Essays on Public, Political and Labor Economics

David Seim

Stockholm University

ISSN 0346-6892
ISBN 978-91-7447-688-0

Cover Picture:
Printed in Sweden by PrintCenter US-AB, Stockholm 2013

Distributor: Institute for International Economic Studies

Doctoral Dissertation
Department of Economics
Stockholm University

Abstract

Real or Evasion Responses to the Wealth Tax? Theory and Evidence from Sweden This paper addresses the behavioral effects of an annual wealth tax. I use Swedish tax records over the period 1999–2006 to estimate bunching at kink points in the progressive tax schedule and find significant estimates of the implied net-of-tax elasticity of taxable net wealth at about 0.3. Exploiting features of the institutional setting, I decompose the effects into a reporting response and a saving response. The results suggest that an increase in the tax is likely to stimulate evasion rather than deter savings. I merge tax records with military enlistment records on cognitive ability and find that high-skilled individuals respond more to the tax. This suggests that the incidence of the tax falls disproportionally on the cognitively less able.

Job Displacement and Labor Market Outcomes by Skill Level This paper investigates the effects of displacement on outcomes such as annual earnings and unemployment. It relies on previously unexplored administrative data on all displaced workers in Sweden in 2002, 2003 and 2004 which are linked to employer-employee matched data at the individual level. By linking the data to military enlistment records, the paper assesses the selection into displacement. The results indicate that workers with low cognitive and noncognitive skills are significantly more likely to be displaced than high-skilled workers, even after controlling for tenure and education. The analysis of displacement effects shows evidence of large and long-lasting welfare costs from displacement. Studying the impacts of job displacement in terms of cognitive and noncognitive skills, reveals that workers with high skills fare better than low-skilled workers in absolute terms, but the recovery rates show no significant differences across skill groups. Finally, by using administrative data on displacements, it is possible to assess quantitatively the bias in earlier studies from not being able to separate quits from layoffs.

Complementary Roles of Connections and Performance in Political Selection in China Who becomes a top politician in China? We focus on provincial leaders, a pool of candidates for top political office, and examine how their chance of being promoted depends on performance – measured by provincial economic growth – and connections with top politicians – measured by past joint work in the same branch of government. A simple theoretical framework suggests that performance and connections may interact, an aspect ignored in the previous literature. Over the period 1993–2009,

we find a positive correlation between promotion and growth that is robustly stronger for connected provincial leaders than for unconnected ones. This evidence indicates that performance and connections are complements in the Chinese political selection process. Auxiliary evidence suggests that the documented promotion pattern does not distort the allocation of talent.

Does the Demand for Redistribution Rise or Fall with Cognitive Ability? What role does cognitive ability play for preferences for redistribution? We investigate this question using a unique Swedish data set that matches data from a survey about the demand for redistribution to administrative data. The latter dataset contains information on cognitive ability for men, collected at military enlistment. On a scale of 0 to 100 percent redistribution, we find that a one standard deviation increase in cognitive ability is associated with a reduction of willingness to redistribute income by 6 percent points. This result holds also when we take the positive correlation between cognitive ability and earnings capacity into account and control for long-run income. Furthermore, we show that the more cognitively able want less redistribution despite the fact that they express stronger altruistic preferences.

To my family

Acknowledgments

In the 1960's, my grandmother, a refugee from Estonia in her thirties, started to work as a librarian in the public libraries in the southern suburbs of Stockholm. She was delighted to discover that one of her colleagues was also of Baltic descent, with roots in Latvia. About ten years later, in 1973, their sons, my father Jaan Seim and Torsten Persson, enlisted in the interpreter division of the Swedish military together. Now, about 50 years after my grandmother, Mall Seim, started her job in the libraries together with Ludmila Persson and 40 years since my father did his military service with Torsten, I am about to finish my PhD with Torsten as my supervisor. They say that in this world of connectivity, everyone is just a couple of handshakes away from the US president. However, these statistically independent events have lead to me to believe that everyone is just a couple of handshakes away from Torsten Persson.

I owe gratitude to many people who made this dissertation possible, but I am particularly thankful to my advisor, Torsten Persson. Since I nervously started to work for him as a research assistant in 2007, Torsten has, with his enthusiasm and ability to inspire, taught me to trust my capabilities. I am truly grateful to Torsten for always being available for discussions, for sharing his vast knowledge and for supporting me at times when I needed it the most.

I would also like to thank Per Krusell, David Strömberg and Raj Chetty, who have been vast sources of wisdom and inspiration. Discussions and feedback from you have truly shaped my view on how to conduct academic research.

Since I started the PhD studies in 2007, some exceptionally bright people have crossed my path. I have learned a lot from discussions with Mark Bernard, Konrad Burchardi, Lars Calmfors, Pamela Campa, Tom Cunningham, Peter Fredriksson, John Hassler, Amanda Jakobsson, Ruixue Jia, Lisa Laun, Sebastian Koehne, Masa Kudamatsu, Martin Olsson, Conny Olovsson, Bei Qin, Andrei Shleifer, Jakob Svensson, Roine Vestman, David Yanagizawa-Drott, Yves Zenou and Robert Östling. Thank you for sharing your time and knowledge.

During my PhD, I had the privilege of spending two semesters at Harvard University and am grateful to everyone at the Department of Economics and in particular to Yehonatan Givati, Johanna Möllerström and László Sándor, who have become close friends. I would also like to thank the creative and inspiring Johannes Haushofer and Petra Persson whom I met at the Summer School in Behavioral Economics in Trento in 2010. Our choir singing has turned into friendship for life.

In 2006 and in 2011, I was given the opportunity to visit the Research Department of the Estonian central bank. I am very grateful for these experiences. Aitäh!

My dissertation would not have been possible without the support of the administrative staff at the IIES. Financial support from Handelsbanken's Research Foundations, Ragnar Söderberg's Foundation, the Royal Academy of Sciences and the Swedish Tax Payers is also gratefully acknowledged.

I would not have had the courage to start a PhD without my nestor and sidekick Andreas Nordvall Lagerås. I am truly grateful to Andreas for showing me the beauty of mathematics, for teaching me mathematical statistics and for inspiring me to think about big questions. *Uno sed leo.*

To my friends outside of academia: thank you for keeping me grounded and providing balance in my life. Special thanks to Christer Benktsson, Maria Fahlén, Erik Klarén and Sara Rosenquist. I look forward to more wine-tastings at Sara's place and pan-flute concertos at Mariatorget.

I would also like to thank my mother Ingrid, my father Jaan and my sister Anna Maret. This dissertation would not exist without the love and warmth that you so unconditionally provide.

Although obtaining the PhD is quite an achievement to take pride in, it does not come close to my greatest accomplishments as a PhD student. Playing soccer with the Department of Economics on Wednesdays turned out to be the best thing I have ever done: this is where I fell in love with Anna, now my wife, with whom I now have a gorgeous daughter, Hilda. Your love, support and patience mean everything to me!

Stockholm, April 2013

Table of Contents

Chapter 1

Introduction

How do government policies affect the economy? How should policies be designed in order to maximize the well-being of citizens? How do political institutions shape the policies actually chosen by politicians? This thesis seeks to contribute to our understanding of these overarching questions. It consists of four self-contained essays related to the fields of public, political and labor economics.

