(0.2197)	(0.1169)	(0.1134)	(0.1140)
Observations	15,193	12,932	10,228	7,181	17,014	17,014	17,014
Mean of dep.	145.6794	145.6854	145.7111	145.7179	145.6649	145.6649	145.6649
Panel C. Making the cut in the outcome tournament - Males							
Making the cut	0.0188*	0.0194*	0.0076	0.0213	0.0248***	0.0248***	0.0268***
	(0.0104)	(0.0116)	(0.0134)	(0.0177)	(0.0093)	(0.0095)	(0.0091)
Observations	32,532	27,938	22,410	15,783	36,119	36,119	36,119
Mean of dep.	0.5471	0.5475	0.5434	0.5452	0.5487	0.5487	0.5487
Panel D. Making the cut in the outcome tournament - Females							
Making the cut	0.0207	0.0075	0.0292	0.0212	0.0260**	0.0206	0.0271**
	(0.0142)	(0.0159)	(0.0186)	(0.0246)	(0.0128)	(0.0130)	(0.0126)
Observations	15,193	12,932	10,228	7,181	17,014	17,014	17,014
Mean of dep.	0.5604	0.5568	0.5548	0.5523	0.5612	0.5612	0.5612
Sample window	[-4,5]	[-3,4]	[-2,3]	[-1,2]	[-5,6]	[-5,6]	[-5,6]
Linear RV	Yes	Yes	Yes	Yes	Yes	Yes	Yes
By treatm. RV	Yes	Yes	Yes	Yes	No	Yes	Yes
Quadratic RV	No	No	No	No	Yes	No	No
By tourn. RV	No	No	No	No	No	No	Yes
Covariates	No	No	No	No	No	Yes	No

Notes: Standard errors are clustered on the strokes * tournament level (in parentheses). */**/*** significant at the 10 /5/1 percent level. RV=running variable.

Essay 3: Rising to the occasion? Youth political knowledge and the voting age

1 Introduction

The right to vote is a fundamental human right (The United Nations, 1948, art. 21). So why is it that not all citizens are allowed to vote?[1] The main argument against lowering or abolishing the voting age is that young individuals, according to proponents of the current voting age, typically lack the appropriate intellectual maturity and political knowledge for voting (see, e.g., Chan and Clayton [2006]). In this paper, motivated by recent evidence, I study whether having the right to vote in itself can stimulate the acquisition of political knowledge, i.e. if young people who are given the right to vote *rise to the occasion*.

Having the right to vote "empowers citizens to influence governmental decision-making and to safeguard their other human rights" (Human Rights Advocates [2009], p. 2). Thus, it must be considered a serious violation of human rights to exclude a section of the population from voting without very strong reasons for doing so. The typical voting age today is 18, meaning that a large share (approx. 25%, [Central Intelligence Agency, 2015]) of the population in most countries lacks the right to vote because they are considered immature in several ways. Can this voting age be justified or should it be lowered? And if it is lowered, how will it affect young individuals who become eligible? While this paper does not give a definite answer to these questions, it aims to contribute to the discussion by providing relevant and credible new evidence.

Obviously, this question has been discussed in many countries in the last ten years because there have been several changes to the voting age from 18 to 16. In 2007, Austria became the first European nation to adopt a voting age of 16 (see Wagner et al. [2012]) and in the Scottish independence referendum in 2014, 16-year-olds were also allowed to vote (UK Government [2015]). Another major example is Argentina, where the voting age was lowered to 16 in 2012 (see The Telegraph [2012]).

The experiences from the Scottish referendum have triggered a discussion in the UK about whether the voting age should be lowered to 16 in all elections. Alex Salmond, Scotland's first minister and leader of the Scottish Nationalist Party (SNP) at the time of the independence referendum, was impressed by the engagement among young voters and delivered the following comment in his first statement to the Scottish Parliament after the referendum (Brooks [2014]):

> There is not a shred of evidence for arguing now 16 and 17-year-olds should not be allowed to vote. Their engagement in this debate, this great constitutional debate, was second to none. They proved themselves to be the serious, passionate and committed citizens we always believed they should be. Eve-

[1] Individuals under the age of 18 are typically not allowed to vote. See list of voting ages around the world compiled by the Central Intelligence Agency (2015).

> ryone in this chamber should be proud of this chamber's decision to widen the franchise. There is an overwhelming, indeed an unanswerable, case for giving 16 and 17-year-olds the vote in all future elections in Scotland, indeed across the United Kingdom. All parties in this parliament I think should make a vow to urge Westminster to make this happen in time for next year's general election.

Given this current debate and the ethical dilemma surrounding the existence of a voting age, it is of major importance to accumulate information about the political knowledge of, primarily, 16 and 17-year-olds in relation to their older fellow citizens. To the best of my knowledge, there have been two major contributions to this topic: Chan and Clayton (2006) and Wagner, Johann and Kritzinger (2012), and they have reached completely different conclusions.

Chan and Clayton (2006) used British survey data from the 1990s and 2001 to study the political interest and knowledge of different age groups. They found a clear age gradient in political maturity, with 16 and 17-year-olds at the bottom of the competence scale. Based on this finding, they argue that the voting age should not be lowered to 16, since this would lead to negative consequences for the quality of democracy.

Wagner, Johann and Kritzinger (2012), on the other hand, using Austrian survey data from 2009 reach a contradicting conclusion. Wagner, Johann and Kritzinger (2012) find that 16 and 17-year-olds are about equally willing to participate in politics and about equally able to make informed voting decisions as their somewhat older fellow citizens (18–21 years of age). Particularly, they do not find the voting choices of 16 and 17-year-olds to be worse in terms of correspondence with preferences than those of somewhat older age groups. Based on this finding, they argue that 16 and 17-year-olds should not be excluded from voting on the grounds of insufficient political knowledge.

Why are the results so different? Wagner, Johann and Kritzinger (2012) offer a potential explanation related to the different voting laws in the UK and Austria. The 16 and 17-year-olds surveyed in Chan and Clayton (2006) did not have the right to vote, while the corresponding group in Wagner, Johann and Kritzinger (2012) actually did have that right. In fact, Austria introduced a voting age of 16 in 2007 and when Wagner, Johann and Kritzinger (2012) conducted their survey in May 2009, the 16 and 17-year-olds in the study were about to vote in their first election (European Parliament Election in June 2009). Wagner, Johann and Kritzinger (2012) argue that citizens who lack the right to vote have small incentives to become politically knowledgeable, while being eligible means that political knowledge can actually be put to use. Thus, according to Wagner, Johann and Kritzinger (2012), 16 and 17-year-olds in Austria are more politically mature relative to their somewhat older fellow citizens than their counterparts in the UK, in part because they have the right to vote. This reasoning suggests the existence of dynamic effects from changing the voting age, where the political maturity of young people endogenously adjusts to the prevailing voting age, i.e. young people "rise to the occasion".

This idea receives additional support from Hart and Atkins (2011) and Zeglovits and Zandonella (2013). Hart and Atkins (2011) argue that the voting age should be lowered to 16 because evidence indicates that opportunities for political socialization in adolescence lead to a deeper civic commitment in early adulthood. Along a similar line, Zeglovits and Zandonella (2013) provide evidence suggesting that the political *interest* of young people responds to changes in the voting age.[2] While the hypothesis that young people react to changes in the voting age is intriguing and consistent with the results in Chan and Clayton (2006) and Wagner, Johann and Kritzinger (2012), it has not been properly evaluated. The diverging results in Chan and Clayton (2006) and Wagner, Johann and Kritzinger (2012) could also potentially be explained by other cross-country differences between Austria and the UK, e.g. the school system, and have nothing to do with the voting age.

Since potential dynamic effects of changing the voting age would be a major argument in favor of lowering it, the question needs a more thorough treatment in a more controlled setting. Conceptually, the real question that must be answered in order to say something more definitive about the effects of lowering the voting age is the following: consider a large sample of identical twins, where one twin in each pair is allowed to vote at 16 and the other at 18. Will the twins who were allowed to vote at 16, on average, exhibit higher levels of political knowledge and/or political interest at age 16 than the twins who were allowed to vote at 18?

Given that very few countries allow 16-year-olds to vote, it is hard to obtain data to answer this exact question, but in this paper, using nationwide Swedish register data, I answer a very similar question. In Sweden, individuals are allowed to vote if they turn 18 on the day of the election at the latest. This law has been practiced in all elections since 1976 (The Swedish Election Authority [2015]). Utilizing the fact that the Swedish register data gives me access to the exact date of birth of all individuals born in Sweden from 1969, and onwards I can employ a regression discontinuity (RD) strategy to estimate the causal effect of having the first voting opportunity at 18, compared with having the first opportunity, on average, three years later, on a range of measures aimed at capturing political knowledge, political interest and civic interest around age 18. The empirical strategy relies on the as-

[2] Zeglovits and Zandonella (2013) looked at the political *interest* of 16- and 17-year-olds in Austria before and after the voting age was lowered to 16 in 2007. They found that political interest among 16 and 17-year-olds surveyed in 2008 was higher than political interest among 16 and 17-year-olds surveyed in 2004. This is consistent with a reaction to the change in the voting age, but Zeglovits and Zandonella (2013) acknowledge that their data makes it impossible to prove a causal relationship.

sumption that random chance determines whether births occurring near the voting eligibility cutoff date for elections 18 years later occur before or after the cutoff. In other words, individuals born just before the cutoff for voting eligibility should be comparable to individuals born just after it (i.e. they should have parents with similar characteristics etc.). Thus, I am able to causally identify the effects of a substantial change in voting age, albeit on a somewhat higher general age level (i.e. 21 to 18 instead of 18 to 16), on measures of political knowledge and interest. It should be noted that with a voting age of 18, the average age at first voting opportunity is around 20. If the voting age was lowered to 16, the average age at first voting opportunity would go down to 18. Thus, this paper investigates voting age effects at a highly relevant margin.

The main measure of political knowledge around age 18 in this paper is the high school grade in Social Studies (*Samhällskunskap*). According to the objectives of this subject, at least 25% of the grade should be based on students' political knowledge. This essentially means that political knowledge is measured with a lot of noise. While troublesome for precision, it should not lead to downward bias in the estimates. At the same time, this imprecise outcome measure is compensated by the fact that I can employ nationwide registers with a large number of observations (relative to survey-based studies like Chan and Clayton [2006] and Wagner, Johann and Kritzinger [2012]).

While the subject objectives suggest that the high school grade in Social Studies is a relevant measure of political knowledge, it is not perfect. To capture different traces of a potential increase in political knowledge and/or civic interest driven by the right to vote that potentially do not show up in the high school grade in Social Studies, I also employ two alternative measures: performance on the General Knowledge section of the Swedish Scholastic Assessment Test (SweSAT) and the orientation of tertiary studies. The performance on the General Knowledge section of the SweSAT captures knowledge in many different fields, but knowledge in Social Science-related questions is particularly important (Stage [1985] estimates that about 15% of the questions are related to political knowledge). Since the test is based on multiple choice questions, it should be cleansed from "classroom skills" which can contaminate school grades. Early participation in tertiary courses within the core subjects of Social Science is taken as an indicator of a high level of political interest. I also study an outcome variable that takes on the value 1 if the individual has a tertiary education ($\geq$ 3 years) within Social Science.

I find that individuals who have the opportunity to vote in an election just after their eighteenth birthday *do not* have higher levels of political knowledge and/or interest around age 18 than comparable individuals whose first voting opportunity takes place, on average, three years later. Individuals with voting rights do not have higher high school grades in Social Studies, nor do they have better results on the General Knowledge section of the SweSAT. If anything, they have worse results, but this cannot be definitively established. I also do not find any significant differences between treated and controls with respect to tertiary education outcomes. Overall, the results suggest that adolescents are unaffected by having the opportunity to vote. This finding weakens the case for lowering the voting age from 18 to 16, but it does not rule out that 16-year-olds, at least in some contexts, have sufficient levels of political knowledge for meaningful voting.

The remainder of this paper is structured as follows. In Section 2, I first discuss studies related to potential effects of changing the voting age. I then elaborate on the concept of political knowledge and explain the measures used in the paper. Section 3 deals with the data and the institutional setting, and in Section 4, I explain the identification strategy. In Section 5, I present the main findings and Section 6 provides a conclusion.

2 Related literature, political knowledge and measurements

In this section, I first provide a discussion of the potential effects of having the right to vote at an early age. This is followed by a short review of the literature on political knowledge. Finally, I discuss the measures of political knowledge, political interest and civic interest that are used in this paper.

2.1 Voting at an early age

Several recent studies have used discontinuities caused by voting age laws to study the effect of early-age voting eligibility on turnout in later elections (see, e.g. Meredith [2009], Coppock and Green [forthcoming] and Bhatti, Hansen and Wass [2016]). These studies have found that individuals who can vote shortly after turning 18 have higher future turnout rates than individuals who turn 18 just too late and whose voting debut therefore takes place some years later. While this strongly suggests that voting is habit-forming, there is still very little direct evidence on the effect of early-age voting eligibility on early-age political knowledge and interest.

There are, however, related studies reporting positive effects of general political socialization (e.g. community service, election-oriented teaching in school and mock voting) during adolescence on civic engagement in early adulthood. First, Hart et al. (2007) find a positive correlation between community service participation in high school and voting and volunteering in early adulthood using longitudinal data on a random sample of US high school students. Second, Syvertsen et al. (2009) performed a randomized trial on a small sample of high school students in the US. Some of the high school students were randomly assigned to an election-oriented curriculum while the other students received the normal curriculum. Syvertsen et al. (2009) then show that the treated students later expressed higher levels of civic interest and self-reported civic disposition than those in the control condition. Thirdly, Meirick and Wackham (2004) provide evidence indicating that the actual act of casting mock votes in school can deepen the civic commitment of students. Overall, Hart and Atkins (2011), who advocate a voting age of 16 in the US, interpret the collected evidence above as suggesting that, in terms of increased civic engagement, "there likely would be considerable benefits to allowing 16- and 17-year-olds to vote" (Hart and Atkins [2011], p. 217).

Since these studies use different measures of civic engagement as the outcome and since they do not look at early age voting right per se, they cannot be taken as evidence in favor of the existence of dynamic changes in political knowledge and interest among young people following a change in the voting age. However, they clearly indicate that teenagers react to different opportunities for political socialization.

2.2 Political knowledge: Definition, importance and measurement

Delli Carpini and Keeter (1996) define political knowledge as "the range of factual information about politics that is stored in long-term memory" (Delli Carpini and Keeter [1996], p. 10). As Mondak (2001) points out, the concept of political knowledge lies at the heart of the research on political behavior. Political knowledge is key within political science, because it has been found to correlate with, e.g., turnout and political beliefs (see Larcinese [2007] for turnout and Delli Carpini and Keeter [1996] for political beliefs). It is also highly relevant for the voting age debate.

Critics of a lower voting age (e.g. Chan and Clayton [2006]) argue that 16 and 17-year-olds lack political knowledge and that it would be detrimental for the quality of democracy to give them the right to vote. If individuals, due to limited political knowledge, are unable to identify which political alternative has the closest correspondence to their own beliefs, it might result in the represented political opinions being substantially different from the actual opinions of the citizens. To exemplify with an extreme case, supporters of lower income taxes might, due to ignorance, vote for a party that wants to raise taxes. Thus, giving the right to vote to individuals who are unable to differentiate between political alternatives is meaningless (see Lau, Andersen and Redlawsk [2008]) or even harmful for democratic legitimacy (see Scharpf [1999]). 16-year-olds might have low levels of political knowledge relative to their older fellow citizens when the voting age is 18, but this would not necessarily be the case if the voting age was 16. In fact, as

Delli Carpini and Keeter (1996) point out, motivation or the desire to learn is an important requisite for the acquisition of political knowledge and it might respond to changes in the voting age.

Political knowledge is typically measured by asking subjects a number of questions on factual political knowledge. The questions are normally formulated as propositions and respondents are asked to determine if they are true or false. Questions can, e.g., be about the name of the prime minister or about the number of parties currently represented in the parliament (see Chan and Clayton [2006] for more examples). While questions of this type do not capture all relevant aspects of political knowledge, the results from exercises of this type are generally considered good predictors of other political abilities.[3] Wagner, Johann and Kritzinger (2012) measure political knowledge by looking at subjects' abilities to place different parties correctly on the left-to-right ideological scale and thus they deviate from the typical factual questions.

2.3 Measures of political knowledge and political interest used in this paper

2.3.1 The Swedish education system

Since many of the measures used in this paper are related to the Swedish school system, it is probably helpful to go through the timing of relevant events. In Sweden, the typical student graduates from junior high school in June of the year the student turns 16.[4] The typical student then starts high school in August of that same year. In high school, students choose between different tracks that normally take three years to complete. Thus, the typical student graduates from high school in the summer of the year the student turns 19.[5] In the fall of that same year, students (typically) have their first opportunity to register for tertiary studies and in the spring of the following year, they have the second opportunity and so on.

The elections that I consider in this paper take place in the fall of the year when the relevant students turn 18, i.e. when they have completed about two thirds of their high school education. I assume that the right to vote starts to matter about one year before the election, when the election campaign begins to unfold. Thus, conditional on this assumption, students can be affected by the right to vote from the second year of high school and onward. The

[3] Butt (2004) shows that high levels of factual political knowledge are associated with a better ability to identify the policy positions of political parties. Thus, individuals with high levels of factual political knowledge should be better at finding parties with opinions that match their own beliefs.

[4] Junior high school is compulsory in Sweden.

[5] It is not compulsory to take part in high school education, but over time it has become more and more common and today it is considered necessary for a successful entrance on the labor market. In the main sample of this paper, about 70% have graduated from high school.

outcome variables in this paper are typically measured at the end of high school or somewhat later (i.e. around age 19).

2.3.2 Main measure: High school grade in Social Studies

The main measure of political knowledge around age 18 in this paper is the high school grade in Social Studies (*Samhällskunskap*). This subject has 8 objectives (see Table A1 in Appendix A). Two of these objectives are directly linked to political knowledge. According to these two objectives, the students should (1) understand how the political system works (at different levels) and (2) know how they can influence the decisions within this system (these objectives correspond to points 3 and 4 in Table A1). In addition, several of the other objectives contain components that are at least related to political knowledge (see objectives 1, 2 and 8 in Table A1).

Thus, at least 25% of the grade in Social Studies should be based on students' competence in political knowledge. Political knowledge is therefore measured with noise which should reduce precision, but not bias the estimates. On the other hand, the high school grade in Social Studies is readily available, which means I have access to a large number of observations. While studies using survey-based measures of political knowledge typically have quite few observations (see Chan and Clayton [2006] and Wagner, Johann and Kritzinger [2012]), I can use nationwide registers to collect data on high school grades for all individuals who graduated from high school. Consequently, this paper uses a measure that should be affected by students' level of political knowledge, that is set around age 18 and that is non-missing for a large share of each cohort in modern times.

2.3.3 Issues with the high school grade in Social Studies

A problem with high school grade in Social Studies is that the grade was potentially set too early for the later cohorts (students who graduated in the years 1997–2010). Students who graduated from high school in 1996 at the latest received a final grade in the *subject* of Social Studies at the time of graduation. In principle, they could affect their grade up to that point, but in practice, it was probably determined somewhat earlier. Thus, since most of the grade was determined after the point when the right to vote (presumably) started to matter, it should be possible to detect a potential voting right effect. Students graduating during the years 1997–2010, on the other, hand were graded in *courses* rather than in *subjects* and when a course was finished (which could be at the end of the first year) the grade could no longer be influenced. Subsequently, a lot of the grades were set long before the actual graduation. During this time, there were three courses in Social Studies: Social Studies A, Social Studies B and Social Studies C. All students (no matter what track) took Social Studies A, since it was mandatory, while (typically) only students in the Social Science track took Social Studies B and C. For this period, the grade in Social Studies A functions as the main measure of knowledge in Social Studies, while Social Studies B and C are used for robustness checks (the B and C courses were taken during the second and third years of high school). A potential problem with using the grade in Social Studies A as a measure of political knowledge is that some students took the course quite early and received the grade after their first year of high school. It is unlikely that having the right to vote influenced those early grades. However, there was significant variation in the timing of Social Studies A across tracks and across schools, making it a relevant measure for some students, but not for others.[6]

In the empirical exercises in Section 5, this late period (1997–2010) is included in the main results, but I also look at the early period separately to tackle the problem of the timing of the measurement. I also examine the late period separately with a focus on the courses Social Studies B and C, since these were taken at the end of high school and thus have better timing. Since the high school grade in Social Studies is not a perfect measure of political knowledge, I also investigate alternative measures of political knowledge and/or interest in society, which I elaborate on in the next section.