Real or Evasion Responses to the Wealth Tax? In the first essay, I analyze the behavioral responses of an annual tax on net wealth, i.e. on assets net of liabilities. Despite the wide use of annual wealth taxes in developed countries (e.g. France, Norway and Spain), investigations of the effectiveness of these policies are notably absent in the academic literature. The aim of this paper is to inject hard evidence of the workings of the wealth tax into the debate between proponents of wealth taxes, who argue that these policies are a direct and effective way of reducing wealth inequality, and the opponents of wealth taxes, who argue that they may hamper capital accumulation in the long run.

Using features of the Swedish institutional setting and data on Swedish tax returns at the individual level between 1999 and 2006, I employ two distinct identification strategies to estimate the elasticity of taxable wealth with respect to the wealth tax. First, I exploit the sharp discontinuities in marginal tax rates that result from the progressive wealth tax scheme. Such discontinuities create kinks in the budget sets of agents and under relatively unrestrictive assumptions on individuals' preferences, we should expect agents to bunch at the kink points, i.e. position their taxable net wealth just below the tax threshold. When applying the methods proposed by Saez [2010] and Chetty et al. [2011] to Swedish wealth data, I find that an increase in the tax by one percentage point leads to a reduction in taxable net wealth by 0.12-0.33 percent.

Second, shifts in the marginal taxes over time generate variation, which I exploit in the vein of Gruber and Saez [2002]. Specifically, I regress the change in taxable net wealth on the change in marginal tax rates. Since the change in the tax rate is an endogenous variable – a positive (negative) shock to wealth may mechanically place the individual in a higher (lower) tax bracket – I instrument for the change in marginal tax rates using the simulated change that would occur if base year wealth levels were held constant. The semi-elasticities estimated using this approach are larger and range from 0.45 to 0.80.

Having obtained these results, I analyze whether the responses to the wealth tax reflect households reducing actual wealth, i.e. whether savings decisions are distorted due to the wealth tax, or if the responses merely reflect households sheltering money from the government. First, analyzing the amount of bunching at kink points over time, I exploit the notion that if savings were distorted by the tax, a shift in the tax brackets would induce a gradual adjustment of the amount of bunching from the old kink to the new kink. Analyzing such shifts empirically reveals that adjustments occur instantaneously. This suggests that responses, to some extent, reflect evasion responses rather than real responses. Second, I exploit that the wealth tax base consisted of two parts. A major part of the tax base consisted of assets and liabilities reported directly to the tax agency by third-parties. Examples of such assets include real estate holdings and financial wealth. In addition, when filing their taxes, the tax payers were supposed to self-report assets and liabilities which were part of the tax base, but not third-party reported. Examples of such assets are cars, boats and informal loans within the family. Using administrative data on car holdings for the population of Swedes during these years – a database that the tax agency did not have access to – I compare the self-reported assets to car values at the individual level and find that only about 30 percent of car owners truthfully report their cars when filing the tax. Finally, to assess the nature of responses, I regress sentenced fines for tax cheating on self-reported downward adjustments of taxable net wealth and find a positive and statistically significant correlation. The larger the downward adjustments of taxable wealth, the larger the fines. These three pieces of evidence indicate that the responses to the wealth tax reflect tax evasion rather than real responses and that the annual wealth tax is likely to be an ineffective instrument.

In the final part of the paper, I investigate whether the wealth tax satisfies the concept of horizontal equity, namely that two individuals with the same tax liabilities are taxed at the same level. Chetty et al. [2009] argue that the salience of a tax matters for how individuals respond to tax changes. Here, I analyze whether there is heterogeneity in this dimension by cognitive skills. To this end, I link the tax records to military enlistment data, which include a measure of cognitive skills for basically all men born between 1951 and 1979. I do find evidence that the responses are larger among the cognitively skilled, also when controlling for basic differences in wealth across high- and low-skilled individuals.

Job Displacement and Labor Market Outcomes by Skill Level In the second essay, I investigate how the incidence and consequences of job loss vary by skills. There is a large literature in labor economics that establishes that individual costs of job displacement are large and long-lasting. Labor market outcomes, such as earnings and unemployment, are dramatically affected (Jacobson et al., 1996, Couch and Placzek, 2010) and mortality increases and health deteriorates (Sullivan and von Wachter, 2009, Eliason and Storrie, 2009). Even children's school grades are negatively affected by parental job loss (Rege et al., 2010).

To understand the mechanisms behind these negative outcomes and to, ideally, guide decision-makers in choosing the right policies, I investigate the role of general skills in the context of job displacement. Cognitive and noncognitive skills have been shown to be extraordinarily important traits for success in the labor market (Lindqvist and Vestman, 2011, Heckman et al., 2006). In this paper, I analyze whether such skills matter for (i) the probability of being displaced when a plant downsizes and (ii) the ability to recover upon losing the job.

I use a large and previously unexplored administrative database. In Sweden, employers wishing to lay off five or more workers simultaneously, or 20 or more workers within a 90 day period, must report this to the Public Employment Service.

Linking this dataset to administrative data, including employer-employee matched records and the military enlistment data used in essay 1, enables me to investigate the selection into job displacement. I find that, within plants that downsize, workers with high cognitive and noncognitive ability are significantly less likely to experience job loss than workers with low skill levels. An increase of one standard deviation in either skill measure decreases the probability of the worker being displaced by one percentage point. In the same analysis, age is shown to matter for the displacement decision: younger workers are significantly more likely be displaced when plants downsize.

Turning to the analysis of displacement effects on economic outcomes, the sample is different from the one above. Following the previous literature, I focus on displaced workers who lost their jobs prior to a plant closure. This is done to minimize the selection into displacement. However, in contrast to the previous literature, where workers separating from closing plants within a time period in before the closure are defined as laid-off, I am able to use administrative data to observe actual layoffs when assessing job displacements. Using the panel structure of the data, I regress different outcomes on a set of dummies indicating whether the worker was displaced and time since the displacement year. In line with the literature, my findings suggest that the labor earnings drop dramatically upon displacement. In the calender year after displacement, annual earnings are estimated at 23.2 percent lower levels compared to what they would have been had the displacement not occurred. These effects also prevail in the long run. Seven years after the displacement, labor earnings are 15 percent lower than the counterfactual.

Turning to the heterogenous effects, I find that the effects of job loss are equal across high- and low-skilled workers. This is true for both cognitive and noncognitive skills. I find some suggestive evidence of skills influencing the willingness to acquire new skills.

Cognitively able and younger workers are more likely to start studying after displacement, but there is no such pattern for noncognitive skills. On the other hand, less able workers are more likely to participate in job-training programs, irrespective of the skill measure used.

In the final part of this paper, I make a methodological contribution to the literature by addressing the bias that may plague previous studies that are unable to distinguish quits from layoffs. I estimate the effects of displacement, defining displacements as in the previous literature and compare them to the effects obtained in my study. With matched employer-employee data, separations are defined as job displacements if they occur within a time window of mass separations, or, within a time window preceding plant closure. My results suggest that the mass separation estimates significantly understate the effects of displacement, while the use of a three-year window prior to plant closure overestimates the effects. The most accurate estimates are obtained by defining displacements as separations occurring during a one-year window preceding plant closure.