2.3.4 Alternative measures

The first alternative measure of political knowledge and/or interest in society is performance on the General Knowledge section of the Swedish Scholastic Assessment Test (SweSAT). The SweSAT is a kind of university admissions test in Sweden given twice a year (each spring and fall). However, it is not compulsory. University admission is primarily based on high school GPA, but for some admissions, students can compete with their SweSAT score. Students (and the rest of the population) can take the SweSAT at any time, but it is uncommon to take it before high school. In this paper I consider tests that were taken the year a student turned 18, 19 or 20 (i.e. when the student could be affected by having the right to vote). If a student took the test multiple times in this period, I keep the first test result. The General Knowledge section was in the test in the years 1977–1995 and consisted of 30 multiple choice questions aimed at capturing general knowledge of society. The questions could, e.g., be about politics, culture and sports (Stage [1985] estimates that about 15% of the questions can be categorized as political knowledge). The main advantage with this measure compared with the high school grade in Social Studies is that it measures factual knowledge which, conceptually, is how political knowledge is defined.

The second alternative measure is really a measure of political interest. I assume, to a greater extent than others, politically interested individuals participate in tertiary studies in the core subjects of Social Science at a young age (before the year they turn 22). These core subjects are: Economics, Polit-

[6] The data does not allow me to see which students took the course early and which students took it late.

ical Science, Economic History, Peace and Development Studies and Social and Economic Geography. I also assume that, to a greater extent than others, politically interested individuals have university degrees in subjects in Social Science. Due to data constraints, a broader definition of Social Science is used for the degrees.

The overall objective with all the discussed measures is to capture different traces of a potential increase in political knowledge and/or civic interest driven by the right to vote around or somewhat after the time of the election.

3 Institutional setting and data

In this section, I present details on the Swedish voting system and provide a list of Swedish elections during the period 1988–2006. I also describe some important details of the sample restrictions and end with a discussion of the data.

3.1 Institutional setting: Elections in Sweden

In Sweden, citizens are eligible to vote if they are 18 years of age on the day of the election at the latest. After several changes in the voting eligibility law in the 1960s, the above law was passed in 1975 and has since been in force (The Swedish Election Authority [2015]). This means that all elections from 1976 onward generate local experiments for individuals close to the eligibility threshold.

The study uses rich Swedish register data with access to complete birth date information for all individuals born in Sweden from 1969 onward. This property of the data allows me to perform regression discontinuity analyses with exact date of birth as the running variable[7] for all elections from 1988 onward. Consequently, data limitations rule out using elections prior to 1988. At the other end of the timeline, I am restricted by the fact that I can only observe register data up to and including the spring semester of 2010. Thus, I can only use elections prior to 2010 (which means 2006 is the last Swedish Parliamentary election). In Table 1, I provide a list of all Swedish elections during the period 1988–2006.

The list contains 6 Swedish Parliamentary elections, 3 European Parliamentary elections and 2 referendums. As we see in Table 1, the elections to the European Parliament generally have substantially lower turnout rates than the other elections. This effectively means they are low-status elections and unlikely to stimulate youth civic engagement and knowledge acquisi-

[7] In RD terminology, the variable that determines whether an individual is treated or not is often called the *running variable*. In this case, date of birth determines whether an individual has the right to vote or not. The term running variable will be used throughout the study.

tion. Thus, elections to the European Parliament will not be included in this study. Additionally, in order for the age at the first voting opportunity to be substantially different between the individuals on the two sides of the cut-offs, I only include elections where the subsequent election takes places at least one year later (this restriction is not relevant if the subsequent election is to the European Parliament).

In summary, as Table 1 indicates, the following seven elections are used in this paper: Swedish Parliament 1988, Swedish Parliament 1991, referendum on EU 1994, Swedish Parliament 1998, Swedish Parliament 2002, referendum on the euro 2003 and Swedish Parliament 2006.

Table 1. Elections in Sweden 1988–2006

Type of election	**Date of election**	**Birth cutoff date**	**Turnout rate**	**Used in the study**
SP election	09/18/1988	09/18/1970	86.0%	Yes
SP election	09/15/1991	09/15/1973	86.7%	Yes
SP election	09/18/1994	09/18/1976	86.8%	No
Referendum[a]	11/13/1994	11/13/1976	83.3%	Yes
EP election	09/17/1995	09/17/1977	41.6%	No
SP election	09/20/1998	09/20/1980	81.4%	Yes
EP election	06/13/1999	06/13/1981	38.8%	No
SP election	09/15/2002	09/15/1984	80.1%	Yes
Referendum[b]	09/14/2003	09/14/1985	82.6%	Yes
EP election	06/13/2004	06/13/1986	37.9%	No
SP election	09/17/2006	09/17/1988	81.9%	Yes

Notes: SP= Swedish Parliament. EP=European Parliament. a=EU, b=EURO. Information on all elections comes from Statistics Sweden (2015).

3.2 Data: Sample restrictions

To avoid potential problems with comparing individuals belonging to different school cohorts, I only study individuals who turn 18 during one of the seven election years. Adopting this restriction implies that the “controls” with respect to the referendum in 1994 are those born from November 14, 1976 to December 31, 1976. This interval consists of 48 days. Note that no other election takes place later in the year; consequently, the interval of 48 days can be used throughout the study, which creates consistency across elections. Thus, each of the seven elections will have a maximum treatment period of 48 days (the 47 days directly before the eligibility cutoff date and the eligibility cutoff date itself), as well as a maximum control period that consists of the 48 days directly following after the eligibility cutoff date. Table 2 clarifies the birthdates that have been used to construct the sample of treated and the sample of controls.

Table 2. Samples of treated and controls

Election	Treated	Controls
SP 1988	08/02/1970 to 09/18/1970	09/19/1970 to 11/05/1970
SP 1991	07/30/1973 to 09/15/1973	09/16/1973 to 11/02/1973
Referendum 1994	09/27/1976 to 11/13/1976	11/14/1976 to 12/31/1976
SP 1998	08/04/1980 to 09/20/1980	09/21/1980 to 11/07/1980
SP 2002	07/30/1984 to 09/15/1984	09/16/1984 to 11/02/1984
Referendum 2003	07/29/1985 to 09/14/1985	09/15/1985 to 11/01/1985
SP 2006	08/01/1988 to 09/17/1988	09/18/1988 to 11/04/1988

Notes: The period consists of 48 days before and after the cutoff dates.

The birthdate of each individual has then been normalized through the subtraction of the respective cutoff dates, generating seven election-specific, normalized birthdates.[8] Only observations with a normalized birthdate within the interval [-47,48] stay in the data. A problem with using the exact date of birth in this context is that previous evidence (see Dickert-Conlin and Elder[(2010]) indicates that there are systematic differences between weekdays with respect to births, i.e. children born on weekends tend to be somewhat different from children born on weekdays (with respect to parental characteristics). To tackle this issue, I aggregate up the data to full weeks. Individuals with a normalized birthdate in the interval [-6,0] will then belong to week 0 and individuals with a normalized birthdate in the interval [1,7] will belong to week 1. Given that I kept individuals with a normalized birthdate in the range [-47,48], the week numbers will go from -6 to 7.[9] Within these boundaries, the bandwidth will be varied in the empirical exercises in Section 5.

3.3 Data: Descriptive statistics

Table 3 describes some descriptive statistics for the two samples specified in Table 2. Overall, the number of individuals born on the dates specified in Table 2 is 172,283. The treated individuals (i.e. those who have the chance to vote just after turning 18) amount to 87,977 and the controls consist of 84,306 individuals. The fact that the sample contains more treated individuals is expected, considering that the number of births is typically lower at the end of a year. As we can see in Table 3, the treated individuals get their first opportunity to vote in a first-order election about three years before the controls (age 18 compared to age 21) which is arguably a substantial difference. From previous evidence it is known that individuals born earlier in the year normally have stronger educational outcomes. This is confirmed in the data

[8] A normalized birthdate of 0 thus indicates that the individual turned 18 on the day of the election. A non-positive number thus indicates eligibility in the election and vice versa.
[9] Note that weeks -6 and 7 only contain 6 days. In the empirical analysis I also present results using the exact normalized birthdate as the running variable to the test robustness of the results. Weekday indicators are then included in the empirical model.

by looking at the junior high school GPA (standardized by graduation year) which is a predetermined covariate.[10] Treated individuals have a higher GPA on average and fewer of them lack[11] information on this variable, indicating higher educational ambition already before the treatment (i.e. the opportunity to vote in a first-order election). The same pattern can be seen for the junior high school grade in Social Studies which constitutes a predetermined measure of political knowledge. There are generally small differences between the two groups with respect to parental characteristics.[12]

Looking at the outcome variables at the bottom of Table 3 we see that treated individuals have a somewhat better high school grade in Social Studies[13] and that they have a slightly higher score on the General Knowledge section in the SweSAT. This is, of course, consistent with a positive treatment effect, but it is clear that a simple comparison of outcomes between the two groups is confounded by unbalanced covariates (i.e. the junior high school grades). This problem can be solved by comparing individuals that are infinitely close to the threshold that separates the two groups, and in the next section I describe this research design in detail. When it comes to the tertiary education outcomes at the very bottom of Table 3, the treated and controls exhibit identical summary statistics.

[10] I assume that the right to vote starts to matter approximately one year before the election, when the election campaign begins to unfold. Thus, conditional on that assumption, all variables that are determined before that point in time can be considered predetermined (i.e. they should be unaffected by the treatment).

[11] I do not have access to junior high school data for the individuals who turned 18 in in 1988.

[12] For those who turned 18 in 1998, 2002, 2003 or 2006, the parental characteristics are measured at age 15. For those who turned 18 in 1988, the parental characteristics are measured at age 20. For those who turned 18 in 1991, the parental characteristics are measured at age 17. For those who turned 18 in 1994, the parental characteristics are measured at age 18. This is due to data constraints. Even if the parental characteristics are measured after age 15, for some individuals, they are arguably good proxies for the parental characteristics at age 15 since it is unlikely that the child's right to vote affects the parents.

[13] To have data on this variable, the individuals must have a grade in Social Studies and they must graduate from high school during the year they turn 18, 19 or 20.

Table 3. Descriptive statistics for the most widely used sample

Characteristics	Data exists	All	Voting right=1	Voting right=0
Demographics				
Male	1	0.514	0.514	0.513
Age at first major voting opportunity	1	19.532	18.065	21.064
Predetermined variables				
JHS Overall GPA	0.812	-0.038	-0.026	-0.050
JHS grade in Social Studies	0.812	-0.027	-0.018	-0.037
Mother tertiary education ≥ 3 years	0.982	0.135	0.135	0.135
Father tertiary education ≥ 3 years	0.945	0.142	0.143	0.142
Mother in work	0.983	0.856	0.858	0.854
Father in work	0.951	0.879	0.879	0.880
Outcome variables				
HS grade in Social Studies	0.677	-0.015	-0.009	-0.020
SweSAT (General Knowledge)*	0.084	-0.248	-0.238	-0.258
Early registration Social Science**	0.847	0.017	0.017	0.017
Tertiary education ≥ 3 years in Social Science***	0.539	0.021	0.021	0.021
Observations	172,283	172,283	87,977	84,306

Notes: JHS=Junior High School. HS=High School. All grade measures are standardized within graduation year. The SweSAT score is standardized within a given test. * Only individuals who turned 18 in 1988, 1991 or 1994 can have data on this variable. ** Individuals who turned 18 in 1988 cannot have data on this variable due to data constraints. *** Only individuals who turned 18 in 1988, 1991, 1994 or 1998 can have data on this variable. This is due to the fact that it is measured in 2009 and cannot be considered a completed education for the later cohorts.

4 Empirical specification

I employ a regression discontinuity design on the data, using birth cutoff dates for voting eligibility to create exogenous variation in the voting age. Under the assumption that births take place randomly around the birth cutoff dates for voting eligibility, this method allows me to estimate the causal effect of early age voting right on the relevant outcomes. The normalized week number is used as the running variable and it is denoted by Z_{ie} where *e* indicates the specific election.[14] Equation (1) specifies the baseline empirical model used in the paper:

$$Y_i = \beta_0 + \beta_1 I[Z_{ie} \leq 0] + \beta_2 Z_{ie} + \beta_3 I[Z_{ie} \leq 0] * Z_{ie} + \delta_e + u_i \tag{1}$$

[14] In this setting we have e=1988, 1991, 1994, 1998, 2002, 2003 and 2006. For some outcomes a subset of these elections are used.

The outcome, denoted Y_i, is a measure aimed at capturing political knowledge and/or political interest. δ_e captures election-specific fixed effects and u_i is an error term. I include separate linear terms on the two sides of the threshold (i.e. $\beta_2 Z_{ie}$ and $\beta_3 I[Z_{ie} \leq 0] * Z_{ie}$) and β_1 captures the difference between the two linear terms as they approach the threshold from below and above respectively. Under the assumption that the underlying confounders are continuous in the running variable (i.e. Z_{ie}) around the threshold, this difference (i.e. β_1) corresponds to the causal effect of early voting eligibility on the relevant outcome (see, e.g., Lee and Lemieux [2010] for a detailed discussion of the identifying assumptions underlying the typical RD design). Essentially, this procedure amounts to running separate local linear regressions on both sides of the threshold and comparing the values of those regressions at the cutoff. The specific data window (bandwidth) within which this is performed should be varied to test the robustness of the results and this is done in Section 5.2.[15] Following the advice of Lee and Card (2008) on the appropriate choice of standard errors when using a discrete running variable, I cluster the standard errors on the week-times-election level.

5 Results

In this section, I first examine the research design to make sure that the underlying identifying assumption is valid. I then present the main results.

5.1 Tests of the identifying assumption

The underlying identifying assumption that is required to make causal interpretations of the treatment effect in this context is that all potential confounders are continuous in the running variable (i.e. the normalized birth week Z_{ie}) across the threshold. This assumption amounts to requiring random assignment to treatment in the immediate proximity of the threshold. Typically, this assumption is tested by investigating whether the running variable evolves smoothly over the cutoff and whether predetermined variables (i.e. variables that were determined before the elections) are continuous around the cutoff conditional on the empirical model, i.e. Equation 1 (see Lee and Lemieux [2010]). Mass points in the running variable close to the threshold raise concerns about the manipulability of the running variable, and jumps in the predetermined variables at the cutoff are yet another indication of manipulation.

[15] The possible windows are: [0,1], [-1,2] … [-6,7]. However, it is considered good practice to have at least four unique values of the forcing variable below the cutoff and four unique values above the cutoff (Schochet et al. [2010]).

Before going into the empirical exercises, let us first consider the running variable and the cutoffs used in this paper. Since the cutoffs refer to voting eligibility birthdate thresholds for elections that take place 18 years later, it seems highly unlikely that parents are able to manipulate births to take place before rather than after a cutoff.[16] Even if the birthdate cutoffs were perfectly predictable, we would arguably not expect strategic timing of births at these cutoffs. Dickert-Conlin and Elder (2010), e.g., do not even find strategic timing of births at thresholds that determine age at school start which is an outcome that arguably carries more significance than early age voting eligibility. Thus, unlike most other RD settings where being on the right side of a threshold is associated with substantial personal benefits (e.g. access to certain schools and eligibility for social benefits), the context in this paper offers small incentives for manipulation of the running variable.

Since the main measure in this paper is the high school grade in Social Studies, the pre-tests are performed on the sample with valid information on this variable (67.7% of the total sample, see Table 3). Figure 1 shows the distribution of the running variable around the voting eligibility birthdate cutoffs. There is no indication that an abnormally high number of individuals were born in the weeks just before the cutoffs (i.e. in weeks -1 and 0). Instead, the number of births during the weeks around the cutoffs evolves smoothly. This visual impression is also confirmed by the McCrary density test, which delivers an insignificant result.[17]

[16] Anecdotally, parents do not behave in that way, and if they did, it would be a difficult practice since election legislation tends to change over time. Examples of such changes can be the time between the elections and the voting age. Thus, it is virtually impossible to predict the timing of elections far in the future.

[17] The p-value is 0.473. The null hypothesis is that the discontinuity in the density of the running variable at the cutoff is zero. See McCrary (2008) for a detailed description of the test.

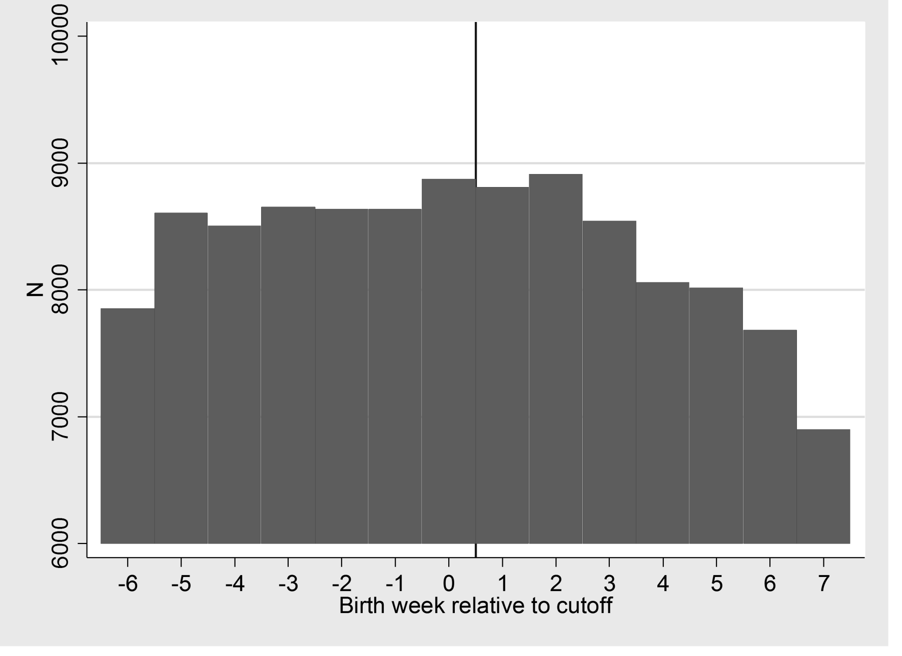


Figure 1. Distribution of the running variable

Notes: The figure is based on individuals with valid information on the high school grade in Social Studies. The sample amounts to 116,713 observations.

As a next step, I examine whether any of the predetermined variables (see Table 3) behave strangely at the cutoff. Figure 2 shows the relationship between the running variable and the junior high school grade in Social Studies.[18] I choose to highlight this relationship since the junior high school grade in Social Studies is a predetermined measure of political knowledge which is the main outcome in this paper and thus a key predetermined variable. As expected, there is a negative relationship between the grade and the running variable, since individuals born late in the year typically have worse educational outcomes. Individuals who can vote shortly after turning 18 (i.e. with non-positive values on the running variable) generally have a higher grade, but importantly, the grade evolves smoothly over the cutoff, suggesting that individuals near the cutoff are similar with respect to the junior high school grade in Social Studies. Similar graphs of all predetermined variables can be found in Figure B1 in Appendix B, and they generally paint a similar picture.

[18] The fact that the standardized grade lies around 0.2 is due to the fact that only individuals who have graduated from high school are studied here. These individuals generally have good pre-high school educational outcomes.

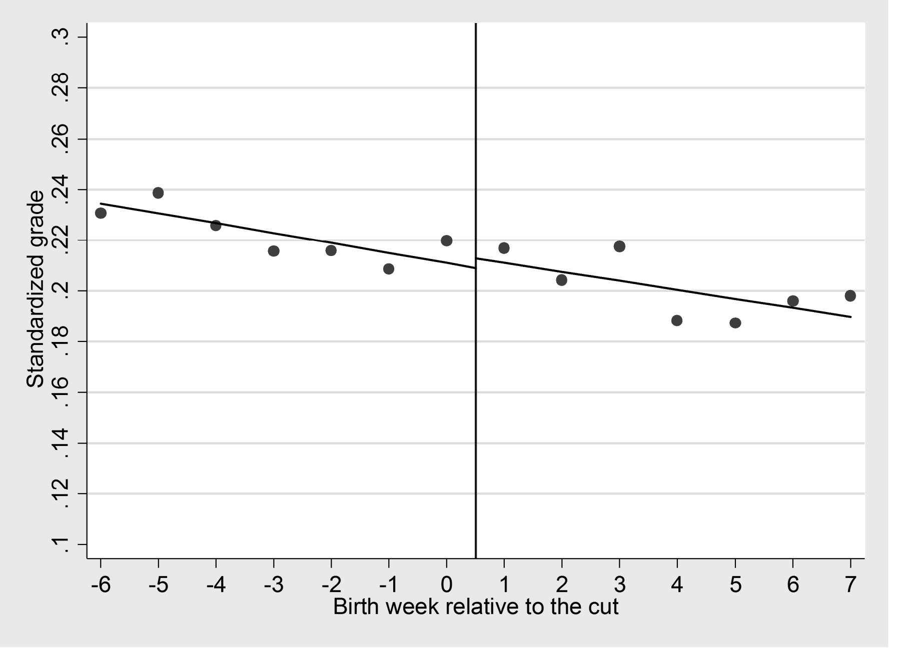


Figure 2. Relationship between JHS grade in Social Studies and the running variable

Notes: The figure is based on individuals with valid information on the high school and junior high school grade in Social Studies. The sample amounts to 102,191 observations.

In Table 4, the predetermined variables are given a more formal treatment within a regression framework. All estimates come from the model specified in Equation (1) and the full bandwidth is used (i.e. the running variable goes from -6 to 7). The estimates are generally small and insignificant and thus consistent with the identifying assumption of random assignment to treatment at the thresholds. Importantly, as can be seen in the bottom of Table 4, the estimates exhibit joint insignificance, which further strengthens the validity of the research design. Table B1 in Appendix B shows estimates from regressions using alternative specifications of the empirical model (i.e. smaller bandwidth and quadratic control for the running variable). The results from these alternative specifications do not substantially deviate from the estimates shown in Table 4, which provides additional evidence of the robustness of the empirical strategy.

Table 4. RD estimates for the predetermined variables

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	JHS total GPA	JHS grade Social St.	Mother highly edu.	Father highly edu.	Mother employed	Father employed
Voting right	-0.0064	-0.0037	-0.0001	0.0043	-0.0031	-0.0050
	(0.0095)	(0.0117)	(0.0052)	(0.0055)	(0.0041)	(0.0037)
Mean of dep.	0.241	0.212	0.166	0.175	0.882	0.901
Observations	102,191	102,191	115,321	111,904	115,405	112,437
	p-value for the test of joint significance of the estimates:0.32					

Notes: The results are based on individuals with valid information on the high school grade in Social Studies and non-missing data on the relevant variable. JHS=Junior High School.

To make sure there are no confounding elements at the cutoffs because of potential seasonal variations, I have also constructed fake cutoffs. These cutoffs are constructed to take place either the year before or the year after the real cutoffs.[19] I then estimate the effect of being on the treatment side of these cutoffs on the high school grade in Social Studies (which constitutes the main outcome) conditional on the empirical model. As expected, the results, which are presented in Table 5, do not include any significant estimates. The estimates are somewhat sensitive to the choice of empirical specification, but generally the results hover around 0. This is reassuring for the research design since it means that any potential true treatment effects should be due to the voting right and not to any confounding seasonal factors.

[19] I have constructed fake birthdate cutoffs in 1969 (real 1970), 1972 (real 1973), 1975 (real 1976), 1979 (real 1980), 1983 (real 1984), 1986 (real 1985) and 1987 (real 1988). The fake cutoffs are set on the same weekday in the same week as the real cutoffs (thus not necessarily on the same date).

Table 5. RD estimates for the high school grade in Social Studies (fake cutoffs)

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Voting right	-0.0021	0.0025	0.0104	0.0148	0.0260	0.0054	0.0000
	(0.0133)	(0.0142)	(0.0143)	(0.0155)	(0.0252)	(0.0117)	(0.0112)
Mean of dep.	-0.0096	-0.0083	-0.0090	-0.0083	-0.0096	-0.0096	-0.0096
Observations	108,518	99,252	83,173	67,180	108,518	108,518	108,518
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variables	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

5.2 Main results

This section contains results from regression discontinuity analyses of three different groups of outcomes all aimed at capturing different aspects of political knowledge and/or civic interest. I discuss the high school grade in Social Studies, the score on the General Knowledge section of the SweSAT and the orientation of tertiary studies in turn.

A substantial part of the main sample lacks data on the high school grade in Social Studies and the score on the General Knowledge section of the SweSAT, and to make sure that there is no selection into having a grade or a score at the cutoff, I have investigated the effect of the right to vote on indicators for having a grade or a score. These results are presented in Table B2 in Appendix B. Some of the estimates for the high school grade in Social Studies are marginally significantly positive, but this result is not robust to reducing the bandwidth or including predetermined variables in the model. Overall, the analysis suggests that we cannot reject the null hypothesis of no selection into having a grade or a score. Any difference in the actual outcomes between the individuals on the respective sides of the cutoff should therefore be attributed to the voting right and not to different probabilities of having a non-missing outcome.

5.2.1 High school grade in Social Studies

I start the presentation of the actual results with investigating potential early age voting right effects on the high school grade in Social Studies, which constitutes my main measure of political knowledge around age 18.[20] Figure 3 shows the relationship between the high school grade in Social Studies and the running variable. Consistent with the picture in Figure 2, we see a negative relationship between the grade and the running variable. Individuals who had the right to vote just after turning 18 have a higher grade than the controls on average, but treated individuals who are right at the threshold actually have lower grades than the corresponding controls. The negative estimate is, however, small (about 1.3% of a standard deviation) and insignificant. Importantly, there is no evidence at all of a positive jump in the outcome at the threshold, which goes against the hypothesis of voting right induced increases in political knowledge (see Wagner, Johann and Kritzinger [2012]).

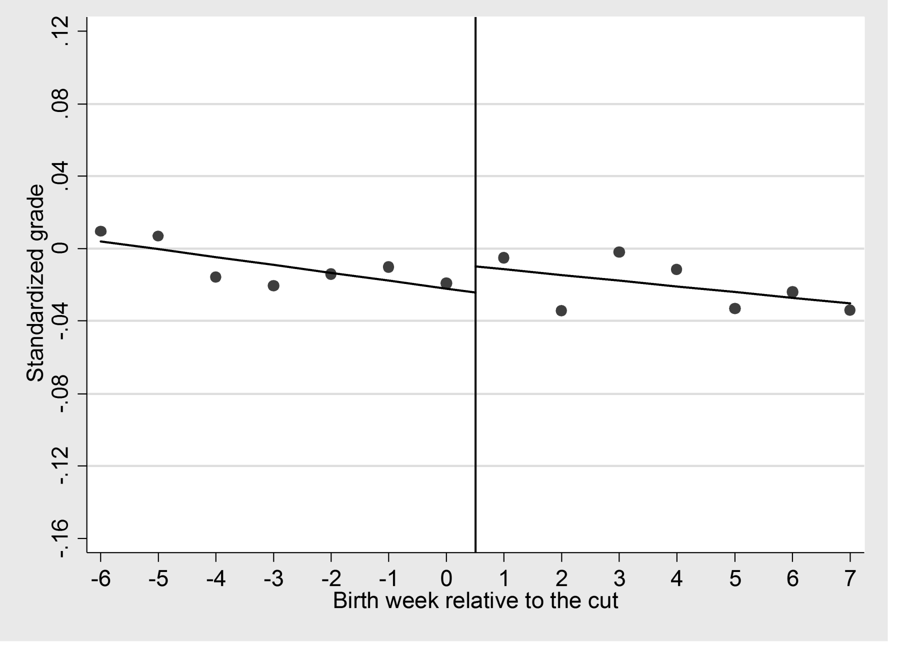


Figure 3. Relationship between HS grade in Social Studies and the running variable

Notes: The figure is based on individuals with valid information on the high school grade in Social Studies. The sample amounts to 116,713 observations.

Table 6 (panel A) provides formal estimates of the effect of having the opportunity to vote at a relatively young age on the high school grade in Social Studies, conditional on seven different empirical specifications where specification (1) corresponds to the difference at the cutoff in Figure 3. In col-

[20] For those who graduated in 1996 at the latest, the measure is based on the grade in the *subject* Social Studies. For those who graduated in 1997–2010, the measure is based on the grade in the *course* Social Studies A.

umns (2–4), the bandwidth is gradually reduced, which leads to a lower estimate in absolute terms. It actually becomes positive in column (4), where I only use four data points on each side of the cutoff. In column (5), I introduce quadratic controls for the running variable which also gets the estimate closer to zero compared to specification (1). In column (6), I include the predetermined variables which increase the precision but only marginally affect the estimate relative to specification (1). In column (7), I use exact date of birth as the running variable instead of the week variable and control for potential differences between the days of the week by including weekday fixed effects. The estimate is virtually the same as in specification (1). Since the estimates from these different specifications are generally on the negative side and never substantially positive, they suggest that there is no positive voting right effect on political knowledge around age 18.

The table also includes alternative measures of the grade in Social Studies (see Section 2.3.3 for a discussion of the timing problem of the grade). In panel B, I focus on the early period (i.e. individuals born in 1970, 1973 or 1976) when the individuals were given a final grade in the *subject* Social Studies. In panels C and D, I focus on the late period (i.e. individuals born in 1980, 1984, 1985 or 1988) when the students could take advanced courses in Social Studies. These courses (i.e. Social Studies B and C) were taken at the end of high school and are thus measured at a good time relative to the voting opportunity.

Panel B gives a similar picture to panel A, but the estimates are generally more negative and in fact never on the positive side. The effect is of a magnitude of about 2% of a standard deviation and it is quite robust to changes in the empirical model. Again, there is absolutely no support for a positive voting right effect. Panels C and D report similar results with the estimates firmly on the negative side. The estimates in column (5), the model with predetermined variables included, are actually significantly negative. But it should be noted that relatively few students take these courses, which leaves us with substantial uncertainty.

The overall message to take away from Table 6 is that we generally find negative but insignificant effects from the right to vote on political knowledge. This suggests that positive effects of practically relevant magnitudes can be ruled out.

Table 6. RD estimates for the high school grade in Social Studies

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Panel A: High school grade in Social Studies, whole period							
Voting right	-0.0131	-0.0094	-0.0058	0.0036	0.0010	-0.0095	-0.0141
	(0.0113)	(0.0121)	(0.0127)	(0.0141)	(0.0164)	(0.0083)	(0.0114)
Mean of dep.	-0.0147	-0.0153	-0.0167	-0.0149	-0.0147	-0.0147	-0.0147
Observations	116,713	101,955	85,668	69,140	116,713	116,713	116,713
Panel B: High school grade in Social Studies, early period							
Voting right	-0.0155	-0.0110	-0.0281	-0.0224	-0.0276	-0.0124	-0.0178
	(0.0181)	(0.0199)	(0.0221)	(0.0228)	(0.0272)	(0.0156)	(0.0178)
Mean of dep.	-0.0310	-0.0330	-0.0339	-0.0303	-0.0310	-0.0310	-0.0310
Observations	44,968	39,222	32,986	26,511	44,968	44,968	44,968
Panel C: High school grade in Social Studies B							
Voting right	-0.0437*	-0.0456	-0.0198	-0.0485	-0.0279	-0.0636***	-0.0416
	(0.0260)	(0.0287)	(0.0316)	(0.0303)	(0.0389)	(0.0201)	(0.0338)
Mean of dep.	0.0105	0.0118	0.0150	0.0101	0.0105	0.0105	0.0105
Observations	12,638	11,060	9,308	7,598	12,638	12,638	12,638
Panel D: High school grade in Social Studies C							
Voting right	-0.0752	-0.0738	-0.0466	-0.0715	-0.0558	-0.0786**	-0.0746
	(0.0486)	(0.0517)	(0.0565)	(0.0600)	(0.0793)	(0.0363)	(0.0469)
Mean of dep.	0.0160	0.0149	0.0183	0.00978	0.0160	0.0160	0.0160
Observations	6,440	5,636	4,740	3,877	6,440	6,440	6,440
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variabl.	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

5.2.2 Score on the General Knowledge section of the SweSAT

The score on the General Knowledge section of the SweSAT offers the possibility of capturing the studied individuals' levels of factual knowledge about society. In that respect, the SweSAT General Knowledge score resembles the definition of political knowledge given by Delli Carpini and Keeter (1996) (see Section 2.2), but it has a wider scope with questions that, besides politics, encompass e.g. news, culture and sports. Stage (1985) estimates that about 15% of the questions in the test are related to political knowledge. The General Knowledge section of the SweSAT was removed after 1995 and subsequently, only the three early elections (i.e. the elections in 1988, 1991 and 1994) can be used to study this outcome.[21]

In Table 7, I provide estimates of the effect of having the opportunity to vote shortly after turning 18, relative to 21, on the SweSAT score in General Knowledge. In column (1), I present the baseline estimate (the same kind of estimate as in Figure 3, but this time for the SweSAT score in General Knowledge). It is negative and amounts to about 3.5% of a standard deviation. When the bandwidth is reduced (columns [2–4]) the estimates come closer to zero relative to the baseline estimate. And when quadratic controls for the running variable are introduced in column (5), we obtain a positive estimate of about 3% of a standard deviation. Including predetermined variables (column [6]) makes the estimate more negative relative to specification (1), while using date of birth as the running variable (column [7]) just barely affects the estimate.

The estimates are generally on the negative side, but the results are clearly sensitive to the choice of empirical model. The conclusion can therefore not be that there is a robust negative effect, but rather that it is unlikely that there is a *positive* effect on General Knowledge measured around age 18 from having the opportunity to vote just after turning 18.

Table 7. RD estimates for the SweSAT score (General Knowledge section)

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Voting right	-0.0343	-0.0257	0.0062	0.0077	0.0310	-0.0524***	-0.0297
	(0.0260)	(0.0290)	(0.0282)	(0.0343)	(0.0378)	(0.0191)	(0.0320)
Mean of dep.	-0.248	-0.245	-0.239	-0.247	-0.248	-0.248	-0.248
Observations	14,538	12,685	10,657	8,480	14,538	14,538	14,538
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variables	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

5.2.3 Tertiary studies within Social Science

If political socialization during adolescence leads to an increased inclination toward civic engagement later in life, as Hart and Atkins (2011) argue, then

[21] I include tests that were taken during the year the individual turned 18, 19 or 20. Thus, for individuals who turned 18 in 1988, I include tests taken in 1988–1990. If the test was taken multiple times, I keep the result from the first test. Due to data constraints, I can only include tests taken in 1994 and 1995 for individuals who turned 18 in 1994.

we would expect early age voters to choose tertiary education programs within Social Science to a greater extent than comparable individuals whose first chance to vote takes place 2–3 years later (when the choice of tertiary education might have already been made). In Table 8, I study two measures of Social Science involvement at university.

The outcome in panel A is an indicator for early registration in a tertiary course within the core subjects of Social Science. Early registration is defined as before the year an individual turns 22, and the core subjects are: Economics, Political Science, Economic History, Peace and Development Studies and Social and Economic Geography. The baseline estimate in column (1) is slightly negative and it is virtually unaffected by the variations in the empirical model to which it is subjected in columns (2–7). In fact, all estimates are on the negative side and they are all insignificant. A 95% confidence interval does include positive estimates, but at a magnitude that can hardly be seen as practically relevant.

The outcome in panel B is an indicator for having completed a tertiary education of at least three years within Social Science. Due to data constraints, Social Science here is more broadly defined than in panel A, and in addition to the subjects mentioned above, it also includes subjects like Sociology and Psychology. The education level is measured in 2009 and since it is supposed to capture completed education, I only study individuals who turned 18 in 1988, 1991, 1994 or 1998. The overall picture is similar to the one in panel A, with the estimates consistently on the negative side. Since the estimate is more or less the same across all seven specifications, it seems unlikely that any reasonable model can generate substantial positive estimates. Therefore, I interpret the collected evidence from this exercise as suggesting that adolescents are virtually unaffected with respect to political/civic interest by having the opportunity to vote shortly after turning 18.

Table 8. RD estimates for tertiary studies within Social Science

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Panel A: Early registration (before age 22) in Social Science							
Voting right	-0.0011	-0.0013	-0.0004	-0.0001	-0.0001	-0.0011	-0.0012
	(0.0012)	(0.0013)	(0.0014)	(0.0016)	(0.0019)	(0.0012)	(0.0014)
Mean of dep.	0.0170	0.0170	0.0171	0.0170	0.0170	0.0170	0.0170
Observations	145,995	127,447	106,948	86,203	145,995	145,995	145,995
Panel B: Tertiary education of at least 3 years within Social Science							
Voting right	-0.0028	-0.0021	-0.0017	-0.0019	-0.0014	-0.0029	-0.0031*
	(0.0019)	(0.0021)	(0.0023)	(0.0027)	(0.0033)	(0.0019)	(0.0017)
Mean of dep.	0.0210	0.0208	0.0206	0.0204	0.0210	0.0210	0.0210
Observations	92,896	81,097	68,053	54,786	92,896	92,896	92,896
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variabl.	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable

5.3 Heterogeneity analyses

Some individuals might be more affected by having the right to vote shortly after turning 18 than others. Experiences from Swedish elections clearly show that females have higher turnout rates than males among young voters. Another robust finding is that individuals with highly educated parents have higher turnout rates than individuals with parents with low levels of education.[22] Individuals with highly educated parents are probably also more likely to be informed about their voting right in the election at an early stage. Based on these observations, I perform a number of exercises separately for males and females and for individuals whose parents have high and low levels of education, respectively.[23] If anything there should be a higher probability of detecting positive voting right effects among women and among individuals with highly educated parents, since they are more likely to be affected by having the right to vote.

[22] See Sveriges Kommuner och Landsting (2009). The report is in Swedish and is written by the Swedish Association of Local Authorities and Regions.
[23] Individuals who have at least one parent with at least three years of tertiary education are defined as having highly educated parents.

I present the main results in Table 9 and robustness checks in Tables B3–B4 in Appendix B. For simplicity and ease of presentation, I restrict the analysis to two outcome variables: high school grade in Social Studies (panel A of Table 9 and Table B3) and SweSAT score on General Knowledge (panel B of Table 9 and Table B4). These two measures are most closely connected to the concept of political knowledge.

With respect to the high school grade in Social Studies (panel A of Table 9), the point estimates are negative for all groups and actually more negative for women and individuals with highly educated parents. Changing the empirical specification (e.g. reducing the bandwidth and including quadratic controls for the running variable) does not affect the estimates for women and individuals with highly educated parents (see Table B3). Thus, it is possible to rule out the existence of positive effects for these groups. For males and individuals without highly educated parents, the results are, however, more sensitive to the empirical model. Models with small bandwidths and with quadratic controls for the running variable generate positive estimates, which makes the conclusion less clear for these groups. But an average over the seven specifications suggests a zero effect.