Complementary Roles of Connections and Performance in Political Selection in China The third essay addresses the selection of top politicians in non-democratic countries. Specifically, who becomes a top politician in China? The spectacular economic performance of the Chinese economy in the past few decades suggests that the politicians selected to rule the country may have been conducive to growth, or at least not detrimental to economic development. Yet, we know relatively little about the selection of top politicians in nondemocratic countries, such as China. This paper seeks to fill that gap by studying the promotions of provincial leaders in China.

Several academic studies argue that the promotion pattern of provincial leaders in China is characterized by meritocracy. For example, Zhiyue [1996, 2002], Maskin et al. [2000] and Li and Zhou [2005] show that party officials in local governments are more likely to be promoted to higher levels of government if they achieve higher economic growth. On the other hand, anecdotal evidence and a more systematic empirical analysis by Shih et al. [2012] suggest that social connections to top leaders influence the chance of promotion.

In this paper, we conduct a systematic analysis of the joint role of these two factors. First, we show, theoretically, that depending on what the role of connections is, we expect different interplays between provincial economic growth and social connections. We introduce the concept of social connections in a simple career-concern model along the lines of Holmstrom [1982], where the Politburo Standing Committee (henceforth PSC) acts as a decision maker when agreeing on which provincial leaders to promote. The theory predicts that connections and economic growth are substitutes if connections convey information on the ability of the provincial leaders to the PSC. On the other hand, if connections instead reflect loyalty between top politicians and provincial leaders, growth and connections are complementary in determining the promotion probability.

To test the predictions of our theory, we collect biographical data on Chinese politicians in office between 1993 and 2009. We use these data to measure connections between provincial leaders and the top seven or nine party officials in the PSC, by whether they used to work in the same branch of the party or the government in the past. To measure performance, we follow the literature and use the real GDP growth of the province ruled by each leader.

We find that connected provincial leaders are, on average, significantly more likely to be promoted than unconnected ones. This difference in the propensity for promotion is driven by a stronger positive correlation between promotion and economic growth for connected officials. Low-performing provincial leaders are unlikely to be promoted irrespective of their connections, while connections increase the likelihood of promotion for high-performing provincial leaders. In other words, connections and performance appear to be complements in the promotion of provincial leaders in China. This evidence is consistent with our theoretical prediction that connections mainly play a role in fostering the loyalty of subordinates to top politicians. Moreover, we find that the complementarity is stronger for the connected pairs in which provincial leaders are substantially younger than the PSC members. Since Communist Party officials compete for high office with similar-aged peers, but not with senior peers, this finding provides further evidence in favor of the loyalty-fostering role of connections.

Does the Demand for Redistribution Rise or Fall with Cognitive Ability? Rising income and wealth inequality is a hotly debated topic among policy makers and scholars alike. Views on the extent to which resources should be redistributed differ greatly. This paper is part of an extensive research project that aims at a deeper understanding of the demand for redistribution, than what the previous literature provides.

In the seminal papers by Meltzer and Richards [1981] and Romer [1975], individuals' income levels determine the demand for redistribution. More recent work, comprising Fong [2001] and Piketty [1995], emphasizes that beliefs and perceptions about social mobility, e.g. the extent to which people think that luck or effort determines outcomes in life, matter for the demand for redistribution. People who think that income heterogeneity arises because of immutable and randomly allocated characteristics prefer redistributional policies as an insurance against bad draws. Gender has also been shown to be an important determinant of redistributional preferences (see Alesina and Giuliano, 2011, Alesina and Ferrara, 2005).

In this paper, we ask what role cognitive ability plays in the demand for redistribution. To this end, we tailor-made a survey with the purpose of eliciting preferences for redistribution and sent it to a representative sample of Swedes. We received responses from $1,549$ individuals and were able to link the survey data to administrative records that, among other variables, include the measure of cognitive ability from the military enlistment used in essays 1 and 2 in the thesis.

We find a negative and strong relationship between cognitive ability – measured at age 18 – and the demand for redistribution measured at ages $33-61$. This relation may be largely driven by differences in incomes. Cognitive skills have been shown to be important determinants of labor market outcomes such as annual earnings, see e.g. Lindqvist and Vestman [2011]. However, exploiting our rich data set that includes administrative records on historical earnings and survey data which detects perceptions of future income levels, we find that the relationship prevails also when controlling for historic, current as well as expected future income.

Trying to identify the mechanisms behind this result, we investigate the role of altruism and risk preferences. Benjamin et al. [2013] show that high-skilled individuals are more prone to take risks, which implies that they should demand less redistribution. On the other hand, Rustichini et al. [2012] provides evidence of cognitively able individuals being more altruistic which, in this context should lead to them to prefer more redistribution. Our survey contains questions aimed at detecting such preferences, developed in the laboratory setting by Falk and Becker [2010]. Neither variable succeeds in explaining the negative relation between cognitive ability and demand for redistribution.

Bibliography

A. Alesina and E. La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89:897–931, 2005.

A. Alesina and P. Giuliano. Handbook of social economics, 2011. A. Bisin, J. Benhabib and M. Jackson.

D. Benjamin, S. Brown, and J. Shapiro. Who is 'behavioral'? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 2013. URL `http://economics.cornell.edu/dbenjamin/iq.pdf`.

R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99:1145–1177, 2009.

R. Chetty, J. N. Friedman, T. Olsen, and L. Pistaferri. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics*, 126:749–804, 2011.

K. A. Couch and D. W. Placzek. Earnings losses of displaced workers revisited. *American Economic Review*, 100:572–589, 2010.

M. Eliason and D. Storrie. Job loss is bad for your health – swedish evidence on cause–specific hospitalization following involuntary job loss. *Social Science and Medicine*, 68: 1396–1406, 2009.

A. Falk and A. Becker. Economic preference module, 2010. University of Bonn.

C. Fong. Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics*, 82:225–246, 2001.

J. Gruber and E. Saez. The elasticity of taxable income: Evidence and implications. *Journal of Public Economics*, 84:1–32, 2002.

J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24:411–482, 2006.

B. Holmstrom. Managerial incentive problems: A dynamic perspective, 1982. Helsinki: Swedish School of Economics.

L. S. Jacobson, R. J. LaLonde, and D. G. Sullivan. Earnings losses of displaced workers. *American Economic Review*, 83:685–709, 1996.

H. Li and L-A. Zhou. Political turnover and economic performance: The incentive role of personnel control in china. *Journal of Public Economics*, 89:1743–1762, 2005.

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

E. Maskin, Y. Qian, and C. Xu. Incentives, information and organizational form. *Review of Economic Studies*, 67:359–378, 2000.

A. Meltzer and S. Richards. A rational theory of the size of the government. *Journal of Political Economy*, 89:914–927, 1981.

T. Piketty. Social mobility and redistributive politics. *Quarterly Journal of Economics*, 110:551–584, 1995.

M. Rege, K. Telle, and M. Votruba. Parental job loss and children's school performance. *Review of Economic Studies*, 78:1462–1489, 2010.

T. Romer. Individual welfare, majority voting and the properties of a linear income tax. *Journal of Public Economics*, 7:163–188, 1975.