The results for the SweSAT score on General Knowledge are more in line with the hypothesis considering that females have the least negative estimate (and several estimates for females in Table B4 are actually significantly positive). On the other hand, individuals with highly educated parents exhibit strongly negative estimates, which is hard to reconcile with the hypothesis. It should be noted that these estimates suffer from bad precision because of small sample sizes, which suggests that the results should be interpreted with caution. But at the very least, I think we can rule out the existence of positive effects for males and individuals with highly educated parents. The picture is less clear for females and individuals without highly educated parents, but it should be noted that specification (6) in Table B4, which arguably has the strongest identification because of the inclusion of predetermined variables, reports negative estimates.

In summary, no single group stands out as being particularly influenced by having the right to vote.

Table 9. Heterogeneity analyses (gender and family background)

Column:	(1)	(2)	(3)	(4)
Bandwidth:	[-6,7]	[-6,7]	[-6,7]	[-6,7]
Group:	Males	Females	HE parents	Not HE parents
Panel A: High school grade in Social Studies				
Voting right	-0.0090	-0.0168	-0.0168	-0.0114
	(0.0158)	(0.0141)	(0.0188)	(0.0113)
Mean of dep.	-0.143	0.0987	0.366	-0.143
Observations	54,735	61,978	29,365	87,348
Panel B: SweSAT score on General Knowledge				
Voting right	-0.0711	-0.0052	-0.0571	-0.0162
	(0.0477)	(0.0294)	(0.0372)	(0.0329)
Mean of dep.	-0.0159	-0.432	-0.0409	-0.371
Observations	6,435	8,103	5,432	9,106
Linear control	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No
Predet. variabl.	No	No	No	No
Day of birth RV	No	No	No	No
Weekday FE	No	No	No	No

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. HE=highly educated.

6 Conclusion

The typical voting age around the world is 18. However, during the last ten years, the appropriateness of this traditional voting age has been challenged. In 2007, Austria lowered the voting age to 16, and 16 and 17-year-olds were also allowed to vote in the Scottish independence referendum in 2014. These events were the results of and have contributed to a discussion of the advantages and disadvantages of lowering the voting age to 16. Proponents of a lower voting age have emphasized that 16 and 17-year-olds might very well respond by becoming more politically knowledgeable (see Wagner, Johann and Kritzinger [2012]) and interested (see Zeglovits and Zandonella [2013]) if they are allowed to vote. It has also been suggested that early age voting leads to deeper civic commitment in adulthood. This claim has appeared in both popular (see Brooks [2014] for the comments by Alex Salmond) and scientific contexts (see Hart and Atkins [2011]). If true, these arguments build a strong case for lowering the voting age to 16. However, given the current evidence, it is hard to validate these claims since they have not been analyzed in a controlled causal framework.

Therefore, in this paper I have used rich nation-wide Swedish register data to contribute with this kind of evidence. Swedish voting laws state that citizens are eligible to vote if they turn 18 on the day of the election at the latest. Utilizing the fact that the Swedish register data gives me access to the exact date of birth of all individuals born in Sweden from 1969 and onward, I can employ a regression discontinuity (RD) strategy to estimate the causal effect of having the first voting opportunity at 18 compared with having the first opportunity, on average, three years later, on measures of political knowledge, political interest and civic interest around age 18. As expected, there is no evidence of manipulation of the running variable at the thresholds that I use in the paper and therefore I cannot reject the null of random assignment to early voting opportunity at the voting eligibility birthdate cutoffs. Thus, I am able to causally identify the effects of a substantial reduction of the voting age, albeit on a somewhat higher general age level (i.e. 21 to 18 instead of 18 to 16), on political knowledge, political interest and civic interest in early adulthood.

Political knowledge and interest is mainly measured using the high school grade in Social Studies, but the score on the General Knowledge section in the SweSAT is also used as an alternative measure which captures a combination of political and civic interest and knowledge. Civic interest is also measured by examining the orientation of tertiary studies (i.e. are the individuals involved in Social Science studies?).

To the extent that my measures can be said to capture political knowledge, political interest and civic interest around and shortly after age 18, the collected evidence from the RD analyses show that individuals who had their first voting opportunity in a first-order election shortly after turning 18 *do not* exhibit higher levels of the aforementioned outcomes than comparable individuals whose first voting opportunity in a major election took place, on average, three years later. This finding strongly indicates that the putative benefits of a lower voting age, in terms of increased political and civic interest, are false or at least severely exaggerated. Thus, this argument should be used with extreme caution in the voting age debate. However, it is of course possible that school curricula would change if the voting age were 16 and that political science material would be introduced at an earlier age. In that case, 16-year-olds would probably become more politically knowledgeable following a decrease in the voting age, but that would be the result of changed teaching rather than a change in the motivation to learn more about politics among 16-year-olds.

Finally, even if the conclusion in this study rejects one of the key arguments for lowering the voting age, it does not imply that it is necessarily wrong to lower the voting age to 16. Rather, it suggests that 16-year-olds will not rise to the occasion and become more politically knowledgeable just because they are given the right to vote. There might still be a case for a voting age of 16, or even lower, on the basis of general human rights considerations as Wagner, Johann and Kritzinger (2012) suggest, and this discussion would surely benefit from further scientific contributions of an empirical as well as a theoretical nature.

References

Bhatti, Y., K. M. Hansen, and H. Wass. 2016. "First-time boost beats experience: the effect of past eligibility on turnout". *Electoral Studies*, 41, 151–158.

Brooks, L. (2014, September 23), "Alex Salmond calls for voting age to be lowered in time for general election", The Guardian, Accessed July 29th 2015. <http://www.theguardian.com/politics/2014/sep/23/alex-salmond-scottish-independence-referendum-exhilarating>.

Butt, S. 2004. "Political knowledge and routes to party choice in the British general election of 2001". *British Elections and Parties Review*, 14(1), 3–17.

Central Intelligence Agency. 2015. "The world factbook". Accessed July 30th 2015. <https://www.cia.gov/library/publications/the-world-factbook/>.

Chan, T. W. and M. Clayton. 2006. "Should the voting age be lowered to sixteen? Normative and empirical considerations". *Political Studies*, 54(3), 533–558.

Coppock, A. and D. P. Green. Forthcoming. "Is voting habit forming? New evidence from experiments and regression discontinuities". *American Journal of Political Science*.

Delli Carpini, M. X. and S. Keeter. 1996. "What Americans know about politics and why it matters". Yale University Press, New Haven.

Dickert-Conlin, S. and T. Elder. 2010. "Suburban legend: school cutoff dates and the timing of births". *Economics of Education Review*, 29(5), 826–841.

Hart, D., T. M. Donnelly, J. Youniss, and R. Atkins. 2007. "High school community service as a predictor of adult voting and volunteering". *American Educational Research Journal*, 44(1), 197–219.

Hart, D. and R. Atkins. 2011. "American sixteen-and seventeen-year-olds are ready to vote". *The ANNALS of the American Academy of Political and Social Science*, 633(1), 201–222.

Human Rights Advocates. 2009. "The right to vote: a basic human right in need of protection", Manuscript, Accessed October 13th 2015. <

http://www.humanrightsadvocates.org/wp-content/uploads/2010/05/The-Right-to-Vote-A-Basic-Human-Right-In-Need-of-Protection.pdf>.

Larcinese, V. 2007. "Does political knowledge increase turnout? Evidence from the 1997 British general election". *Public Choice*, 131(3–4), 387–411.

Lau, R. R., D. J. Andersen, and D. P. Redlawsk. 2008. "An exploration of correct voting in recent US presidential elections". *American Journal of Political Science*, 52(2), 395–411.

Lee, D. S. and D. Card. 2008. "Regression discontinuity inference with specification error". *Journal of Econometrics*, 142(2), 655–674.

Lee, D. S. and T. Lemieux. 2010. "Regression discontinuity designs in economics". *Journal of Economic Literature*, 48(2), 281–355.

McCrary, J. 2008. "Manipulation of the running variable in the regression discontinuity design: a density test". *Journal of Econometrics*, 142(2), 698–714.

Meirick, P. C. and D. B. Wackman. 2004. "Kids voting and political knowledge: narrowing gaps, informing votes". *Social Science Quarterly*, 85(5), 1161–1177.

Meredith, M. (2009). "Persistence in political participation". *Quarterly Journal of Political Science*, 4(3), 187-209.

Mondak, J. J. 2001. "Developing valid knowledge scales". *American Journal of Political Science*, 45(1), 224–238.

Scharpf, F. 1999. "Governing in Europe: effective and democratic?". Oxford University Press, Oxford.

Schochet, P., T. Cook, J. Deke, G. Imbens, J. R. Lockwood, J. Porter, and J. Smith. 2010. "Standards for regression discontinuity designs". Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.

Stage, C. 1985. "Group differences in test results. The significance of test item contents for sex differences in results on vocabulary and general knowledge". Doctoral dissertation at the Department of Education, University of Umeå.

Statistics Sweden. Accessed August 5th 2015. <http://www.scb.se>.

Swedish Association of Local Authorities and Regions. 2009. "Valdeltagande bland förstagångsväljare (Turnout among first time voters)". (in Swedish).

Syvertsen, A. K., M. D. Stout, C. A. Flanagan, D. L. Mitra, M. B. Oliver, and S. S. Sundar. 2009. "Using elections as teachable moments: a randomized evaluation of the student voices civic education program". *American Journal of Education*, 116(1), 33–67.

The Swedish Election Authority. 2015. Accessed August 5th 2015. <http://www.val.se/det_svenska_valsystemet/historik/index.html>.

The Swedish National Agency for Education. (2015). Accessed August 5th 2015. <http://www.skolverket.se/laroplaner-amnen-och-kurser/gymnasieutbildning/gymnasieskola/kursplaner-fore-2011/subjectKursinfo.htm?subjectCode=SH&courseCode=SH1201&lang=sv#anchor_SH1201>

The Telegraph. (2012, November 1), "Argentina lowers voting age to 16", Accessed October 14th 2015. <www.telegraph.co.uk/news/worldnews/southamerica/argentina/9647629/Argentina-lowers-voting-age-to-16.html>.

The United Nations. 1948. "Universal Declaration of Human Rights", Accessed October 13th 2015. <http://www.un.org/Overview/rights.html>.

UK Government. Accessed July 30th 2015. <https://www.gov.uk/government/topical-events/scottish-independence-referendum/about>.

Wagner, M., D. Johann, and S. Kritzinger. 2012. "Voting at 16: turnout and the quality of vote choice". *Electoral studies*, 31(2), 372–383.

Zeglovits, E. and M. Zandonella. 2013. "Political interest of adolescents before and after lowering the voting age: the case of Austria", *Journal of Youth Studies*, 16(8), 1084–1104.

Appendix A: Additional information on measures

Table A1 Course objective for Social Studies A (*Samhällskunskap A*)

The students should:
1. Have knowledge about the evolution and function of democracy and be able to apply democratic working methods (ha kunskap om demokratins framväxt och funktion samt kunna tillämpa ett demokratiskt arbetssätt)
2. Be able to understand how political, economic, geographical and social conditions have shaped and continue to shape our own society as well as the international community (kunna förstå hur politiska, ekonomiska, geografiska och sociala förhållanden har format och ständigt påverkar såväl vårt eget samhälle som det internationella samhället)
3. Have knowledge about the function of the political system at the local, regional and national levels as well as in the EU (ha kunskaper om det politiska systemets funktion på lokal, regional, nationell och EU-nivå)
4. Be able to understand how one can influence political decisions at the local, regional and national levels as well as in the EU and internationally (kunna förstå hur man kan påverka politiska beslut på lokal, regional och nationell nivå, inom EU samt internationellt)
5. Be able to formulate, understand and reflect upon social issues using historical as well as future perspectives (kunna formulera, förstå och reflektera över samhällsfrågor ur såväl historiska som framtida perspektiv)
6. Be able to apply ethical and environmental perspectives on different social issues (kunna lägga etiska och miljömässiga perspektiv på olika samhällsfrågor)
7. Be able to use different sources of knowledge and methods when working with social issues (kunna använda olika kunskapskällor och metoder vid arbetet med samhällsfrågor)
8. Understand how opinions and attitudes come about and be aware of how values and stances are formed (känna till hur åsikter och attityder uppstår samt vara medveten om hur värderingar och ställningstaganden formas)

Notes: This information comes from The Swedish National Agency for Education (2015). The original Swedish formulations are in parentheses.

Appendix B: Additional results

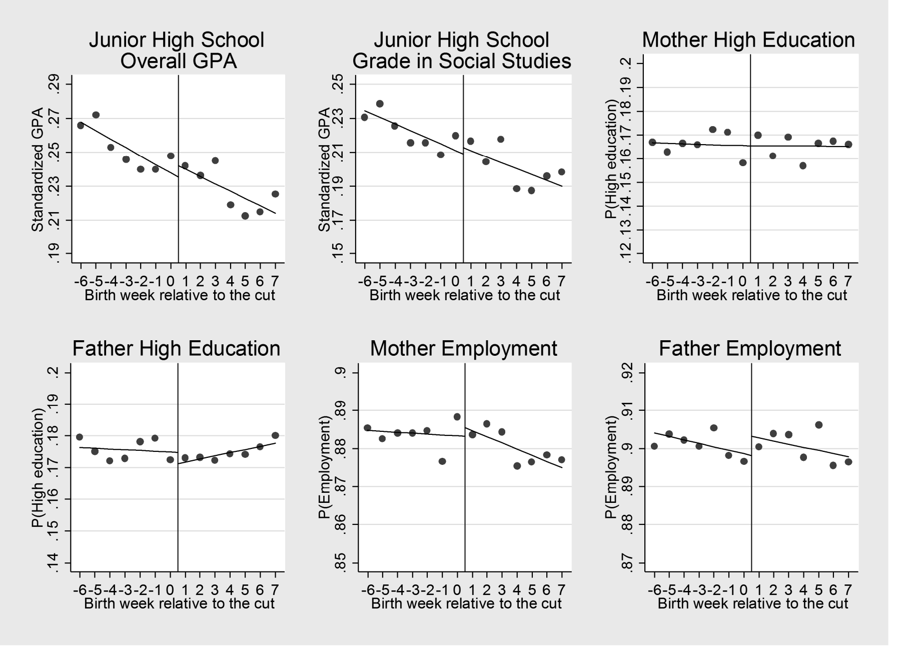


Figure B1. The relationship between the predetermined variables and the running variable

Notes: The figures are based on individuals with valid information on the high school grade in Social Studies and non-missing data on the relevant variable.

Table B1. RD estimates for the predetermined variables

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	JHS Overall GPA	JHS grade in Social Studies	Mother highly educated	Father highly educated	Mother employed	Father employed
Panel A. Linear control, full sample [-6,7]						
Voting right	-0.0064	-0.0037	-0.0001	0.0043	-0.0031	-0.0050
	(0.0095)	(0.0117)	(0.0052)	(0.0055)	(0.0041)	(0.0037)
Mean of dep.	0.241	0.212	0.166	0.175	0.882	0.901
Observations	102,191	102,191	115,321	111,904	115,405	112,437
p-value for the test of joint significance of the estimates: 0.32						
Panel B. Linear control, reduced sample [-5,6]						
Voting right	-0.0112	-0.0080	-0.0005	0.0043	-0.0033	-0.0051
	(0.0101)	(0.0126)	(0.0058)	(0.0060)	(0.0044)	(0.0039)
Mean of dep.	0.240	0.212	0.166	0.174	0.882	0.901
Observations	89,227	89,227	100,736	97,757	100,811	98,229
p-value for the test of joint significance of the estimates: 0.23						
Panel C. Linear control, reduced sample [-4,5]						
Voting right	-0.0086	-0.0097	-0.0037	0.0041	-0.0046	-0.0026
	(0.0112)	(0.0135)	(0.0064)	(0.0068)	(0.0048)	(0.0041)
Mean of dep.	0.239	0.210	0.166	0.174	0.883	0.901
Observations	74,958	74,958	84,636	82,146	84,701	82,554
p-value for the test of joint significance of the estimates: 0.54						
Panel D. Linear control, reduced sample [-3,4]						
Voting right	-0.0038	-0.0072	-0.0087	0.0032	-0.0047	-0.0059
	(0.0109)	(0.0143)	(0.0070)	(0.0076)	(0.0056)	(0.0044)
Mean of dep.	0.240	0.211	0.166	0.174	0.883	0.901
Observations	60,453	60,453	68,307	66,259	68,364	66,589
p-value for the test of joint significance of the estimates: 0.36						
Panel E. Quadratic control, full sample [-6,7]						
Voting right	-0.0102	-0.0138	-0.0096	0.0012	-0.0040	-0.0024
	(0.0144)	(0.0190)	(0.0091)	(0.0094)	(0.0068)	(0.0051)
Mean of dep.	0.241	0.212	0.166	0.175	0.882	0.901
Observations	102,191	102,191	115,321	111,904	115,405	112,437
p-value for the test of joint significance of the estimates: 0.79						

Notes: The results are based on individuals with valid information on the high school grade in Social Studies and non-missing data on the relevant variable. JHS=Junior High School. Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level.

Table B2. Is there a selection into having a grade or a test result?

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Panel A: Indicator for having a high school grade in Social Studies							
Voting right	0.0065*	0.0082*	0.0096**	0.0056	0.0103	0.0057	0.0073
	(0.0039)	(0.0042)	(0.0045)	(0.0054)	(0.0064)	(0.0035)	(0.0045)
Mean of dep.	0.677	0.677	0.678	0.678	0.677	0.677	0.677
Observations	172,283	150,566	126,383	101,923	172,283	172,283	172,283
Panel B: Indicator for having a SweSAT result							
Voting right	-0.0016	-0.0008	-0.0050	-0.0015	0.0017	-0.0019	-0.0014
	(0.0063)	(0.0066)	(0.0077)	(0.0081)	(0.0092)	(0.0045)	(0.0051)
Mean of dep.	0.195	0.195	0.195	0.193	0.195	0.195	0.195
Observations	74,527	65,103	54,640	43,908	74,527	74,527	74,527
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variabl.	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

Table B3. Heterogeneity: high school grade in Social Studies

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Panel A: Males							
Voting right	-0.0090	0.0009	0.0017	0.0174	0.0228	-0.0003	-0.0090
	(0.0158)	(0.0173)	(0.0181)	(0.0199)	(0.0242)	(0.0128)	(0.0169)
Mean of dep.	-0.143	-0.142	-0.145	-0.142	-0.143	-0.143	-0.143
Observations	54,735	47,840	40,308	32,524	54,735	54,735	54,735
Panel B: Females							
Voting right	-0.0168	-0.0173	-0.0146	-0.0112	-0.0212	-0.0173	-0.0183
	(0.0141)	(0.0147)	(0.0162)	(0.0184)	(0.0206)	(0.0122)	(0.0153)
Mean of dep.	0.0987	0.0972	0.0972	0.0981	0.0987	0.0987	0.0987
Observations	61,978	54,115	45,360	36,616	61,978	61,978	61,978
Panel C: Individuals with at least one highly educated parent							
Voting right	-0.0168	-0.0189	-0.0060	-0.0188	-0.0135	-0.0251	-0.0148
	(0.0188)	(0.0202)	(0.0223)	(0.0269)	(0.0316)	(0.0157)	(0.0226)
Mean of dep.	0.366	0.368	0.363	0.366	0.366	0.366	0.366
Observations	29,365	25,630	21,528	17,366	29,365	29,365	29,365
Panel D: Individuals with no highly educated parent							
Voting right	-0.0114	-0.0058	-0.0044	0.0145	0.0098	-0.0055	-0.0131
	(0.0113)	(0.0118)	(0.0121)	(0.0139)	(0.0160)	(0.0094)	(0.0121)
Mean of dep.	-0.143	-0.144	-0.144	-0.143	-0.143	-0.143	-0.143
Observations	87,348	76,325	64,140	51,774	87,348	87,348	87,348
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variabl.	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

Table B4. Heterogeneity: SweSAT score (General Knowledge)

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bandwidth:	[-6,7]	[-5,6]	[-4,5]	[-3,4]	[-6,7]	[-6,7]	[-6,7]
Panel A: Males							
Voting right	-0.0711	-0.0676	-0.0404	-0.0277	-0.0041	-0.0761*	-0.0705
	(0.0477)	(0.0512)	(0.0519)	(0.0598)	(0.0641)	(0.0429)	(0.0526)
Mean of dep.	-0.0159	-0.0163	-0.0110	-0.0234	-0.0159	-0.0159	-0.0159
Observations	6,435	5,637	4,745	3,774	6,435	6,435	6,435
Panel B: Females							
Voting right	-0.0052	0.0154	0.0588*	0.0604*	0.0919**	-0.0274	0.0018
	(0.0294)	(0.0320)	(0.0306)	(0.0340)	(0.0381)	(0.0277)	(0.0425)
Mean of dep.	-0.432	-0.428	-0.423	-0.426	-0.432	-0.432	-0.432
Observations	8,103	7,048	5,912	4,706	8,103	8,103	8,103
Panel C: Individuals with at least one highly educated parent							
Voting right	-0.0571	-0.0439	-0.0135	-0.0856*	-0.0197	-0.0805**	-0.0431
	(0.0372)	(0.0442)	(0.0436)	(0.0451)	(0.0567)	(0.0339)	(0.0490)
Mean of dep.	-0.0409	-0.0338	-0.0212	-0.0240	-0.0409	-0.0409	-0.0409
Observations	5,432	4,727	3,934	3,140	5,432	5,432	5,432
Panel D: Individuals with no highly educated parent							
Voting right	-0.0162	-0.0119	0.0209	0.0672	0.0634	-0.0353	-0.0167
	(0.0329)	(0.0383)	(0.0370)	(0.0458)	(0.0524)	(0.0246)	(0.0416)
Mean of dep.	-0.371	-0.371	-0.367	-0.378	-0.371	-0.371	-0.371
Observations	9,106	7,958	6,723	5,340	9,106	9,106	9,106
Linear control	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Election FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quad. control	No	No	No	No	Yes	No	No
Predet. variabl.	No	No	No	No	No	Yes	No
Day of birth RV	No	No	No	No	No	No	Yes
Weekday FE	No	No	No	No	No	No	Yes

Notes: Standard errors are clustered on week * election level. */**/*** significant at the 10/5/1 percent level. RV=running variable.