A. Rustichini, C. DeYoung, J. Anderson, and S. Burks. Toward the integration of personality theory and decision theory in the explanation of economic and health behavior, 2012.

E. Saez. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2:180–212, 2010.

V. Shih, C. Adolph, and M. Liu. Getting ahead in the communist party: Explaining the advancement of central committe members in china. *American Political Science Review*, 106:166–187, 2012.

D. Sullivan and T. von Wachter. Job displacement and mortality: An analysis using administrative data. *Quarterly Journal of Economics*, 124:1265–1306, 2009.

B. Zhiyue. Economic performance and political mobility: Chinese provincial leaders. *Journal of Contemporary China*, 5:135–154, 1996.

B. Zhiyue. *Chinese Provincial Leaders: Economic Performance and Political Mobility Since 1949*. M.E. Sharpe, 2002.

Chapter 2

Real or Evasion Responses to the Wealth Tax? Theory and Evidence from Sweden*

2.1 Introduction

Wealth taxes are surprisingly widespread given the lack of theoretical consensus on how wealth should be taxed, and the sparse knowledge about their effects. This paper addresses the latter problem. The effects that I document suggest a couple of reasons for why wealth taxes might be ineffective, inefficient, and unjust: the Swedish annual wealth tax was subject to evasion, in particular by households with higher cognitive ability.

Historically, many western economies have taxed net wealth and a number of countries, including France, Norway and Spain, still uphold these policies. Yet, wealth taxes remain controversial. Opponents of wealth taxation argue that the introduction of distortive taxes may hamper capital accumulation in the long run. In addition, a range of well-known problems are associated with the implementation of wealth taxes, including the difficulties to define a comprehensive tax base, appraise assets and prevent tax evasion (Adam et al., 2011, Brown, 1991). Proponents of wealth taxation, on the other hand, argue that it is a direct and effective way of reducing inequality. The question of whether or not wealth taxes are desirable is thus subject to a continuous debate that is far from settled.

In recent years, there has been renewed academic interest in the normative aspects of capital and wealth taxation. The proposition that capital taxes should be set to zero, as

*I am grateful to Torsten Persson, Konrad Burchardi, Raj Chetty, Tom Cunningham, John N. Friedman, Nathan Hendren, Per Krusell, Sebastian Koehne, Anna Larsson Seim, Petra Persson, László Sándor, Andrei Shleifer and David Strömberg, and seminar participants in Stockholm, Copenhagen, Oslo, Helsinki, Bergen, Uppsala, Berlin, Mannheim, Toulouse, Barcelona and Toronto for comments and suggestions. I also thank Christina Lönnblad for editorial assistance.

argued by Atkinson and Stiglitz [1976], Chamley [1986], Judd [1985], has been scrutinized and challenged.[1] Within this growing strand of literature, Diamond and Saez [2011] question assumptions made in the optimal taxation literature and propose a rationale for positive capital income taxes. Piketty and Saez [2012] derive expressions for optimal policy as a function of empirically estimable parameters.[2]

The empirical contribution of this paper is to analyze how wealth taxation works in practice. To quantify welfare costs of taxes, Feldstein [1995, 1999] shows that the tax elasticity of the tax base is a so-called sufficient statistic for welfare analysis since it incorporates all responses. When approaching wealth taxation from an empirical perspective, our first objective should thus be to estimate the effect of the wealth tax on taxable net wealth, a sufficient statistic for the ensuing utility flow. However, even if such an elasticity can be credibly estimated, it only provides part of the picture. To understand the workings of a wealth tax, margins of adjustments might still matter, not just the composite, reduced form response. Taxation typically affects not only the incentives to save, but may also trigger undesirable behavior in terms of tax sheltering. Chetty [2009] challenges the results in Feldstein [1995, 1999] and shows that a distinction between real responses and evasion responses becomes crucial for welfare analysis if the marginal tax dollar lost to evasion is tantamount to a transfer across agents.[3] As noted by Kopczuk [2013], this is an area where evidence is particularly scant. The issue of tax sheltering may be particularly relevant in the context of wealth taxation. Many taxes in advanced economies are based on information solicited from third parties, which makes them largely impervious to abuse (Kleven et al., 2011). The wealth tax base, however, also involves an element of self-reporting, which makes the wealth tax susceptible to lower compliance rates and tax evasion.

Assessing the effects on the tax base and decomposing these into real and reporting responses may help us understand important aspects of policy. But even with perfect knowledge of these effects, the story would remain incomplete. If the ability to comprehend the tax system differs across individuals, they may be asymmetrically affected by tax reforms. Traditional models of optimal taxation generally assume that all agents are rational and respond optimally to policy. In practice, opaque reforms and salient policies may trigger widely different responses.[4] For instance, the tendency to confound marginal

[1] Atkinson and Stiglitz [1976] argue that, in the presence of skill heterogeneity, the possibility of using non-linear income taxes implies that there is no role for capital taxes. Chamley [1986], Judd [1985] show that optimal capital taxes should be zero in the long run. Here the result depends on the infinite-horizon setting and a tax distortion which grows exponentially over time and cannot be optimal from the social planner's point of view.

[2] Boadway et al. [2010], Banks and Diamond [2010], Kopczuk [2013] provide extensive surveys of the literature on optimal wealth and capital income taxation.

[3] This decomposition hinges on evasion costs being transfers across agents. The effect on the tax base remains a sufficient statistic for welfare analysis when evasion is present and costs are not transfers but, for example, moral costs.

[4] See Chetty et al. [2009] for policy responses and salience.

and average tax rates indicate that individuals may not necessarily respond as predicted to a progressive wealth tax.[5] If comprehension of the tax scheme is unevenly distributed, the tax burden may fall disproportionately on the less able.

Motivated by this general background, I investigate the behavioral effects of the Swedish annual wealth tax, using a unique dataset, comprising about 58 million observations in a panel of individual taxpayers. I quantify the effects of the tax by estimating the tax elasticities of taxable net wealth and decompose the response into a reporting effect and an intertemporal substitution effect. Using a measure of cognitive skills, I also test the hypothesis that high-skilled individuals understand the tax system better and respond differently to tax reforms than do low-skilled individuals.

Sweden introduced a progressive tax on net wealth in 1947. From 1991, when Sweden implemented an extensive tax reform, until 2007, when the tax was repealed, it had two brackets, separated by a threshold.[6] The marginal tax rate was zero below the threshold, and 1.5 percent above it and the threshold was changed a number of times. The design of the tax schedule gives rise to two sources of variation that can be exploited empirically. First, the threshold for taxable wealth creates a kink in the budget set which, under general assumptions about individuals' behavior, makes them bunch at the kink point. Second, the change in the threshold over time enables a difference-in-difference strategy.

Swedish taxpayers annually received a prepopulated tax return based on the net wealth reported to the tax authority by third parties, such as banks and financial institutions. However, they were required to report any omitted assets and liabilities themselves. Third-party reported net wealth corrected for the taxpayers' self-reported adjustments thus constituted taxable net wealth. With access to both third-party reported net wealth and self-reported net wealth, I can study the nature of household responses.

By linking the wealth tax records to military enlistment data, which include a measure of cognitive ability, I study how responses differ across skill groups. Since military enlistment was mandatory for males in the sample cohorts, this close proxy for actual ability is not plagued by selection bias.

The institutional setting thus makes Sweden a promising testing ground for identifying the behavioral effects of the wealth tax. I first assess the data imposing as few assumptions as possible about preferences and behavior, and then gradually add more structure to unravel the mechanisms behind the observed outcomes.

To my knowledge, this paper is the first to assess the tax elasticity of taxable net wealth. Using the variation in the marginal tax rate across tax brackets, I start by estimating bunching at the kink point, i.e., the excess mass in the distribution at the tax threshold. When applying the methods proposed by Saez [2010], Chetty et al. [2011] to

[5] On the confusion of marginal and average taxes, see Liebman and Zeckhauser [2004], Feldman and Katuščàk [2006].

[6] On the reform in 1991, see Agell et al. [1996].

Swedish wealth data for 1999-2006, I find strong evidence of bunching at the kink point. The implied tax semi-elasticity of taxable net wealth lies in the absolute range $[0.12, 0.33]$, depending on the chosen bunching estimate. In other words, an increase in the tax by one percentage point leads to a reduction in taxable net wealth by 0.12-0.33 percent. I then exploit movements in the tax bracket over time in a difference-in-difference strategy. Using this approach, I find larger tax semi-elasticities of taxable wealth, ranging from 0.45 to 0.80 depending on the chosen time frame. The larger responses could be attributed to the larger salience of the tax on those once already close to the kink.

By comparing the distribution of third-party reported net wealth to taxable net wealth, I find that all bunching occurs along the self-reported margin. Kleven et al. [2011] argue that the high tax-compliance rates in modern tax systems are a result of third-party reported tax liabilities that are difficult to evade. Evasion is thus more likely to occur along the self-reported margin. I examine whether tax evasion is the source of bunching in two different ways. First, I study dynamic responses to a change in the threshold. If a change with a new threshold were to reflect real intertemporal responses, the stock of wealth would adjust only gradually and bunching at the new threshold would therefore manifest itself over time. However, I find that bunching occurs immediately upon moving the threshold and there is no systematic increase over time. Second, I empirically relate maximal pecuniary penalties for cheating to the difference between third-party reported net wealth and taxable net wealth and find that downward adjustments of taxable net wealth due to self-reporting are positively and significantly correlated with penalties. Third, cars were part of the tax base and were self-reported. Having gained access to register data on the car holdings (brand, model and vintage) of all Swedes during the years under study – data that the Tax Agency did not have access to – I cross-check the sum of self-reported assets against car holdings. I find that close to the exemption threshold, the fraction of car owners who report more assets than their cars are worth, is only 15-20 percent. These findings suggest that the self-reported margin is a good proxy for wealth tax evasion.

Since the data not only support the occurrence of bunching but also suggest that changes to the tax scheme affect evasion rather than savings decisions, I impose some structural assumptions in order to address the underlying mechanisms at work. I start with a simple static model, where agents choose how much money to shelter from the government, subject to a pecuniary convex cost of evasion consistent with the work of Allingham and Sandmo [1972], Slemrod [2001].[7] I use bunching at kink points and data on penalties to estimate two parameters of the evasion cost function. Translating these findings into tax semi-elasticities of taxable net wealth, I find that they are in the

[7]Slemrod and Yitzhaki [2002] provide an extensive review of the literature on tax evasion. Theoretical predictions of the effect of a tax increase on sheltering behavior are shown to be highly model-dependent. Yitzhaki [1974] shows that if the penalty is proportional to the tax rate, as is the case for modest amounts of cheating in many countries, the amount of sheltering is independent of the tax rate.

range of $[0.18, 0.42]$. I then extend the model to a dynamic setting where agents are motivated to save in order to smooth consumption. Parameterizing the model, I estimate uncompensated tax semi-elasticities of wealth in the range $[0.20, 0.40]$.

I finally investigate whether some individuals are more prone to evade than others. If cognitively able individuals are indeed more able to understand the tax system, do they respond more? To allow for heterogenous effects in cognitive skills, I use the framework in Chetty et al. [2007] to specify a bounded-rationality model where agents compare the utility from taking the tax rate into account, to the utility from ignoring it. If the utility difference exceeds some exogenous cost, agents do internalize the tax rate. Assuming that the cost depends negatively on cognitive ability, the model predicts more bunching among high-skilled households. This prediction is confirmed by the data. Exploring various explanations for this finding, I argue that agents with high ability understand the tax system better than agents with low ability. I find no evidence for alternative explanations, such as access to different technologies for evasion or differences in risk and time preferences.[8]

To ease the exposition, the different theoretical approaches are collected in Section 2.2. Section 2.3 presents the institutional setup and data. Section 2.4 provides the estimates of bunching at kink points along with an argument for how to empirically gauge evasion. Section 2.5 uses the theory in Section 2.2 and the estimates of bunching in Section 2.4 to compute the elasticities of the tax base to the wealth tax. Key results are discussed in Section 2.6 and Section 3.6 concludes.

2.2 Theory

This section outlines a theoretical framework for studying intertemporal substitution and reporting responses to a wealth tax. This is done in four steps. The framework in the first step refrains from making functional-form assumptions about individual behavior. The second step features a sheltering decision in a parametrized, but static framework. The third step explores a dynamic model, allowing also savings to respond to the tax. The fourth step presents a simple bounded-rationality framework which incorporates heterogeneity in skills.

2.2.1 Small Kink Analysis

Individuals trade off consumption today, c, against savings, s (consumption in the future). They are heterogenous with respect to preferences and savings technologies, which

[8] Although the data used for this investigation has unique coverage, the sample is small and the results should therefore be interpreted with caution. The literature seems to have settled on a positive relation between cognitive ability and patience and a negative relation between cognitive ability and risk aversion. See Frederick [2005], Benjamin et al. [2013].

are distributed according to some continuous and differentiable cumulative distribution function. With a linear tax on wealth (stock of savings), denoted by τ, and no uncertainty, individuals' wealth, Z, will be distributed according to a smooth density function $h(Z)$. This is true also when allowing for a tax-sheltering opportunity subject to a convex sheltering cost.

Introducing a kink in the budget set at wealth level z^*, by a higher marginal tax rate $\tau + \mathrm{d}\tau$ to the right of the kink, will trigger a savings response and, potentially, a reporting response for agents with a taxable net wealth level above z^*. Since wealth is a stock variable, the initial effect should represent reporting responses rather than intertemporal-substitution effects.[9] Agents that under the linear tax scheme chose taxable net wealth levels in some interval $[z^*, z^* + \mathrm{d}z]$, will bunch at the kink point. The number of households that bunch is thus $B = h(z^*)\,\mathrm{d}z^*$.[10] Individuals who chose higher wealth levels in the absence of the higher tax, reduce their taxable wealth to the point where their indifference curves are tangent to the budget line under the higher tax (which has the slope $1 - \tau - \mathrm{d}\tau$).