Essay 4: The strength of the weakest link: Sickness absence, internal substitution and worker-firm matching

(With Lena Hensvik)

1 Introduction

Many countries struggle with high sickness absence rates and large associated costs for firms.[1] Yet little is known about how sickness absence affects key labor market outcomes, such as access to jobs, worker mobility and career trajectories. In addition, we know next to nothing about which strategies firms use to minimize the costs of employee absence.

This paper examines the idea that firms' production losses caused by temporary work absence depend on the internal substitutability of workers. Thus, firms should have incentives to keep absence low in jobs with few substitutes. Using Swedish administrative matched employer-employee data linked to information about individual sickness absence for almost 6 million worker-year observations, we document a robust positive relationship between employee absence and the number of internal substitutes defined by detailed occupations. The difference in the absence probability between more and less substitutable employees is substantial, even conditioning on establishment and occupation fixed effects: it is roughly equal to the average difference in work absence between young labor market entrants and middle-aged workers or between workers with and without children.[2] This pattern holds whether we look at employees' own sickness absence or absences among parents caused by child sickness. Parents in jobs with few internal substitutes seem to shift part of their care leave for children to their spouse instead.

We then use several additional analyses to probe the mechanisms behind our results. About half of the effect remains after the inclusion of worker fixed effects, which indicates that sorting and on-the-job adjustments in absence behavior are of equal importance for the observed association between sickness absence and internal substitutability. Further investigations of the selection mechanism show that workers hired for jobs with few internal substitutes have significantly lower pre-hire sickness absence than other new hires. They also display higher turnover rates caused by realized absence. Together, these two results highlight that sorting occurs both via the entry and exit margin.

In addition, we find stronger selection effects and weaker separation responses among hires with a strong pre-hire employment record, previous

---

[1] It is reported that 131 million working days were lost due to sickness absence in the UK in 2013 (Office for National Statistics (UK), 2014). Another report from the UK estimates that employers pay GBP 9 billion (USD 12 billion) a year in sick pay and associated costs (Black and Frost, 2011). In Germany it is reported that employers spend about EUR 25 billion (USD 28 billion) per year on sick pay. This number is more than 1 percent of the total GDP in Germany (German Federal Statistical Office, 2011). Numbers for Sweden suggest that employers spent SEK 21 billion (USD 2.6 billion) on sick pay and associated costs in 2012 (Previa, 2013).

[2] The reported difference reflects the differential absence rate between employees with no internal substitutes and employees with more than five substitutes.

employment at another site in the same firm or a coworker connection to an incumbent employee. We interpret this as suggestive evidence that sorting is more pronounced when there is more information about the workers' absence-types ex ante, and that learning about match quality is an important determinant of turnover rates.

This paper contributes to several strands of the current literature. The idea that firms try to find the right employees for the right jobs is motivated by the notion that worker and firm heterogeneity can lead to match-specific gains in productivity.[3] But despite the theoretical foundations for match-specificity, there is still little empirical evidence on cross-firm differences in hiring and the importance of worker-firm complementarities. One reason is that it is inherently difficult to measure, ex ante, how well a worker matches a particular job. Thus, researchers have mainly been restricted to infer the effects of match quality based on how wages and separations vary with tenure and job mobility (Nagypál [2007]; Lazear and Oyer [2007]).[4]

In addition, the discussion about match quality is often focused on complementarities in terms of worker skills (or human capital) and the skill requirements (or technology) of different jobs.[5] But it is equally likely to be important complementarities between other dimensions of employee attributes and firm technology that can affect the sorting of workers over jobs and, in turn, their subsequent labor market outcomes.[6] Our results highlight a previously overlooked, but seemingly important, dimension of match quality related to complementarities between workers' absence rates and firms' possibilities for internal replacement.

A few recent papers specifically point to the importance of internal labor substitution for worker and firm outcomes. Jäger (2015) shows that internal labor markets are important for firms' replacements of sudden employee exits (caused by deaths), suggesting that firms face significant search frictions in the external labor market. Our results complement his findings by highlighting the importance of internal substitution for insuring firms against temporary production disruptions caused by employee absence. In this sense,

[3] See Sattinger (1975) and Tinbergen (1956) for the original work on the problem of assigning heterogeneous workers to heterogeneous jobs.
[4] Two exceptions are Jackson (2013), who shows that teacher-school match effects explain a quarter of the variation in teacher quality, and Fredriksson et al. (2015), who show that wages and job separations depend on how well workers' cognitive abilities and personality traits match the abilities of the existing workforce.
[5] See, for example, Abowd et al. (2007) on how different components of skills are related to firms' technological inputs; Andersson (2009) on the relationship between firms' product market segment and the demand for worker innovation skills in the software industry or Lazear (2009) on firm-level heterogeneity in skill-weights.
[6] For example, Lazear (1998) argues that the match-quality between a worker and a given firm depends on the riskiness of workers and firm-level characteristics such as expected time-horizon and the degree of private information.

high-absence workers are weak links in jobs with few internal replacements.[7] Furthermore, Goldin and Katz (forthcoming) argue that the possibilities for employee substitution is a key factor behind the wage penalties associated with shorter hours, and in turn the gender pay gap.[8] The observed link between low internal substitutability and low probability of being absent to care for children is clearly consistent with their argument that the ease with which employees can substitute for each other affects individual absence costs.

In addition, employee selection and hiring strategies are still something of a black box (Oyer and Schaefer [2011]). Limited evidence suggests that employers are reluctant to hire applicants with a history of sickness absence, but remain uninformative of why (Eriksson, Johansson and Langenskiöld [2012]). Our results suggest that firms' ability to internally substitute for absent workers is a key aspect in this process. But we also shed light on the role of information in the hiring decision. The fact that job separations respond to realizations of absence indicates that employment relationships are formed under uncertainty, as in the seminal model of Jovanovic (1979). Consistent with several studies showing that firms rely on signals or informal search channels in order to screen for the right workers, our findings suggest that pre-hire screening serves as a tool for firms to achieve an allocation of low-absence workers in unique positions.[9]

From the worker's perspective, these results imply that episodes of sickness absence affect the chances of accessing and retaining unique positions, which account for a non-trivial share of the labor market. Hence, workers have strong incentives to keep absence low in jobs with low internal substitutability, which they do by e.g. shifting the care for children to their partners.

The remainder of the paper is structured as follows. In Section 2, we describe the data and clarify crucial definitions. The empirical specification and the results are presented in Section 3. Section 4 concludes.

[7] Our findings also relate to the studies documenting a positive association between sickness absence rates and firm size in the cross-section, which is consistent with the argument that production in small firms should be particularly sensitive to individual sickness absence (Barmby and Stephen [2000]), Dionne and Dostie [2007], Ose [2005] and Lindgren [2012]). However, it is possible that this relationship also reflects other between-firm differences related to size. By exploiting variation in the number of substitutes within narrowly defined job cells, the present paper provides a more credible assessment of the direct relationship between sickness absence and employee substitutability.

[8] Their paper specifically looks at the pharmacist occupation and argues that enhanced substitutability (due to technological change and increased standardization) has decreased the wage penalty from shorter hours for women with children, and in turn the gender pay gap relative to other professions in the US labor market.

[9] Empirical studies in this literature suggest that employers use observable signals such as education (Farber and Gibbons [1996]; Altonji [2005]; Lange [2007]; Schönberg [2007]), unemployment status (Eriksson and Rooth [2014]), and referral ability (Hensvik and Nordstrom Skans, forthcoming) to form expectations about prospective workers' productivity.

2 Data

2.1 Definitions and measurements

We use Swedish register data from 1997 to 2007. These data are drawn from registers administered by Statistics Sweden that follow all Swedish workers from 1985-2010, with unique person, firm and establishment identifiers. In the main analysis we restrict the sample to jobs in the private sector. The reason is that the definition of the establishment is more precise in the private sector.[10] To these data, we add demographics from a population-wide dataset and information on occupation codes, which is available from 1997-2010 for a large sample of private establishments covering almost 50 percent of private sector workers.[11]

2.1.1 Measuring internal substitutability

We define employee substitutability as the number of other workers within the same combination of establishment and occupation (ISCO-88, 3-digit level) in a given year. For example, an administrator at an establishment that employs four administrators in total will have three substitutes. In order to focus on regular workers, we drop employees in managerial positions. We also drop employees at very small establishments (less than three employees).

Our definition of employee substitutability is supported by Jäger (2015) who shows that when an employee exits (due to death), firms increase their demand for the remaining workers in the same occupation, but not in other occupations, as the deceased. This clearly indicates that firms regard employees within the same narrowly defined occupations as closer substitutes than employees in other occupations. In most specifications, we let an indicator for having 0–5 substitutes define low internal substitutability, but we also show results from more flexible models.

It is likely that the number of substitutes will be measured with error. Specialization within occupations could lead to an overstatement of the true level of substitutability. But it is also possible that some coworkers have overlapping skills even if they occupy different jobs, in which case we may understate the true number of substitutes. We address this issue in the robustness section using alternative definitions of employee substitutability that e.g. takes firm size into account.

[10] In the public sector, all individuals employed by the same municipality are sometimes registered as belonging to the same establishment.

[11] We start the observation period in 1997 since this is the first year that we can observe occupations in our data. The reason for ending already in 2007 is that in some cases, we want to follow workers for a 3-year follow-up period. In terms of sampling, a new random sample is drawn each year and the establishments are stratified by firm size and industry. Table A3 shows the distribution of establishments (col. 2) and employees (col. 3) with respect to establishment size.

2.1.2 Measuring sickness absence

We add sickness absence spells from the Swedish Social Insurance Agency. These data include all spells longer than two weeks.[12] Sickness absence will generally be defined as an indicator for having at least one such spell in a given year. But in some specifications we will also consider absence on the intensive margin by using the log of sickness benefits as the outcome of interest.

The fact that we cannot observe shorter spells is obviously a limitation of the data, and we will therefore complement our analysis with short-term work absence due to care leave for sick children as an alternative outcome measure. In Sweden, parents with small children (0–10 years old) can be absent from work to care for sick children (who are too sick or infectious to be in school or in daycare).[13] The parent that stays home receives Temporary Parental Benefits from the Social Insurance system from day one, meaning that these benefits data also pick up short term absence spells.[14]

2.1.3. Defining hires, pre-hire and realized absence

We examine the role of worker sorting in more detail using a dataset consisting of new hires. We define new hires as employees observed in an establishment in a given year, but not in the same establishment *or* in the same firm in any of the five preceding years. For each hire, we measure their pre-hire sickness absence as the average incidence of having at least one sickness absence spell (longer than two weeks) per year in the three years prior to employment. In order for all new hires to have at least three pre-hire years, we restrict the sample to workers with at least 4 years of labor market experience.[15] We will also examine the probability of job separation when the worker absence type is revealed. To this end, we define realized absence of new hires as the average sickness absence probability in the hiring year and the year after entry.

2.1.4 Measuring uncertainty

Part of our empirical analysis aims to contrast realized matches between workers and firms where the hiring decision was based on more or less in-

[12] The data include all spells for which the individual was entitled to sickness benefits from the social insurance system. Since spells shorter than two weeks are paid by employers, these are not available in our data.

[13] 90 percent of all parents in Sweden have their children between 3-6 years of age in subsidized child care (Mörk et al. [2013]).

[14] Parents may claim benefit compensation for up to 120 days per year. The replacement rate is 80 percent of lost earnings up to a monthly wage ceiling of SEK 37,000. The benefit compensation data contain information on the total amount of child sick benefits received each year, from which we construct an indicator for having at least one child sick spell in a given year.

[15] In other words, labor market entrants who graduated from their highest education less than four years ago are excluded.

formation about worker absence type. To this end, we use three different proxies for the amount of information about the employees in the matching stage:

I *Pre-hire employment*: an indicator for whether the new employee was employed in t-3, t-2 and t-1.

II *Firm connection*: an indicator for whether the new employee was employed in the same firm but in another establishment sometime between t-5 and t-1.

III *Coworker connection*: an indicator for whether the new employee was ever employed in the same establishment (at another firm) as at least one of the incumbent employees of the hiring establishment.[16]

These information proxies are all based on the notion that the employment history of a worker provides information about his or her future absence behavior. Hence, when new hires fulfill one of the three criteria above, we assume that the hiring decision was based on a more precise signal about the prospective hire's absence type.

Although it is clear that these measures are far from perfect, several studies support our choice of information proxies. Work by Eriksson and Rooth (2014) shows that employers are reluctant to hire people from non-employment, which indicates that non-employment is associated with some degree of uncertainty about worker type. Thus, we find it reasonable to expect that there is more information available about workers with a strong attachment to the labor market.[17] Schönberg (2007) further shows that hard-to-observe characteristics of college graduates are more easily assessed by the current firms than by outside firms. Under this assumption, we expect that matches involving workers with an earlier connection to the recruiting firm are based on better information about the worker absence type.

Finally, there is recent evidence that incumbent employees can provide valuable information about the productivity of prospective hires with whom they have worked in the past (Dustmann et al. [2015] and Hensvik and Nordström Skans, forthcoming). Based on these findings, we assume that firms

[16] We construct dyads for each hire-incumbent combination (i.e. if a new worker comes to an establishment with 10 incumbent workers we create 10 dyads). For each dyad we add information on the full history of employers for both agents back to 1985. A coworker connection is defined as having overlapping employment spells in the same establishment.
[17] Farber and Gibbons (1996), Altonji (2005) and Lange (2007) show that employers overprice formal credentials (and underprice hidden talents) among inexperienced workers, which further supports that there is less information about worker type for employees with weaker labor market experience.

can make better predictions about the absence type of former coworkers to their current employees.

2.2 Descriptive statistics

Tables A1 and A2 in the Appendix contain descriptive statistics on the sample of all workers and on the sample of new hires, respectively. There are 6 million observations in the full sample (Table A1) and 400,000 new hires (Table A2). About 20 percent of these occupy jobs with 0-5 substitutes, which suggests that positions with low employee substitutability account for a significant share of the labor market.[18] About 4 percent of the workers have truly unique jobs (i.e. 0 substitutes) and about 11 percent of the employees have at least one sickness absence spell that is longer than two weeks in a given year.[19] Consistent with our hypothesis, the incidence of sickness absence is lower for workers with relatively few substitutes (0–5), but these workers differ in other aspects as well; they are, for example, employed in smaller establishments and in more skilled professions with higher wages, suggesting that they have key positions within the firms.[20] Workers in relatively unique positions are also older and more often women, although education levels appear similar to other employees.

The image of the new hires is very much in line with the full sample. Importantly, positions without substitutes are present in all occupational skill levels (the note for Table A2 gives the distribution). In contrast to our hypothesis, however, the pre-hire sickness absence rate is higher for workers who entered relatively unique positions, while wages are about the same. But as noted before, it is important to account for other aspects that differ systematically between individuals in more/less unique positions before we can draw conclusions about the relationship between employee absence and internal substitutability.

3 Empirical strategy and findings

3.1 Empirical specification

We start by exploring the association between present sickness absence and the number of internal substitutes among all private sector workers by estimating Eq. (1) by OLS:

[18] 1,050,017 (73,366) out of the 5,863,497 (387,901) workers (hires) have jobs with 0–5 substitutes.

[19] The figure on sickness absence is confirmed by estimates from Statistics Sweden (Statistics Sweden, 2007).

[20] The summary statistics show the distribution of workers/hires across a broader set of occupations (1-digit level). When defining the number of substitutes we use more detailed occupation codes (3-digit level).

$$A_{ieot} = \gamma S_{ieot} + \alpha_e + \alpha_o + \theta_t + \beta X_{it} + \delta Z_{et} + \varepsilon_{ieot} \quad (1)$$

where the outcome A_{ieot} is the incidence of sickness absence for worker i in establishment e and occupation o in year *t*. S_{ieot} measures employee substitutability within each job, defined by the interaction between the establishment and the 3-digit occupation code.[21] α_e and α_o are establishment and occupation fixed effects, respectively. We also include year fixed effects, θ_t, to account for e.g. business cycle swings potentially correlated with firms' organization of work and individual sickness absence. The worker characteristics X_{it} consist of gender, age, education, country of origin and an indicator for having children under the age of three.[22] Finally we include establishment size Z_{et}. ε_{ieot} is the error term.

The parameter of interest is γ, which aims to capture the relationship between the number of internal substitutes and work absence.[23] It should be noted that the model is fairly rich as it accounts for unobserved characteristics of both occupations and establishments that could generate a spurious correlation between employee substitutability and absence.

We also want to disentangle to which extent γ captures behavioral responses and/or employee selection on the entry and exit margin. As a first step, we therefore add worker fixed effects to Eq. (1), which means that we account for the selection of workers over jobs with few/many substitutes. Second, we estimate Eq. (1) separately for new hires and replace the outcome with an indicator for the *pre-hire* sickness absence, defined as the average incidence of having at least one sick leave spell longer than two weeks per year in the three years prior to entry. Since pre-hire sickness absence is potentially correlated with past employment, we also control for the employment probability in the same time period. Finally, we examine the separation response to realized sickness absence among new hires by estimating the following equation:

$$Separation_{ieot+2} = \mu \bar{A}_{ieot} + \alpha_e + \alpha_o + \theta_t + \beta X_{it} + \delta Z_{et} + \varepsilon_{ieot} \quad (2)$$

where $Separation_{ieot+2}$ is an indicator for if worker i hired in year t left establishment e between year *t*+1 and *t*+2, and $\bar{A}_{ieot}$ is the realized absence of entrant i measured as the averaged incidence of absence over the entry year (year *t*) and the first year into the employment spell (*t*+1) (we focus on

[21] In our main specifications, we consider jobs with 0-5 substitutes as jobs with low substitutability. Sometimes we also use models with slightly different specifications, which we then state clearly.
[22] We group individuals by their country of origin into the following six categories: Sweden, rest of the Nordic countries, rest of Europe, North America, South America, and the rest of the world.
[23] The baseline analysis focuses on sickness absence on the extensive margin. As a robustness check, we also consider outcomes that capture the intensive margin.

entrants that stayed for at least one year in order to be able to observe their realized sickness absence). The controls are the same as in Eq. (1) and ε_{ieot} is the error term.