Under the assumption that the tax rate change at the kink is sufficiently small, so that the associated income effects are negligible, the response of taxable wealth can be interpreted by way of a compensated semi-elasticity. The semi-elasticity refers to the percentage change in wealth arising from a one-percentage point increase in the tax rate, $\varepsilon_{W,\tau} = -\frac{\mathrm{d}z}{\mathrm{d}\tau}\frac{1}{z}$. Combining this with the expression for bunching at the kink, I obtain

$$\frac{B}{h(z^*)\,z^*} = -\varepsilon^{K}_{W,\tau}\mathrm{d}\tau, \tag{2.1}$$

where W denotes taxable net wealth. As the response can differ over time, estimating bunching dynamically can potentially identify a vector of elasticities until the steady-state wealth distribution is obtained along with a long-run elasticity. With heterogeneous elasticities, bunching identifies the average elasticity.

2.2.2 Static Model

In practice, agents may respond to the wealth tax by saving less or by reporting less wealth. In this subsection I focus on reporting behavior.

Assume that agents aim at maximizing their budget given some pecuniary cost of tax sheltering. In principle, the framework can be reinterpreted as a maximization problem where utility is linear in consumption and agents pay a convex moral cost of tax evasion. However, in order to estimate the parameters, I regard the cost as monetary. Although this distinction may appear peripheral, it has important implications for welfare. If the

[9] This is true for unanticipated and immediately implemented tax changes.

[10] The expression is approximate, as it assumes a constant density over the net wealth interval of bunchers. The accuracy of the approximation increases as $\mathrm{d}z^*$ decreases.

cost is a sunk resource cost, Feldstein [1999] shows in a labor income tax setting that whether agents respond through behavior or through reporting is not important from the viewpoint of the social planner. According to Chetty [2009], this is no longer true when the costs take the form of transfers across agents (via tax sheltering penalties for example).

Formally, an agent faces the following maximization problem:

$$\max_{e\leq s}\left(1-\tau\right)\left(s-e\right)+e-C\left(e,s\right), \tag{2.2}$$

where s denotes exogenous savings, e tax sheltering activities and τ a linear tax. The cost of evasion, $C\left(e,s\right)$ is assumed to be increasing in evasion and decreasing in savings.[11] In the application to the Swedish wealth tax, s denotes true wealth and $s-e$ taxable (net) wealth. Following Slemrod [2001], I parametrize the cost function as follows:

$$C\left(e,s\right)=\left(\frac{e}{s}\right)^{\frac{1}{\gamma}}\frac{pe}{1+\frac{1}{\gamma}}, \tag{2.3}$$

where γ is the constant tax elasticity of evasion, and p is a parameter which can be interpreted as a linear penalty including fines, costs of going to court and other transfers.[12] The cost function is not specific to the wealth tax andÂ Slemrod [2001] employs it in a labor-income tax setting. The agent's solution to problem (2.2) is given by

$$e^{*}=\left(\frac{\tau}{p}\right)^{\gamma}s, \tag{2.4}$$

and the agent's taxable net wealth by

$$s-e^{*}=\left(1-\left(\frac{\tau}{p}\right)^{\gamma}\right)s. \tag{2.5}$$

The cost function implies that evaded amounts are proportional to true wealth. This is a prediction that I empirically assess in Section 2.5.2. Assuming that s is distributed according to some continuous and differentiable CDF $F\left(s\right)$, the choices of e^{*} under a linear tax imply that the distribution of taxable net wealth is also described by a continuous and differentiable function $H\left(s-e\right)$.

If a kink is introduced in the budget set at taxable net wealth z^{*}, so that $\tau=\tau_{0}$ below the kink and $\tau=\tau_{1}>\tau_{0}$ above the kink, agents who chose a taxable net wealth level in $[z^{*},z^{*}+\Delta z]$ will now choose to locate exactly at the kink point. The agent with the highest savings level s, who is bunching, had a taxable net wealth under the

[11] Allingham and Sandmo [1972] formulate this problem as a gamble. Mayshar [1991] shows that the gamble can be represented with a monetary cost of evasion, where the cost is the certainty-equivalent of the gamble that causes extra utility loss as the risk of audit is increased.

[12] Yitzhaki [1974] shows that if p is linear in τ, the tax rate has no effect on sheltering behavior.

linear tax rate given by $(s-e)^U = z^* \left(1-(\tau_0/p)^\gamma\right) / \left(1-(\tau_1/p)^\gamma\right)$. Hence, the number of households that bunch at the kink point is given by $H\left((s-e)^U\right) - H(z^*)$. Using the counterfactual density together with the fact that $\log(1+x) \approx x$ for small x, I obtain:

$$\frac{B}{h(z^*)\, z^*} \approx \log\left(\frac{1-\left(\frac{\tau_0}{p}\right)^\gamma}{1-\left(\frac{\tau_1}{p}\right)^\gamma}\right), \tag{2.6}$$

where B denotes the number of households that bunch at the kink point, and $h(z^*)$ is the density at the kink point under the linear tax scheme. Thus, $B/h(z^*)$ is the mass at the kink point in excess of the counterfactual density. If $\tau_0 = 0$, I approximate equation (2.6) as

$$\frac{B}{h(z^*)\, z^*} \approx \left(\frac{\tau_1}{p}\right)^\gamma. \tag{2.7}$$

Equation (2.7) allows me to identify the tax elasticity of evasion, γ, as a function of the observable parameters τ_1 and z^*, the estimable cost parameter p and bunching at the kink, $B/h(z^*)$. This formula has the intuitive feature that the constant fraction of wealth that is sheltered from the government (the right-hand side) and estimated excess mass are proportional to each other. As the tax rate increases, the fraction of evaded savings goes up and more households bunch at the kink.

The relevant semi-elasticities take the following form:

$$\begin{aligned}
\varepsilon^S_{e,\tau} &= \frac{\gamma}{\tau} \\
\varepsilon_{W,\tau} &= -\frac{\gamma}{\tau}\frac{\left(\frac{\tau}{p}\right)^\gamma}{1-\left(\frac{\tau}{p}\right)^\gamma},
\end{aligned}$$

where W again represents taxable net wealth. The tax elasticity of tax evasion is decreasing in the tax rate and for $p > \tau$, the tax elasticity of taxable net wealth is increasing in the tax rate.

2.2.3 Dynamic Model

Section 2.2.2 assumes that all responses to the tax rate occur through tax sheltering. This section considers a dynamic model where agents also respond to tax changes by way of savings. The model abstracts from labor supply responses to the wealth tax.

In the dynamic model, agents trade off consumption in two periods. Agents pay a tax on their accumulated assets (which in this framework are equal to savings). However, they can shelter money from the government through the same technology as in Section 2.2.2.

Since a change in the wealth tax may trigger a slow dynamic response of wealth, rather than a one-off adjustment, Appendix A.1 lays out the infinite-horizon version of this dynamic model. This extension allows for a transition path of wealth to the new steady state.

The agent faces the following maximization problem:

$$\max_{s,e\leq s} U\left(c_1, c_2\right) = \max_{s,e\leq s} \frac{c_1^{1-\frac{1}{\sigma}} - 1}{1-\frac{1}{\sigma}} + \beta \frac{c_2^{1-\frac{1}{\sigma}} - 1}{1-\frac{1}{\sigma}} \tag{2.8}$$

subject to

$$c_1 = y - s$$
$$c_2 = (1-\tau)(s-e) + e - \left(\frac{e}{s}\right)^{\frac{1}{\gamma}} \frac{pe}{1+\frac{1}{\gamma}},$$

where c_t is consumption in period t, β the discount factor, σ the elasticity of intertemporal substitution and y is heterogenous income distributed according to a continuous and differentiable CDF $G(y)$.[13] The cost of evasion is assumed to take the same functional form as in Section 2.2.2, implying that the first-order condition that governs the tax sheltering response is given by equation (2.4), restated here for convenience: $e^* = (\tau/p)^\gamma s$. Substituting this into the Euler equation which determines the savings response to the tax, I obtain:

$$c_1^{-\frac{1}{\sigma}} = \beta\left(1-\tau\left(1-\left(\frac{\tau}{p}\right)^\gamma \frac{1}{1+\gamma}\right)\right)^{1-\frac{1}{\sigma}} c_2^{-\frac{1}{\sigma}}. \tag{2.9}$$

An increase in the tax rate has three effects. First, the fraction of savings evaded from tax goes up. The magnitude of this response is given by the structural parameter γ and the penalty cost, p. Second, the return to saving is negatively affected by a tax increase and parameter σ determines the relative importance of the income and substitution effects associated with a tax increase. With $\sigma < 1$, the income effect dominates the substitution effect.[14] An increase in the tax rate actually raises savings. When $\sigma > 1$, the substitution effect dominates the income effect and an increase in the tax rate lowers savings. Third, the cost function possesses the feature that higher savings lower the marginal cost of evasion. Slemrod [2001] refers to this as the *avoidance-facilitating* effect. The distortionary effect of an increased tax rate on savings is thus attenuated by agents evading a fraction of their savings.

In the general version of this economy, the Euler equation determines the balanced growth path. From a growth-enhancing policy perspective, tax evasion thus weakens the distortionary effects of the tax on long-run growth.

[13]For simplicity, I assume that the gross interest rate is zero, but this can easily be relaxed. In the estimation procedure, the choice of the interest rate does not have a large effect on estimated entities.

[14]However, both the uncompensated and income-compensated effects on consumption by an increase in the tax rate are negative.

The agent chooses s^* according to

$$s^* = f(\tau)\, y, \tag{2.10}$$

where

$$f(\tau) = \frac{\beta^\sigma \left(1 - \tau \left(1 - \left(\frac{\tau}{p}\right)^\gamma \frac{1}{1+\gamma}\right)\right)^{\sigma-1}}{1 + \beta^\sigma \left(1 - \tau \left(1 - \left(\frac{\tau}{p}\right)^\gamma \frac{1}{1+\gamma}\right)\right)^{\sigma-1}}, \tag{2.11}$$

and taxable net wealth becomes

$$s^* - e^* = f(\tau) \left(1 - \left(\frac{\tau}{p}\right)^\gamma\right) y. \tag{2.12}$$

Taxable net wealth is proportional to exogenous income y. Therefore, it is again distributed according to some continuous and differentiable CDF denoted by $K(s - e)$ under the linear tax rate. Increasing the marginal tax rate above threshold z^*, such that $\tau = \tau_0$ for taxable net wealth levels below the kink and $\tau = \tau_1 > \tau_0$ above z^*, leads agents close to the kink to adjust their taxable net wealth levels downwards and bunch at the threshold. This could be done either by savings (real response) or by evasion (reporting response), or a combination of the two.

Identifying the interval of bunchers as in the static case, I can relate the bunching at the kink point to the parameters of the model:

$$\frac{B}{k(z^*)\, z^*} \approx \left(\frac{f(\tau_0)}{f(\tau_1)} \frac{\left(1 - \left(\frac{\tau_0}{p}\right)^\gamma\right)}{\left(1 - \left(\frac{\tau_1}{p}\right)^\gamma\right)} - 1\right). \tag{2.13}$$

In (2.13), $k(z^*)$ denotes the density of the distribution of taxable net wealth at the kink point with a linear tax rate. Equation (2.13) is a generalized version of equation (2.6). The left-hand side is the excess mass at the kink point z^*. The right-hand side is the interval of taxable net wealth values under the linear tax where wealth holders bunch at the kink point when under the progressive tax.

If $\tau_0 = 0$, the following approximation holds:

$$\frac{B}{z^* k(z^*)} \approx \log\left(\frac{\beta^\sigma}{1+\beta^\sigma}\right) - \log\left(\frac{\beta^\sigma \left(1 - \tau_1 \left(1 - \left(\frac{\tau_1}{p}\right)^\gamma \frac{1}{1+\gamma}\right)\right)^{\sigma-1}}{1 + \beta^\sigma \left(1 - \tau_1 \left(1 - \left(\frac{\tau_1}{p}\right)^\gamma \frac{1}{1+\gamma}\right)\right)^{\sigma-1}}\right) + \left(\frac{\tau_1}{p}\right)^\gamma. \tag{2.14}$$

The log difference on the right-hand side captures the discrepancy in savings rates between the left and the right side of the threshold. A large positive discrepancy implies more bunching at the kink point. The third term on the right-hand side captures the fraction of evaded savings, where higher evasion adds to bunching.[15] Equation (2.14) illustrates intuitively how bunching can arise through adjusted savings as well as evasion. In the static model, a higher penalty rate p always implies lower overall bunching. Here, the impact is less clear. Higher penalty rates still lower evasion but raise the difference in savings rates between the two sides of the kink. If $\sigma < 1$, the log-difference is actually negative, and the fraction evaded increases to reconcile the estimated amount of bunching.

According to equation (2.14), bunching depends on: (i) observable tax parameters (the tax rate and the kink point); (ii) preference parameters determining the real response (the discount factor and the elasticity of intertemporal substitution); and (iii) evasion cost parameters (the convexity of the cost function and the penalty).

The uncompensated tax semi-elasticities of taxable net wealth and evasion now take the following form:

$$\varepsilon^D_{W,\tau} = -\left((1-\lambda)(\sigma-1)\left(1-\tau+\tau\frac{1}{1+\gamma}\left(\frac{\tau}{p}\right)^{\gamma}\right)^{-1}(1-f(\tau)) + \lambda\frac{\gamma}{\tau\left(1-\left(\frac{\tau}{p}\right)^{\gamma}\right)}\right) \tag{2.15}$$

$$\varepsilon^D_{e,\tau} = -(1-\lambda)(\sigma-1)\left(1-\tau+\tau\frac{1}{1+\gamma}\left(\frac{\tau}{p}\right)^{\gamma}\right)^{-1}(1-f(\tau)) + \frac{\gamma}{\tau} \tag{2.16}$$

where $\lambda = (\tau/p)^{\gamma}$ denotes the fraction evaded. These expressions are sums of the real and the evasion response. In fact, the first part on the right-hand side of both equations denotes the tax semi-elasticity of actual wealth, or savings. If the tax rate goes up (or the penalty rate goes down), agents evade more (λ goes up) and the elasticity of taxable net wealth is relatively more affected by evasion than savings. If $\sigma < 1$, the income effect is stronger than the substitution effect and agents save a larger fraction of income upon the tax change. This effect arises as agents are only aiming at consumption smoothing. In Piketty and Saez (2012), agents' preferences are defined over wealth, bequests and consumption and therefore real savings responses to tax changes reflect many wealth accumulation motives.