The aim of μ is to capture the separation response to the realized absence behavior among newly hired workers. To examine whether this response depends on the internal substitutability of employees, we also estimate versions where the model in Eq. (2) is fully interacted with the indicator variable for low employee substitutability (i.e. with S_{ieot} in Eq. [1]).

3.2 Baseline results: employee substitutes and absence

Figure 1 shows the estimates from Eq. (1) when we include dummy variables for having up to 5 substitutes (employees with more than 5 substitutes constitute the reference category). The estimates are all statistically negative on the 1-percent level, ranging between 1 and 2 percentage points. Hence, workers with few close substitutes have lower absence rates. Interestingly, the estimates become smaller in absolute value as the number of employees performing the same job increases, which is consistent with the idea that the costs of employee absence, in terms of production disruptions, increase as the possibilities of internal substitution decrease.

In the Appendix (Figure A1) we show the same relationship for up to 10 substitutes (employees with more than 10 substitutes constitute the reference category). These results suggest that there is a significant jump in the absence probability when the number of substitutes increases from 0 to 1. Beyond that, there is a fairly linear relationship between employee absence and the number of substitutes.[24] The magnitudes of the estimates are substantial, especially for the coefficients on 0 and 1 substitutes: the difference in sickness absence between jobs with more than 5 substitutes and jobs with 0 substitutes, conditional on the model, is roughly equivalent to the estimated difference in absence rates between workers in their 20s and 40s, or between workers with and without small children (0–3 years of age).

[24] The difference in absence probability between jobs with ten and more than ten substitutes is around 0.5 percentage points. This remaining difference may seem surprising but is probably due to the fact that we have measurement error in the possibilities of replacing an absent employee which is likely to decrease with the size of the job-cell (in large cells there is a greater chance that at least some workers are perfect substitutes for each other).

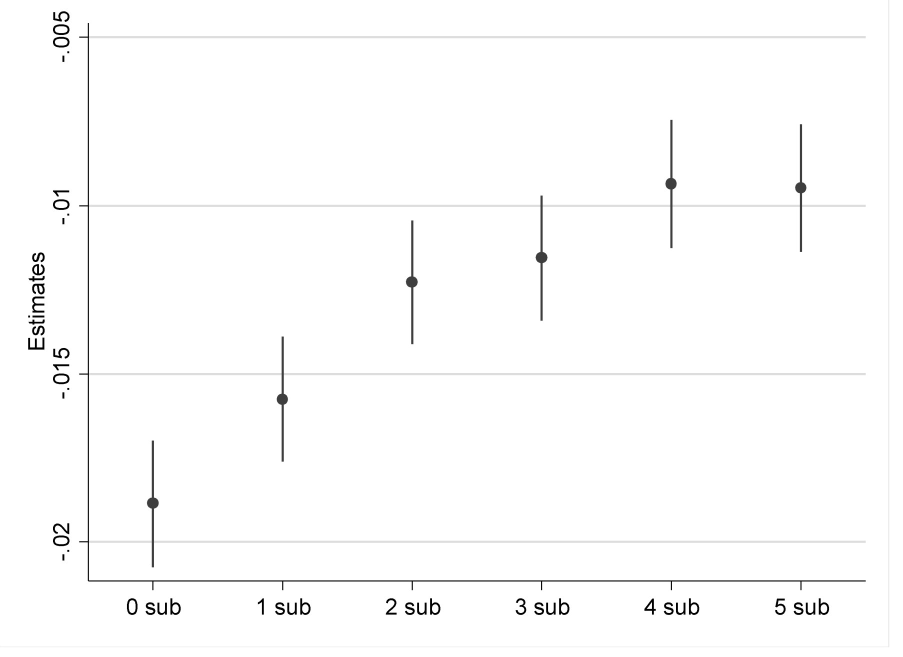


Figure 1. Sickness absence and the number of internal substitutes.

Notes: The figure shows the estimated coefficients on dummies for 0–5 substitutes in Eq. (1). The reference category is employees with more than 5 substitutes and the background controls are gender, age, education, birth country, having children aged 0–3 and establishment size. The model also includes year, occupational and establishment fixed effects. Standard errors are clustered on the establishment level.

Table 1 shows the point estimates (with and without worker characteristics) when we, for simplicity, only use one indicator for low substitutability, defined as having 0–5 substitutes. As before the reference category is employees with more than 5 substitutes. Overall, these results show a strong negative correlation between low internal substitutability and work absence.

Table 1. Sickness absence and internal substitutability

Outcome: Sickness absence in *t*	(1)	(2)
Low substitutability	-0.0104***	-0.0131***
	(0.0006)	(0.0006)
Number of observations	5,863,497	5,863,497
Mean of dependent variable	0.109	0.109
Background controls	No	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes: The standard errors are clustered on the establishment level. Low substitutability is defined as having 0–5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having children aged 0-3 and establishment size. *, **, and *** denote statistical significance at the 10-, 5-, and 1 percent level.

3.3 Evidence from child sick spells

A limitation of our data is that we only observe absence spells longer than two weeks. To test if our results also extend to shorter absence spells, we therefore include an alternative absence measure: *Care leave for sick children*, which also includes short-term work absence.[25] 65 percent of the parents have at least one absence spell according to our definition, which suggests that this type of absence is a first-order concern for firms that employ workers with small children in the household.[26]

We restrict this analysis to individuals with at least one child between 0 and 10 years old (these are the children with whom parents are entitled to be at home) and use an indicator for having positive Temporary Parental Benefits in a given year as the outcome. To see if employee substitutability affects the division of care for sick children within the family, we also use the corresponding absence measure for the partner as an outcome (the sample is then further restricted to individuals with a cohabiting partner).

Table 2 presents the results from this exercise using the model described by Eq. (1). The estimate in column (1) clearly suggests that workers in jobs with few substitutes are significantly less likely to be absent due to care leave for sick children. Thus, the results are in line with our general findings in Table 1, although compared to the baseline they are smaller in magnitude.[27] Interestingly, the partners of employees with few substitutes are more

[25] The reason is that parents receive benefits from the Social Insurance System from day one.
[26] Following women in Sweden who had their first child in 1994, Boye (2015) shows that the average woman is absent from work for 5 days per year and the average man is absent 2.5 days per year during the first 10 years of the child's life, with higher absence rates for children in daycare ages.
[27] The weaker relationship may reflect that a large share of parents have at least one child sickness spell, which is likely to make the extensive margin less relevant. In Table A7, which

likely to be home caring for sick children (see column [2]) and the magnitude of the estimate is almost equal to the estimate in column (1).[28] Thus, children of workers with few internal substitutes are no less sick than other children; instead, these workers seem to avoid work interruptions by shifting work absence to their partners.[29]

Table 2. Evidence from child sick spells

Column:	(1)	(2)
Outcome:	Absence to care for sick child	
	Own absence	Partner's absence
Low substitutability	-0.0115***	0.0087***
	(0.0017)	(0.0018)
Number of observations	1,911,734	1,767,118
Mean of dependent variable	0.654	0.553
Background controls	No	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes: The standard errors are clustered on the establishment level. Low substitutability is defined as having 0–5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having children aged 0–3 and establishment size. In column (1) we restrict the sample to individuals with at least one child younger than 11 years of age and in column (2) we further restrict the sample to individuals with cohabiting partners. We further control for the number of children in the following categories: 0–3 years, 4–6 years and 7–10 years. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

3.4 Robustness checks

3.4.1 Alternative measures of employee substitutability

Our baseline measure of internal substitutability is the number of employee substitutes in the same occupation. But it is quite possible that the substitutability of workers could interact with the size and the organization of the establishment. We may, for example, overstate the degree of substitutability in large establishments if employees are organized in different departments that make substitution difficult. At the same time, more coworkers in general could imply that employees are more substitutable, as there is a higher likelihood that some workers have overlapping skill sets even though they occupy different jobs.

we discuss below we show that the effect on the intensive margin is very similar as for own sickness absence.

28 The estimate in column (1) is almost identical if we use the same sample as in column (2) (i.e. employees with cohabitating partners).

29 As a robustness exercise, we have also looked at the relationship between own substitutability and the partner's *own* sickness absence. The estimate is close to zero and precisely estimated (-0.0020 [0.0010]). The fact that the partner's response is concentrated to child leave days is reassuring, as these (but not own sick leave days) can be shifted between parents.

It is therefore not clear how to (and if we should) adjust the number of substitutes to establishment size. As a starting point, column (4) of Table A3 shows how much of the identifying variation in the variable *Low substitutability* that comes from establishments of different sizes. The figures are based on the squared residuals from a regression of S_{ieot} on the full covariate set in Eq. (1). It is clear that small to medium establishments account for a large share of the variation: 41 percent comes from establishments with 3–49 employees and 40 percent comes from establishments with 50–249 employees. To test how relevant our results are for establishments of different sizes, we therefore re-estimate the baseline model separately for those with 3–49, 50–249, 250–500 and more than 500 employees.[30] The estimates from this exercise are plotted in Figure A2. All four estimates are significantly negative on the 1 percent level and the magnitudes of the estimates are roughly similar to the estimate presented in column (2) of Table 1. Thus, our measure of low substitutability is relevant for both small and large establishments.

However, for completeness we have also tested two other definitions of low substitutability based on the logic that employees are less substitutable in larger establishments (for a given number of substitutes). First, we define low substitutability as a situation when one of the following criteria is fulfilled: (i) no substitutes in establishments with 3–49 employees; (ii) <4 substitutes in establishments with 50–249 employees; (iii) <7 substitutes in establishments with 250–500 employees or (iv) <10 substitutes in establishments with more than 500 employees. The second alternative definition is based on the number of substitutes divided by establishment size and defines low substitutability as a situation when this share is below 0.03 (which corresponds to the tenth percentile).

These definitions are of course arbitrary, but offer a way of relating the notion of substitutability to the overall size of the establishment (columns [5–6] of Table A3 show that more of the identifying variation now comes from larger establishments). However, when we re-estimate Eq. (1) using these two alternatives, we obtain virtually identical estimates as before (see Table A4).[31] Overall, we conclude that the link between employee substitutability and absence is relevant for establishments of all sizes (rather than only relatively small ones) and that our results are robust to different variations in the definition of low internal substitutability.

[30] The division is based on a classification that Statistics Sweden uses when they collect data from firms.

[31] Column (4) also shows the linear relationship between the share of substitutes and absence, which suggests that a standard deviation increase in the share of substitutes is associated with a 0.8 percentage point higher probability of absence (0.3*2.7).

3.4.2 Alternative explanations and specification checks

The strong association between sickness absence and substitutability naturally raises the relevant question of whether wages differ between more and less substitutable jobs. Table A6 in the Appendix suggests that is indeed the case. We obtain these estimates by replacing sickness absence as the outcome in Eq. (1) with the log of the monthly full-time wage.[32] The results in column (1) suggest that employees with 0–5 substitutes have 1.3 percent higher wages on average relative to employees with more than 5 substitutes, conditional on establishment and occupation fixed effects.[33]

This wage premium could reflect both that unique jobs are more productive and/or that the employees in unique jobs have more productive (unobserved) skills that are correlated with their absence type. In column (2) of Table A6, we show the baseline estimate when we, as a robustness check, hold the wage constant. Even if it is potentially problematic to control for the wage (as the wage is likely to be endogenous to the level of absence), it is reassuring to see that this only has a minor impact on the main estimates.

Column (3) of Table A6 shows the estimate when we add the public sector employees to our sample. This estimate is somewhat smaller, but still significant and of important magnitude, suggesting that the relationship between the number of internal substitutes and sickness absence holds in the full economy. In column (4) of Table A6 we use data on private sector employees for the years 2005–2007. In these years the occupational code is available on a 4-digit level and thus we can test if our main results in Table 1, which are based on a 3-digit occupational code, are robust to finer definitions of occupations. The estimate is very similar to the one in column (2) of Table 1 and confirms the general picture of low sickness absence in jobs with few substitutes.

Finally, in Table A7 we use the log of annual sickness benefits instead of an indicator for sickness absence as the outcome, which picks up the length and number of absence spells. Conditional on being absent for at least two weeks, employees with fewer substitutes have roughly two percent fewer absence days (Panel A) and almost four percent fewer care leave days due to child sickness (Panel B) compared to employees with more substitutes (the received benefits are closely related to the number of leave days). The results thus seem qualitatively robust to variations in the way we measure sickness

[32] The wage is the wage the employee had during the sampling week expressed in full-time monthly equivalents. The variable includes all fixed wage components, including piece-rate and performance pay as well as fringe benefits. Overtime pay or paid leave is, however, not included. The monthly wage is adjusted to full-time for part-time workers by Statistics Sweden. For blue-collar workers the wage is typically obtained by the hourly pay rate times the number of hours that correspond to full-time employment. For white-collar workers it reflects the September wage adjusted by the share of part-time work during the same month.

[33] Estimating the same model for new hires we find an identical wage premium.

absence and suggest that workers with few substitutes have lower absence rates on both the extensive and the intensive margin.

3.5 Behavior vs. entry and separation responses

3.5.1 Behavior and entry

The documented relationship between sickness absence and internal substitutability may both reflect a selection effect (systematic sorting into and out of jobs with few substitutes) and a behavioral effect (workers adjusting their absence behavior when they have few substitutes). To examine the relevance of these two explanations, we exploit variation in the number of substitutes for the same worker over time by adding worker fixed effects to the baseline specification. The estimate presented in column (2) of Table 3 is roughly halved compared to the baseline estimate in column (1) but remains significantly negative on the 1 percent level. Hence, workers do adjust their work absence depending on the number of internal substitutes.[34] Taken together, these results suggest that the correlation between internal substitutability and sickness absence entails both a selection component and a behavioral component that appear to be of similar importance.

In the third column of Table 3, we replace present sickness absence with the *pre-hire* sickness absence described in Section 2.1 in a sample of new hires. Consistent with our earlier results, this estimate clearly suggests that workers hired into positions with fewer substitutes are more likely to be low-absence types. Reassuringly, this estimate (0.4 percentage points) is similar to the difference between the estimates with and without worker fixed effects in the full sample, which supports the interpretation that workers with few absence spells sort into jobs with low substitutability.

[34] This could reflect that employers spend more money on employee wellness for unique employees or increase the pressure not to be absent.

Table 3. Behavior vs. sorting into jobs

Column:	(1)	(2)	(3)
Sample:	All workers	All workers	New hires
Outcome:	Present absence	Present absence	Pre-hire absence
Mechanism:	Baseline	Behavior	Selection
Low substitutability	-0.0131***	-0.0058***	-0.0043***
	(0.0006)	(0.0007)	(0.0014)
Number of observations	5,863,497	5,863,497	387,901
Mean of dep. variable	0.109	0.109	0.116
Background controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes
Worker fixed effects	No	Yes	No

Notes: The standard errors are clustered on the establishment level in columns (1) and (3) and on the worker level in column (2). Low substitutability is defined as having 0-5 substitutes (i.e. the reference category is employees with more than 5 substitutes). The background controls are gender, age, education, birth country, having children aged 0–3 and establishment size. In column (3) we also control for the pre-hire employment status of the new hire. *, ***, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

3.5.2 Separations

So far, we have focused on employee absenteeism and selection into jobs. But in Table 4 we complement the analysis by asking how the realized sickness absence among new hires (measured as the average sickness absence probability in t and t+1) affects (i) the probability of exiting the employment relationship as well as (ii) the probability of receiving more substitutes within three years after entry. We study the first question by estimating Eq. (2) using a sample of new hires that stay in the establishment for at least one year (in order to be able to observe realized sickness absence). The outcome is an indicator for exiting the establishment between t+1 and t+2 (Panel A).[35] We study the second question using a sample of new hires that are observed in the establishment at least until t+3. The outcome is an indicator for having more substitutes in t+3 than in t (Panel B).

The results suggest that higher realized sickness absence is generally associated with significantly higher turnover rates (Panel A, column [1]), and a higher likelihood of receiving more substitutes (Panel B, column [1]).[36] This relationship is particularly strong for workers employed in jobs with low

[35] As a robustness check we have also used an indicator for not being observed in the establishment in either t+2 or t+3. This does not substantially change the results.

[36] Interestingly, when we condition on being observed in t+1 and t+2 and use the average sickness absence probability in t+1 and t+2 as an explaining variable for leaving the establishment in t+3 the estimate in Panel A, column (1), is substantially lower. This is consistent with the notion that the marginal effect of exhibiting bad properties (in this case high sickness absence), in relation to the job, on job separation probability should decrease with tenure (see Kwon [2005] for an interesting contribution on this topic).

internal substitutability (columns [2–4]), suggesting that sorting on the basis of sickness absence also occurs via the exit margin.

Table 4. Realized sickness absence and post-hire outcome

Column:	(1)	(2)	(3)	(4)
	All	Unique jobs	Not unique jobs	Difference
Outcome:		A: Separation in t+2		
Realized absence	0.1100***	0.1284***	0.1072***	0.1072***
	(0.0037)	(0.0097)	(0.0040)	(0.0040)
Realized abs. * Low subst.				0.0212**
				(0.0105)
Number of observations	336,026	63,624	272,402	336,026
Mean of dep. variable	0.270	0.280	0.267	0.270
Outcome:		B: More substitutes in t+3		
Realized absence	0.0145*	0.0576**	0.0120	0.0120
	(0.0075)	(0.0249)	(0.0079)	(0.0079)
Realized abs. * Low subst.				0.0456*
				(0.0261)
Number of observations	110,869	18,838	92,031	110,869
Mean of dep. variable	0.487	0.446	0.496	0.487
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes: In Panel A the sample is restricted to new hires that are observed in the establishment in *t*+1. Low substitutability is defined as having 0–5 substitutes (i.e. the reference category is employees with more than 5 substitutes). In Panel B the sample is restricted to new hires that are observed in the establishment in *t*+1, *t*+2 and *t*+3. The standard errors are clustered on the establishment level. The background controls are gender, age, education, birth country, having children aged 0–3 and establishment size. In column (4) all variables are interacted with the variable indicating 0–5 substitutes. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

3.6 The role of information

The fact that job separations respond to realizations of sickness absence, suggests that matches are formed under some remaining uncertainty. In this section, we examine the direct role of information for the selection into and out of jobs. We use the information proxies described in Section 2 to assess the degree of uncertainty in the hiring stage: (i) an indicator for being employed in *t*-1 to *t*-3 (*Pre-hire employment)*, (ii) an indicator for previous employment in another establishment within the same firm (*Firm connection)* and (iii) an indicator for previous employment in the same establishment as an incumbent employee (*Coworker connection)*.[37]

[37] When we use the previous firm connection as the information proxy we relax the new hire definition and include new hires in the workplace with a history within the firm.

If employers are reluctant to hire applicants with an observable history of sickness absence, we expect that better information should be associated with lower pre-hire absence among new hires. This is also what we see in the first row of panel A of Table 5: hires have between one and two percentage points lower pre-hire absence for two out of the three information measures. Importantly, the negative relationship between information availability and pre-hire sickness appears about twice as strong in jobs with few substitutes as in jobs with many substitutes. Thus, when employers are recruiting for positions with few internal substitutes they react even stronger to information about worker absence type, which suggests that screening for low absence workers seems to be more important when there is low internal substitutability of workers.

Panel B shows how job separations induced by realized sickness absence are related to the information available in the hiring stage. Intuitively, separations should respond more to realized sickness absence if there is less information about absence type beforehand. For simplicity, we restrict this analysis to jobs with few substitutes (≤ 5) and interact realized absence with our information proxies. Consistent with the results in Panel A of Table 4, there is a strong relationship between realized sickness absence and the probability of job separation. However, this relationship is weaker when the match was based on more precise information (suggested by the interaction terms). Depending on the information proxy, the point estimates are between 4 and 7 percentage points lower when there was more information, although the difference is not statistically significant when we use pre-hire employment as the information proxy (see column [1]). In sum, these findings suggest that matches formed with less precise information are more likely to be affected by revelations of worker absence type.

Table 5. The role of information

Column:	(1)	(2)	(3)
Information proxy:	Pre-hire employment	Firm connection	Coworker connection
Outcome:	A: Pre-hire sickness absence		
Sample:	All new hires		
Better informed (base)	-0.0208***	-0.0093***	-0.0008
	(0.0015)	(0.0011)	(0.0010)
Better inf. * Low subst.	-0.0249***	-0.0065***	-0.0045
	(0.0043)	(0.0024)	(0.0029)
Number of observations	387,901	586,994	387,901
Mean of dep. variable	0.116	0.115	0.116
Outcome:	B: Separation in t+2		
Sample	New hires with 0–5 substitutes		
Realized absence (base)	0.1575***	0.1284***	0.1318***
	(0.0390)	(0.0097)	(0.0113)
Realized abs. * Better inf.	-0.0441	-0.0777***	-0.0481*
	(0.0404)	(0.0157)	(0.0273)
Number of observations	63,624	95,236	63,624
Mean of dep. variable	0.280	0.276	0.280
Background controls	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes

Notes: The standard errors are clustered on the establishment level. The background controls are gender, age, education, birth country, having children aged 0–3, pre-hire employment status (not included in column [1] of Panel A or in Panel B) and establishment size. Panel A: All variables are interacted with the indicator for low substitutability, defined as having 0–5 substitutes (i.e. the reference category is employees with more than 5 substitutes). Panel B: All variables are interacted with the information proxy. The sample corresponds to the sample in column (2) of Panel A in Table 4. In column (2), we relax the new hire definition and include new hires with previous employment in another establishment within the same firm, which explains why the sample size is larger than in columns (1) and (3). *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

4 Conclusions

We have documented that workers matched to jobs with few internal substitutes are significantly less absent from work compared with other workers in the same narrowly defined occupations. The difference is substantial and holds regardless of whether if we look at employees’ own sickness absence or absences among parents caused by child sickness. Parents working in jobs with lower employee substitutability shift part of their child sickness absence spells to their partners.

About half of the correlation remains when we account for worker fixed effects, suggesting that both sorting based on pre-hire absence types and on-the-job changes in absence behavior are important mechanisms behind the strong association between sickness absence and employee substitutability. But sorting also occurs via the exit margin, as job separations respond to realizations of absence among new hires, particularly when they have few substitutes.

In addition, we find suggestive evidence that employee selection is more pronounced when there was more information about the workers' absence types beforehand. Thus, screening leads to more efficient matching between workers of different absence types and jobs with different opportunities for internal replacement. Finally, we find that the separation response due to realized sickness absence among workers in jobs with few substitutes is negatively related to the amount of information in the hiring stage, suggesting that learning about match quality is an important determinant of turnover rates, as in Jovanovic (1979).

Overall, our results highlight the importance of internal labor markets for firms to handle the costs of production disruptions caused by work absence. For jobs with low internal substitutability, sickness absence is a significant determinant in the selection process of new workers. But the difficulties of perfectly predicting the absence propensity of new employees leads to mismatches between workers and firms and, in turn, job separations. Our findings thus validate previous theoretical and empirical work on the importance of sorting and point at sickness absence as a previously unexplored dimension of match quality.

From the worker's perspective, our findings suggest that episodes of sickness absence affect the chances of accessing and retaining unique positions, which account for a significant share of the labor market. Hence, workers have strong incentives to keep absence low in jobs with low internal substitutability, which they do by e.g. shifting child care to their partners. In future work, it would be valuable to further explore if there is more scope for statistical discrimination against workers with above-average sickness absence rates at the group level in unique positions; e.g. women with children or workers from the upper part of the age distribution. Further explorations of the interplay between job characteristics and the allocation of time within the household could also potentially enhance our understanding of the systematic gender pay differences in modern labor markets.

References

Abowd, J. M., J. Haltiwanger, J. Lane, K. L. McKinsey, and K. Sandusky. 2007. "Technology and the demand for skill: an analysis of within and between firm differences". NBER Working Paper 13043.

Altonji, J. 2005. "Employer learning, statistical discrimination and occupational attainment". *American Economic Review*, 95(2), 112–117.

Andersson, F., M. Freedman, J. Haltiwanger, J. Lane and K. Shaw. 2009. “Reaching for the stars: who pays for talent in innovative industries?”. *The Economic Journal*, 119(538), F308–F332.

Barmby, T. and G. Stephan. 2000. “Worker absenteeism: why firm size may matter”. *The Manchester School*, 68(5), 568–577.

Black, C. and D. Frost. 2011. “Health at work – and independent review of sickness absence”. Presented to the UK parliament in November, 2011.

Boye, K. 2015. “Care more, earn less? The association between care leave for sick children and wage among Swedish parents”. IFAU Working Paper 2015:18.

Dionne, G. and B. Dostie. 2007. “New evidence on the determinants of absenteeism using linked employer-employee data”. *Industrial & Labor Relations Review*, 61(1), 108–120.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker. 2016. “Referral-based job search networks”. *The Review of Economic Studies*, 83(2), 514–546.

Eriksson, S., P. Johansson, and S. Langenskiöld. 2012. “What is the right profile for getting a job? A stated choice experiment of the recruitment process”. IFAU Working Paper 2012:13.

Eriksson, S. and D. Rooth. 2014. “Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment”. *American Economic Review*, 104(3), 1014–1039.

Farber, H. and R. Gibbons. 1996. “Learning and wage dynamics”. *The Quarterly Journal of Economics*, 111(4), 1007–1047.

Fredriksson, P., L. Hensvik, and O. Nordström Skans. 2015. “Mismatch of talent: evidence on match quality, job mobility and entry wages”. IFAU Working Paper 2015:26.

German Federal Statistical Office. 2011. “Statistical Yearbook 2011 for the Federal Republic of Germany”. Stuttgart: Metzler-Poeschel.

Goldin, C. and L. Katz. Forthcoming. “The most egalitarian of all professions: pharmacy and the evolution of a family-friendly occupation”. *Journal of Labor Economics*.

Hensvik, L. and O. Nordström Skans. Forthcoming. “Social networks, employee selection and labor market outcomes”. *Journal of Labor Economics*.

Jackson, K. 2013. “Match quality, worker productivity, and worker mobility: direct evidence from teachers”. *Review of Economics and Statistics*, 95(4), 1096–1116

Jovanovic, B. 1979. “Job matching, and the theory of turnover“. *Journal of Political Economy*, 87, 972–990.

Jäger, S. 2015. “How substitutable are workers? Evidence from worker deaths”.

Kwon, I. 2005. “Threat of dismissal: incentive or sorting?”. *Journal of Labor Economics*, 23(4), 797–838.

Lange, F. 2007. “The speed of employer learning”. *Journal of Labor Economics*, 25(1), 1–35.

Lazear, E. 1998. “Hiring risky workers”. In *Internal Labour Markets, Incentives, and Employment*. Ed. by I. Ohashi and T. Tachibanki. St. Martin’s Press.

Lazear, E. 2009. “Firm-specific human capital: a skill-weights approach”. *Journal of Political Economy*, 117(5), 914–940.

Lazear, E. and P. Oyer. 2012. “Personnel Economics”. In *Handbook of Organizational Economics*. Ed. by R. Gibbons and J. Roberts. Princeton University Press, 479–519.

Lindgren, K. 2012. “Workplace size and sickness absence transitions”. IFAU Working Paper 2012:26.

Mörk, E., A. Sjögren, and H. Svaleryd. 2013. “Childcare costs and the demand for children—evidence from a nationwide reform”. *Journal of Population Economics*, 26(1), 33–65.

Nagypál, É. 2007. “Learning by doing vs. learning about match quality: can we tell them apart?”. *The Review of Economic Studies*, 74(2), 537–566.

Office for National Statistics (UK). 2014. “Full report: sickness absence in the labour market, February 2014”.

Ose, S. 2005. “Working conditions, compensation and absenteeism”. *Journal of Health Economics*, 24(1), 161–188.

Oyer, P. and S. Schaefer. 2011. “Personnel economics: hiring and incentives”. In *Handbook of Labor Economics*. Ed. by O. C. Ashenfelter and D. Card. Vol. 4B. Elsevier, 1769–1823.

Previa (private company in the Swedish health industry). 2013. “Arbetsgivarnas kostnader för sjukfrånvaro ökade med en miljard”. Press release, April 23, 2013.

Sattinger, M. 1975. “Comparative advantage and the distribution of earnings and abilities“. *Econometrica*, 43, 455–468.

Schönberg, U. 2007. “Testing for asymmetric employer learning”. *Journal of Labor Economics*, 25(4), 651–691.

Statistics Sweden. 2007. “Labour income and public transfers in the Nordic countries in the 1990’s”. Economic Welfare 2007:1.

Tinbergen, J. 1956. “On the theory of income distribution”. *Weltwirtschaftliches Archiv*, 77, 156–173.

Appendix A: Additional figures and tables

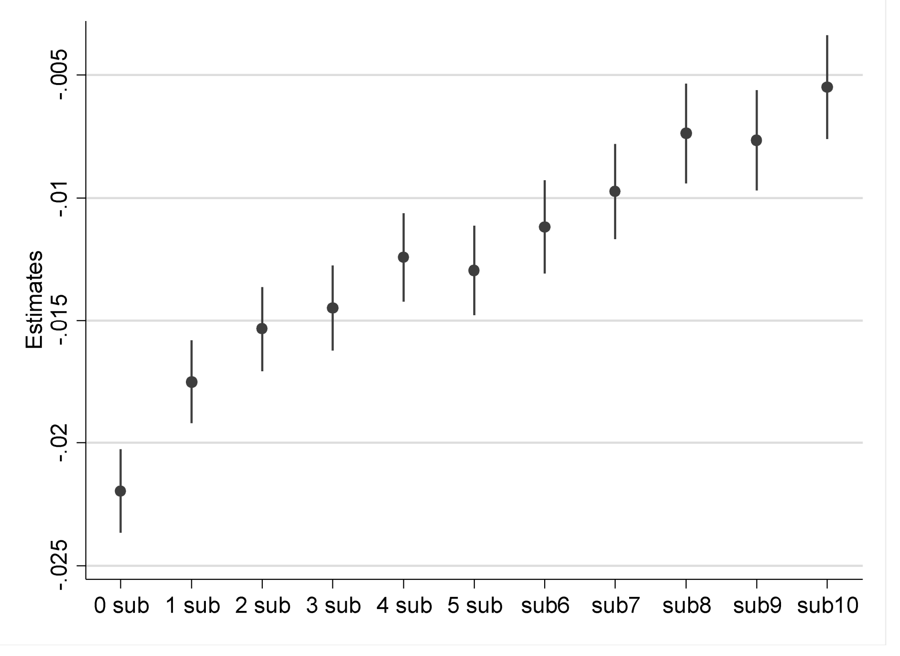


Figure A1. Sickness absence and the number of internal substitutes.

Notes: The figure shows the estimated coefficients on dummies for 0–10 substitutes in Eq. (1). The reference category is employees with more than 10 substitutes and the background controls are gender, age, education, birth country, having children aged 0–3 and establishment size. The model also includes year, occupational and establishment fixed effects. Standard errors are clustered on the establishment level.

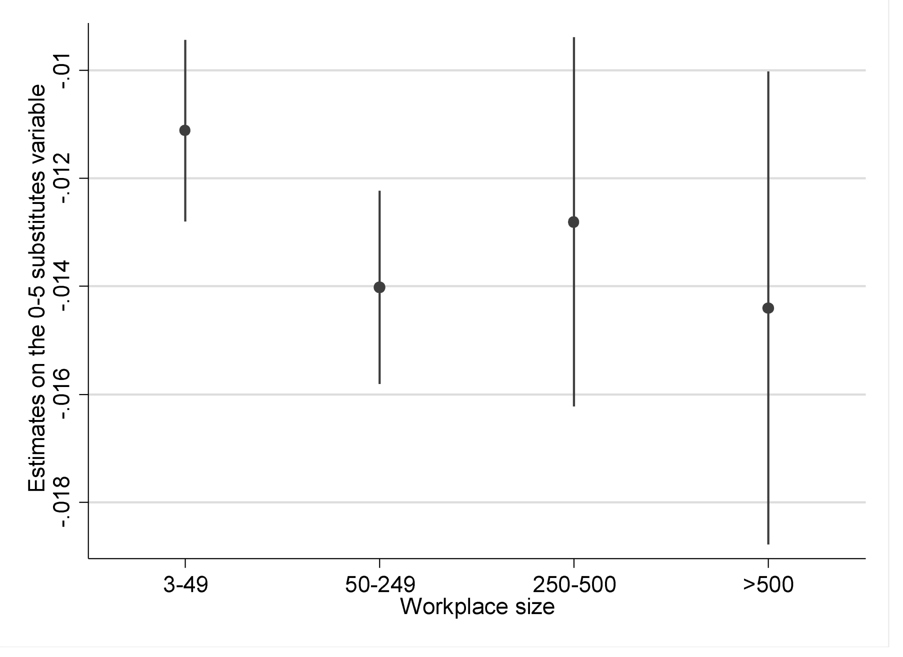


Figure A2. Sickness absence and workplace size.

Notes: The figure shows the results from separate estimations of Eq. (1) by establishment size. The independent variable of interest is an indicator for having 0–5 internal substitutes. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. The model also includes year fixed effects, occupational fixed effects and establishment fixed effects. The standard errors are clustered on the establishment level.

Table A1. Descriptive statistics for all employees

	All	0–5 substitutes	>5 substitutes
Establishment characteristics			
Unique position	0.036	0.204	0.000
Establishment size	570.9	90.8	675.7
Demographics			
Age	40.6	42.6	40.2
Male	0.629	0.495	0.658
Number of children age 0–17	0.829	0.836	0.828
Country of origin			
Sweden	0.913	0.939	0.908
Rest of Nordic countries	0.032	0.026	0.033
Rest of Europe	0.026	0.018	0.027
North America	0.001	0.001	0.001
South America	0.006	0.003	0.007
Rest of the world	0.022	0.012	0.024
Education			
Pre high school education (< 9 years)	0.025	0.021	0.026
Pre high school education ((≥ 9 years)	0.054	0.042	0.057
High school education max 2 years	0.413	0.386	0.418
High school education 2–3 years	0.225	0.233	0.223
Post high school education (< 3 years)	0.143	0.169	0.138
Post high school education (≥ 3 years)	0.131	0.143	0.128
Postgraduate education	0.008	0.004	0.008
Wage and Benefits			
Monthly wage in *t* (SEK)	23,657	22,824	23,839
Sickness benefit recipient in *t*	0.109	0.098	0.111
Professions			
Professionals	0.165	0.191	0.159
Technicians	0.245	0.315	0.229
Clerks	0.124	0.212	0.105
Service workers and shop sales	0.088	0.093	0.087
Skilled agricultural and fishery	0.005	0.010	0.004
Craft and related trades workers	0.116	0.090	0.121
Plant and machine operators	0.196	0.043	0.230
Elementary occupations	0.062	0.047	0.065
Number of observations	5,863,497	1,050,017	4,813,480

Notes: The sample is based on private sector employees in Sweden in 1997–2007. Managers and labor market entrants are excluded. The distribution across occupations is reported at the 1-digit level of the occupation code.

Table A2. Descriptive statistics for new hires

	All	0–5 substitutes	>5 substitutes
Establishment characteristics			
Unique position	0.039	0.207	0.000
Establishment size	406.8	77.0	483.7
Demographics			
Age	35.7	37.4	35.3
Male	0.599	0.501	0.622
Number of children age 0–17	0.807	0.890	0.787
Country of origin:			
Sweden	0.898	0.928	0.891
Rest of Nordic countries	0.024	0.023	0.025
Rest of Europe	0.030	0.021	0.032
North America	0.001	0.001	0.001
South America	0.010	0.006	0.011
Rest of the world	0.065	0.041	0.071
Education			
Pre high school education (< 9 years)	0.008	0.006	0.008
Pre high school education ((≥ 9 years)	0.089	0.059	0.097
High school education max 2 years	0.343	0.332	0.346
High school education 2–3 years	0.270	0.270	0.270
Post high school education (< 3 years)	0.133	0.158	0.127
Post high school education (≥ 3 years)	0.150	0.172	0.145
Postgraduate education	0.006	0.004	0.006
Wages and Benefits			
Monthly wage in *t* (SEK)	22,413	22,226	22,457
Pre-hire sickness benefit recipient	0.116	0.125	0.114
Professions			
Professionals	0.171	0.186	0.168
Technicians	0.227	0.293	0.212
Clerks	0.128	0.207	0.110
Service workers and shop sales	0.133	0.113	0.137
Skilled agricultural and fishery	0.006	0.010	0.004
Craft and related trades workers	0.091	0.082	0.094
Plant and machine operators	0.165	0.052	0.192
Elementary occupations	0.079	0.058	0.084
Number of observations	387,901	73,366	314,535

Notes: The sample is based on private sector new hires in Sweden in 1997–2007. Managers and labor market entrants are excluded. The distribution across occupations is reported at the 1-digit level of the occupation code.

Table A3. Descriptive statistics with respect to establishment size

(1)	(2)	(3)	(4)	(5)	(6)
Establishment size	Share of establish-ments	Share of employees	Share of variation (0–5 subs.)	Share of variation (alt. 1)	Share of variation (alt. 2)
3–49	0.853	0.290	0.413	0.305	0.267
50–249	0.124	0.323	0.395	0.401	0.309
250–499	0.014	0.128	0.097	0.135	0.138
≥500	0.008	0.259	0.094	0.159	0.286

Notes: Columns (2) and (3) show the distribution of establishments and employees over establishment size. In columns (4)–(6) we show the share of the variation in having few employee substitutes, conditional on all covariates in Eq. (1) by establishment size. In column (4) we use our baseline definition (i.e. the number of coworkers in the same establishment and 3-digit occupation is 0–5). In column (5) we count workers as having few employee substitutes if they (i) have 0 substitutes in a workplace with 3–49 employees or (ii) have less than four substitutes in a workplace with 50–249 employees or (iii) have less than seven substitutes in a workplace with 250–500 employees or (iv) have less than 10 substitutes in a workplace with more than 500 employees. In column (6) we divide the number of substitutes by the total number of employees in the establishment and require that quotient to be lower than 0.03 for an employee to be regarded as having few substitutes. Both these alternative definitions of low internal substitutability generate a situation where more of the identifying variation comes from larger establishments.

Table A4. Alternative definitions of low internal substitutability

Column:	(1)	(2)	(3)	(4)
Outcome:	Present sickness absence			
Definition of low subst.	Baseline	Alt. def. 1	Alt. definition 2	
Low substitutability	-0.0131***	-0.0123	-0.0131***	
	(0.0006)	(0.0006)	(0.0007)	
Share of substitutes				0.027***
				(0.0012)
Number of observations	5,863,497	5,863,497	5,863,497	5,863,497
Mean of dep. variable	0.109	0.109	0.109	0.109
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes: The standard errors are clustered on the establishment level. See Table A3 for the alternative definitions of low internal substitutability. The background controls are gender, age, education, birth country, having small children, and establishment size. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

Table A5. Substitutability and wages

Column:	(1)	(2)
Outcome:	Log of monthly wage in t	
Low substitutability	0.0134***	0.0109***
	(0.0011)	(0.0010)
Number of observations	5,863,497	5,863,497
Mean of dependent variable	10.01	10.01
Background controls	No	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes: The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children, pre-hire employment status and establishment size. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

Table A6. Robustness checks

Column:	(1)	(2)	(3)	(4)
Outcome:		Present sickness absence		
	Baseline	With wage control	Including public sector	4-digit occupation code
Low substitutability	-0.0131***	-0.0128***	-0.0079***	-0.0112***
	(0.0006)	(0.0006)	(0.0004)	(0.0010)
Number of observations	5,863,497	5,863,497	12,160,539	1,656,960
Mean of dep. variable	0.109	0.109	0.125	0.105
Background controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes
Establishment fixed effects	Yes	Yes	Yes	Yes

Notes: The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. Column (1) repeats the estimate from Table 1, col (2). In column (2) wage is included in the model. In column (3) we include public sector employees. In column (4) we calculate the number of substitutes based on a 4-digit occupational code which is available for the years 2005–2007. The occupational fixed effects are also based on the 4-digit occupational code. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

Table A7. Absence on the extensive and intensive margin

Column:	(1)	(2)
Margin:	Extensive	Intensive
Outcome:	Incidence (baseline)	Log of benefits
	A. Present sickness absence	
Low substitutability	-0.0131***	-0.0195**
	(0.0006)	(0.0078)
Number of observations	5,863,497	638,409
Mean of dependent variable	0.109	4.822
	B. Own care leave	
Low substitutability	-0.0115***	-0.0381***
	(0.0017)	(0.0038)
Number of observations	1,911,734	1,249,558
Mean of dependent variable	0.654	3.605
	C. Partner's care leave	
Low substitutability	0.0087***	0.0143***
	(0.0018)	(0.0043)
Number of observations	1,767,118	977,270
Mean of dependent variable	0.553	3.427
Background controls	Yes	Yes
Year fixed effects	Yes	Yes
Occupation fixed effects	Yes	Yes
Establishment fixed effects	Yes	Yes

Notes: The standard errors are clustered on the establishment level. The reference category is employees with more than 5 substitutes. The background controls are gender, age, education, birth country, having small children and establishment size. The estimates in column (2) are conditional on having positive benefits (i.e. at least one spell) according to the measure of interest. *, **, and *** denote statistical significance at the 10-, 5-, and 1-percent level.

Economic Studies

1987:1 Haraldson, Marty: To Care and To Cure. A linear programming approach to national health planning in developing countries. 98 pp.

1989:1 Chryssanthou, Nikos: The Portfolio Demand for the ECU: A Transaction Cost Approach. 42 pp.

1989:2 Hansson, Bengt: Construction of Swedish Capital Stocks, 1963-87: An Application of the Hulten-Wykoff Studies. 37 pp.

1989:3 Choe, Byung-Tae: Some Notes on Utility Functions Demand and Aggregation. 39 pp.

1989:4 Skedinger, Per: Studies of Wage and Employment Determination in the Swedish Wood Industry. 89 pp.

1990:1 Gustafson, Claes-Håkan: Inventory Investment in Manufacturing Firms. Theory and Evidence. 98 pp.

1990:2 Bantekas, Apostolos: The Demand for Male and Female Workers in Swedish Manufacturing. 56 pp.

1991:1 Lundholm, Michael: Compulsory Social Insurance. A Critical Review. 109 pp.

1992:1 Sundberg, Gun: The Demand for Health and Medical Care in Sweden. 58 pp.

1992:2 Gustavsson, Thomas: No Arbitrage Pricing and the Term Structure of Interest Rates. 47 pp.

1992:3 Elvander, Nils: Labour Market Relations in Sweden and Great Britain. A Comparative Study of Local Wage Formation in the Private Sector during the 1980s. 43 pp.

12 Dillén, Mats: Studies in Optimal Taxation, Stabilization, and Imperfect Competition. 1993. 143 pp.

13 Banks, Ferdinand E.: A Modern Introduction to International Money, Banking and Finance. 1993. 303 pp.

14 Mellander, Erik: Measuring Productivity and Inefficiency Without Quantitative Output Data. 1993. 140 pp.

15 Ackum Agell: Susanne: Essays on Work and Pay. 1993. 116 pp.

16 Eriksson, Claes: Essays on Growth and Distribution. 1994. 129 pp.

17 Banks, Ferdinand E.: A Modern Introduction to International Money, Banking and Finance. 2nd version, 1994. 313 pp.

18 Apel, Mikael: Essays on Taxation and Economic Behavior. 1994. 144 pp.

19 Dillén, Hans: Asset Prices in Open Monetary Economies. A Contingent Claims Approach. 1994. 100 pp.

20 Jansson, Per: Essays on Empirical Macroeconomics. 1994. 146 pp.

21 Banks, Ferdinand E.: A Modern Introduction to International Money, Banking, and Finance. 3rd version, 1995. 313 pp.

22 Dufwenberg, Martin: On Rationality and Belief Formation in Games. 1995. 93 pp.

23 Lindén, Johan: Job Search and Wage Bargaining. 1995. 127 pp.

24 Shahnazarian, Hovick: Three Essays on Corporate Taxation. 1996. 112 pp.

25 Svensson, Roger: Foreign Activities of Swedish Multinational Corporations. 1996. 166 pp.

26 Sundberg, Gun: Essays on Health Economics. 1996. 174 pp.

27 Sacklén, Hans: Essays on Empirical Models of Labor Supply. 1996. 168 pp.

28 Fredriksson, Peter: Education, Migration and Active Labor Market Policy. 1997. 106 pp.

29 Ekman, Erik: Household and Corporate Behaviour under Uncertainty. 1997. 160 pp.

30 Stoltz, Bo: Essays on Portfolio Behavior and Asset Pricing. 1997. 122 pp.

31 Dahlberg, Matz: Essays on Estimation Methods and Local Public Economics. 1997. 179 pp.

32 Kolm, Ann-Sofie: Taxation, Wage Formation, Unemployment and Welfare. 1997. 162 pp.

33 Boije, Robert: Capitalisation, Efficiency and the Demand for Local Public Services. 1997. 148 pp.

34 Hort, Katinka: On Price Formation and Quantity Adjustment in Swedish Housing Markets. 1997. 185 pp.

35 Lindström, Thomas: Studies in Empirical Macroeconomics. 1998. 113 pp.

36 Hemström, Maria: Salary Determination in Professional Labour Markets. 1998. 127 pp.

37 Forsling, Gunnar: Utilization of Tax Allowances and Corporate Borrowing. 1998. 96 pp.

38 Nydahl, Stefan: Essays on Stock Prices and Exchange Rates. 1998. 133 pp.

39 Bergström, Pål: Essays on Labour Economics and Econometrics. 1998. 163 pp.

40 Heiborn, Marie: Essays on Demographic Factors and Housing Markets. 1998. 138 pp.

41 Åsberg, Per: Four Essays in Housing Economics. 1998. 166 pp.

42 Hokkanen, Jyry: Interpreting Budget Deficits and Productivity Fluctuations. 1998. 146 pp.

43 Lunander, Anders: Bids and Values. 1999. 127 pp.

44 Eklöf, Matias: Studies in Empirical Microeconomics. 1999. 213 pp.

45 Johansson, Eva: Essays on Local Public Finance and Intergovernmental Grants. 1999. 156 pp.

46 Lundin, Douglas: Studies in Empirical Public Economics. 1999. 97 pp.

47 Hansen, Sten: Essays on Finance, Taxation and Corporate Investment. 1999. 140 pp.

48 Widmalm, Frida: Studies in Growth and Household Allocation. 2000. 100 pp.

49 Arslanogullari, Sebastian: Household Adjustment to Unemployment. 2000. 153 pp.

50 Lindberg, Sara: Studies in Credit Constraints and Economic Behavior. 2000. 135 pp.

51 Nordblom, Katarina: Essays on Fiscal Policy, Growth, and the Importance of Family Altruism. 2000. 105 pp.

52 Andersson, Björn: Growth, Saving, and Demography. 2000. 99 pp.

53 Åslund, Olof: Health, Immigration, and Settlement Policies. 2000. 224 pp.

54 Bali Swain, Ranjula: Demand, Segmentation and Rationing in the Rural Credit Markets of Puri. 2001. 160 pp.

55 Löfqvist, Richard: Tax Avoidance, Dividend Signaling and Shareholder Taxation in an Open Economy. 2001. 145 pp.

56 Vejsiu, Altin: Essays on Labor Market Dynamics. 2001. 209 pp.

57 Zetterström, Erik: Residential Mobility and Tenure Choice in the Swedish Housing Market. 2001. 125 pp.

58 Grahn, Sofia: Topics in Cooperative Game Theory. 2001. 106 pp.

59 Laséen, Stefan: Macroeconomic Fluctuations and Microeconomic Adjustments: Wages, Capital, and Labor Market Policy. 2001. 142 pp.

60 Arnek, Magnus: Empirical Essays on Procurement and Regulation. 2002. 155 pp.

61 Jordahl, Henrik: Essays on Voting Behavior, Labor Market Policy, and Taxation. 2002. 172 pp.

62 Lindhe, Tobias: Corporate Tax Integration and the Cost of Capital. 2002. 102 pp.

63 Hallberg, Daniel: Essays on Household Behavior and Time-Use. 2002. 170 pp.

64 Larsson, Laura: Evaluating Social Programs: Active Labor Market Policies and Social Insurance. 2002. 126 pp.

65 Bergvall, Anders: Essays on Exchange Rates and Macroeconomic Stability. 2002. 122 pp.

66 Nordström Skans, Oskar: Labour Market Effects of Working Time Reductions and Demographic Changes. 2002. 118 pp.

67 Jansson, Joakim: Empirical Studies in Corporate Finance, Taxation and Investment. 2002. 132 pp.

68 Carlsson, Mikael: Macroeconomic Fluctuations and Firm Dynamics: Technology, Production and Capital Formation. 2002. 149 pp.

69 Eriksson, Stefan: The Persistence of Unemployment: Does Competition between Employed and Unemployed Job Applicants Matter? 2002. 154 pp.

70 Huitfeldt, Henrik: Labour Market Behaviour in a Transition Economy: The Czech Experience. 2003. 110 pp.

71 Johnsson, Richard: Transport Tax Policy Simulations and Satellite Accounting within a CGE Framework. 2003. 84 pp.

72 Öberg, Ann: Essays on Capital Income Taxation in the Corporate and Housing Sectors. 2003. 183 pp.

73 Andersson, Fredrik: Causes and Labor Market Consequences of Producer Heterogeneity. 2003. 197 pp.

74 Engström, Per: Optimal Taxation in Search Equilibrium. 2003. 127 pp.

75 Lundin, Magnus: The Dynamic Behavior of Prices and Investment: Financial Constraints and Customer Markets. 2003. 125 pp.

76 Ekström, Erika: Essays on Inequality and Education. 2003. 166 pp.

77 Barot, Bharat: Empirical Studies in Consumption, House Prices and the Accuracy of European Growth and Inflation Forecasts. 2003. 137 pp.

78 Österholm, Pär: Time Series and Macroeconomics: Studies in Demography and Monetary Policy. 2004. 116 pp.

79 Bruér, Mattias: Empirical Studies in Demography and Macroeconomics. 2004. 113 pp.

80 Gustavsson, Magnus: Empirical Essays on Earnings Inequality. 2004. 154 pp.

81 Toll, Stefan: Studies in Mortgage Pricing and Finance Theory. 2004. 100 pp.

82 Hesselius, Patrik: Sickness Absence and Labour Market Outcomes. 2004. 109 pp.

83 Häkkinen, Iida: Essays on School Resources, Academic Achievement and Student Employment. 2004. 123 pp.

84 Armelius, Hanna: Distributional Side Effects of Tax Policies: An Analysis of Tax Avoidance and Congestion Tolls. 2004. 96 pp.

85 Ahlin, Åsa: Compulsory Schooling in a Decentralized Setting: Studies of the Swedish Case. 2004. 148 pp.

86 Heldt, Tobias: Sustainable Nature Tourism and the Nature of Tourists' Cooperative Behavior: Recreation Conflicts, Conditional Cooperation and the Public Good Problem. 2005. 148 pp.

87 Holmberg, Pär: Modelling Bidding Behaviour in Electricity Auctions: Supply Function Equilibria with Uncertain Demand and Capacity Constraints. 2005. 43 pp.

88 Welz, Peter: Quantitative new Keynesian macroeconomics and monetary policy 2005. 128 pp.

89 Ågren, Hanna: Essays on Political Representation, Electoral Accountability and Strategic Interactions. 2005. 147 pp.

90 Budh, Erika: Essays on environmental economics. 2005. 115 pp.

91 Chen, Jie: Empirical Essays on Housing Allowances, Housing Wealth and Aggregate Consumption. 2005. 192 pp.

92 Angelov, Nikolay: Essays on Unit-Root Testing and on Discrete-Response Modelling of Firm Mergers. 2006. 127 pp.

93 Savvidou, Eleni: Technology, Human Capital and Labor Demand. 2006. 151 pp.

94 Lindvall, Lars: Public Expenditures and Youth Crime. 2006. 112 pp.

95 Söderström, Martin: Evaluating Institutional Changes in Education and Wage Policy. 2006. 131 pp.

96 Lagerström, Jonas: Discrimination, Sickness Absence, and Labor Market Policy. 2006. 105 pp.

97 Johansson, Kerstin: Empirical essays on labor-force participation, matching, and trade. 2006. 168 pp.

98 Ågren, Martin: Essays on Prospect Theory and the Statistical Modeling of Financial Returns. 2006. 105 pp.

99 Nahum, Ruth-Aïda: Studies on the Determinants and Effects of Health, Inequality and Labour Supply: Micro and Macro Evidence. 2006. 153 pp.

100 Žamac, Jovan: Education, Pensions, and Demography. 2007. 105 pp.

101 Post, Erik: Macroeconomic Uncertainty and Exchange Rate Policy. 2007. 129 pp.

102 Nordberg, Mikael: Allies Yet Rivals: Input Joint Ventures and Their Competitive Effects. 2007. 122 pp.

103 Johansson, Fredrik: Essays on Measurement Error and Nonresponse. 2007. 130 pp.

104 Haraldsson, Mattias: Essays on Transport Economics. 2007. 104 pp.

105 Edmark, Karin: Strategic Interactions among Swedish Local Governments. 2007. 141 pp.

106 Oreland, Carl: Family Control in Swedish Public Companies. Implications for Firm Performance, Dividends and CEO Cash Compensation. 2007. 121 pp.

107 Andersson, Christian: Teachers and Student Outcomes: Evidence using Swedish Data. 2007. 154 pp.

108 Kjellberg, David: Expectations, Uncertainty, and Monetary Policy. 2007. 132 pp.

109 Nykvist, Jenny: Self-employment Entry and Survival - Evidence from Sweden. 2008. 94 pp.

110 Selin, Håkan: Four Empirical Essays on Responses to Income Taxation. 2008. 133 pp.

111 Lindahl, Erica: Empirical studies of public policies within the primary school and the sickness insurance. 2008. 143 pp.

112 Liang, Che-Yuan: Essays in Political Economics and Public Finance. 2008. 125 pp.

113 Elinder, Mikael: Essays on Economic Voting, Cognitive Dissonance, and Trust. 2008. 120 pp.

114 Grönqvist, Hans: Essays in Labor and Demographic Economics. 2009. 120 pp.

115 Bengtsson, Niklas: Essays in Development and Labor Economics. 2009. 93 pp.

116 Vikström, Johan: Incentives and Norms in Social Insurance: Applications, Identification and Inference. 2009. 205 pp.

117 Liu, Qian: Essays on Labor Economics: Education, Employment, and Gender. 2009. 133 pp.

118 Glans, Erik: Pension reforms and retirement behaviour. 2009. 126 pp.

119 Douhan, Robin: Development, Education and Entrepreneurship. 2009.

120 Nilsson, Peter: Essays on Social Interactions and the Long-term Effects of Early-life Conditions. 2009. 180 pp.

121 Johansson, Elly-Ann: Essays on schooling, gender, and parental leave. 2010. 131 pp.

122 Hall, Caroline: Empirical Essays on Education and Social Insurance Policies. 2010. 147 pp.

123 Enström-Öst, Cecilia: Housing policy and family formation. 2010. 98 pp.

124 Winstrand, Jakob: Essays on Valuation of Environmental Attributes. 2010. 96 pp.

125 Söderberg, Johan: Price Setting, Inflation Dynamics, and Monetary Policy. 2010. 102 pp.

126 Rickne, Johanna: Essays in Development, Institutions and Gender. 2011. 138 pp.

127 Hensvik, Lena: The effects of markets, managers and peers on worker outcomes. 2011. 179 pp.

128 Lundqvist, Heléne: Empirical Essays in Political and Public. 2011. 157 pp.

129 Bastani, Spencer: Essays on the Economics of Income Taxation. 2012. 257 pp.

130 Corbo, Vesna: Monetary Policy, Trade Dynamics, and Labor Markets in Open Economies. 2012. 262 pp.

131 Nordin, Mattias: Information, Voting Behavior and Electoral Accountability. 2012. 187 pp.

132 Vikman, Ulrika: Benefits or Work? Social Programs and Labor Supply. 2013. 161 pp.

133 Ek, Susanne: Essays on unemployment insurance design. 2013. 136 pp.

134 Österholm, Göran: Essays on Managerial Compensation. 2013. 143 pp.

135 Adermon, Adrian: Essays on the transmission of human capital and the impact of technological change. 2013. 138 pp.

136 Kolsrud, Jonas: Insuring Against Unemployment 2013. 140 pp.

137 Hanspers, Kajsa: Essays on Welfare Dependency and the Privatization of Welfare Services. 2013. 208 pp.

138 Persson, Anna: Activation Programs, Benefit Take-Up, and Labor Market Attachment. 2013. 164 pp.

139 Engdahl, Mattias: International Mobility and the Labor Market. 2013. 216 pp.

140 Krzysztof Karbownik. Essays in education and family economics. 2013. 182 pp.

141 Oscar Erixson. Economic Decisions and Social Norms in Life and Death Situations. 2013. 183 pp.

142 Pia Fromlet. Essays on Inflation Targeting and Export Price Dynamics. 2013. 145 pp.

143 Daniel Avdic. Microeconometric Analyses of Individual Behavior in Public Welfare Systems. Applications in Health and Education Economics. 2014. 176 pp.

144 Arizo Karimi. Impacts of Policies, Peers and Parenthood on Labor Market Outcomes. 2014. 221 pp.

145 Karolina Stadin. Employment Dynamics. 2014. 134 pp.

146 Haishan Yu. Essays on Environmental and Energy Economics. 132 pp.

147 Martin Nilsson. Essays on Health Shocks and Social Insurance. 139 pp.

148 Tove Eliasson. Empirical Essays on Wage Setting and Immigrant Labor Market Opportunities. 2014. 144 pp.

149 Erik Spector. Financial Frictions and Firm Dynamics. 2014. 129 pp.

150 Michihito Ando. Essays on the Evaluation of Public Policies. 2015. 193 pp.

151 Selva Bahar Baziki. Firms, International Competition, and the Labor Market. 2015. 183 pp.

152 Fredrik Sävje. What would have happened? Four essays investigating causality. 2015. 229 pp.

153 Ina Blind. Essays on Urban Economics. 2015. 197 pp.

154 Jonas Poulsen. Essays on Development and Politics in Sub-Saharan Africa. 2015. 240 pp.

155 Lovisa Persson. Essays on Politics, Fiscal Institutions, and Public Finance. 2015. 137 pp.

156 Gabriella Chirico Willstedt. Demand, Competition and Redistribution in Swedish Dental Care. 2015. 119 pp.

157 Yuwei Zhao de Gosson de Varennes. Benefit Design, Retirement Decisions and Welfare Within and Across Generations in Defined Contribution Pension Schemes. 2016. 148 pp.

158 Johannes Hagen. Essays on Pensions, Retirement and Tax Evasion. 2016. 195 pp.

159 Rachatar Nilavongse. Housing, Banking and the Macro Economy. 2016. 156 pp.

160 Linna Martén. Essays on Politics, Law, and Economics. 2016. 150 pp.

161 Olof Rosenqvist. Essays on Determinants of Individual Performance and Labor Market Outcomes. 2016. 151 pp.