2.2.4 Bounded Rationality

The framework considered so far may be appropriate for studying how the wealth tax affects savings, evasion, and aggregate wealth inequality. However, it does not allow for innate heterogeneity in the ability to fathom the tax system across groups.

To allow for heterogenous responses to the tax rate, based on the ability to perceive the tax, I follow the approach in Chetty et al. [2007]. I formulate a model where agents rationally ignore the tax if the utility gain from including the tax in the optimization

[15] Since the tax rate is zero to the left of the threshold, there is no evasion behavior among households to the left of the kink. If the tax rate to the left of the kink was positive, the amount of bunching would be increasing in the difference between evasion rates on the two sides of the threshold.

decision is lower than some exogenously given cost.[16] This is in line with the bounded-rationality literature, which assumes that it is costly to acquire or process information and that agents therefore overlook information that would lead to optimal choices in the absence of these costs.[17]

Since an inattentive agent ignores the tax when optimizing, he evades no money in the static model. The gain – measured in monetary units – from taking the tax rate into account is thus given by

$$G(\tau) = s\left(1 - \tau + \tau\frac{1}{1+\gamma}\left(\frac{\tau}{p}\right)^{\gamma}\right) - s(1-\tau) = \tau\frac{1}{1+\gamma}\left(\frac{\tau}{p}\right)^{\gamma} s. \tag{2.17}$$

Agents optimize taking the tax rate into account if $G(\tau) > c$. In the following, I assume that c is heterogeneously distributed in the population and that c is negatively correlated with cognitive ability. If the gains from knowing the tax rate are small, aggregate inattention is higher. When the tax rate increases, the amount of inattention decreases.

The wealth tax considered in this paper is low which implies that the gains are small. For example, if the fraction evaded is 10 percent, $\gamma = 0.5$, $p = 0.86$ and $\tau = 0.015$ for wealth above SEK 1 million, an agent with SEK $1,5$ million loses SEK 665, about USD 100, by ignoring the tax. However, if less able individuals are also myopic, ignoring the tax every year, small losses may grow to large differences over time. With an annual interest of 5 percent, the future value of the augmented losses, computed as an annuity, amounts to SEK $80,211$ over a 40 year period. This is roughly equal to a third of median annual earnings in Sweden. Seemingly low gains from being attentive may thus produce amplified inequalities across skill groups. A natural policy intervention for mitigating the inequality effects of the tax would be to increase the penalty. Higher penalties decrease the redistributional distortions as evasion goes down and hence, the utility loss from ignoring the tax is lowered.

In the dynamic framework, an inattentive agent chooses how much to save while ignoring the tax rate. The perceived budget constraints are thus:

$$c_1 = y - s$$
$$c_2 = s.$$

However, the inattentive optimal-savings choice will entail an infeasible consumption plan since the agents are nonetheless tax liable. Consumption has to adjust in view of

[16] This optimization problem has the perplex feature that the decision of whether to remain inattentive or not requires not only knowing the utility from full optimization but also the utility from inattentive optimization. Conlisk [1996] refers to this phenomenon as the *regression problem*. In practice, individuals will presumably know the losses from ignoring the tax, at least approximately.

[17] In this vein, Reis [2006] shows how agents optimally choose the interval length between updates of their financial portfolio. In a series of papers, Sims [2003, 2006] develops the idea that agents are instead limited in the amount of information they can process.

the optimization mistake. A natural assumption is that households choose first-period consumption according to their first-order condition and period-two consumption equals the residual after the tax has been paid.[18] Then actual consumption in period 2 is:

$$c_2 = s\left(1-\tau\right). \tag{2.18}$$

An inattentive agent's first-order condition reads $c_1^{-\frac{1}{\sigma}} = \beta c_2^{-\frac{1}{\sigma}}$ and hence, $\hat{c}_1 = \left(1-\beta^{\sigma}/\left(1+\beta^{\sigma}\right)\right)y$ and $\hat{c}_2 = \left(\beta^{\sigma}/\left(1+\beta^{\sigma}\right)\right)\left(1-\tau\right)y$. Agents rationally ignore the tax if the utility from this consumption plan is close to the utility obtained from optimizing with the tax rate. This is true if

$$G\left(\tau\right) = \frac{c_1^{*1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} + \beta\frac{c_2^{*1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \frac{\hat{c}_1^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \beta\frac{\hat{c}_2^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} \tag{2.19}$$

is lower than some cognitive cost c. The above expression can be restated as

$$G\left(\tau\right) = \Big(\left(1-f\left(\tau\right)\right)^{1-\frac{1}{\sigma}} - \left(1-f\left(0\right)\right)^{1-\frac{1}{\sigma}}\Big)\frac{y^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} + \tag{2.20}$$

$$\beta\left(\left(\left(1-\tau+\tau\left(\frac{\tau}{p}\right)^{\gamma}\frac{1}{1+\gamma}\right)f\left(\tau\right)\right)^{1-\frac{1}{\sigma}} - \left(\left(1-\tau\right)f\left(0\right)\right)^{1-\frac{1}{\sigma}}\right)\frac{y^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}, \tag{2.21}$$

where I have used the fact that the savings rule of an inattentive agent is the same as that of a rational agent when the tax rate is zero.[19]

As before, a discontinuity in the marginal tax rate heterogeneously triggers bunching depending on the utility gain from bunching at the kink as opposed to ignoring the tax rate. Let us assume that the cost of internalizing the tax rate is heterogeneously distributed within the groups of low- and high-skilled, *LS* and *HS* respectively, and assume that $\mathrm{E}_{LS}\left[c\right] > \mathrm{E}_{HS}\left[c\right]$. Then, the fraction of agents who optimize under the tax rate close to the kink is larger among high-skilled than low-skilled. As the tax rate increases, the utility difference increases and more individuals rationally cease to ignore the tax. A testable prediction of this framework is thus heterogenous bunching across skill groups. Importantly, in this model, heterogeneity in bunching does not reflect preferential differences or discrepancy in the evasion technology. With high tax rates, the cost of ignoring the tax is large and aggregate behavior is similar to that of a rational agent.

---

[18] The assumption of when the tax is payed does not influence the results.

[19] If agents with low ability, i.e. high costs, are also more myopic than agents with high ability, with $\beta_L < \beta_H$, low-skilled agents may require an even larger utility gain than high-skilled in order to optimize with the tax rate. This result is an effect of the assumed budget rule with period two consumption adjusting to the optimization mistake and it is obtained when the savings rate of an attentive agent is higher than that of an inattentive agent.

2.3 Institutional Background and Data

In a comprehensive tax reform dated 1947, the Swedish parliament supplemented the existing inheritance and gift tax with a separate progressive annual wealth tax and an estate tax.[20] The adoption of the reform was preceded by intense debate. In the opening speech of a meeting with the Swedish Economic Association in 1947, Eli Heckscher criticized the higher taxes on wealth not only for their distortive effects on private savings but also for the risk of increased tax avoidance.[21] The difficulties associated with the legal implementation of an annual wealth tax are also recognized in the recent work by Adam et al. [2011]. The proper valuation of assets is difficult and impractical when defining a broad tax base, which includes assets and liabilities that are not traded or priced on a regular basis. In addition, some taxable assets and liabilities are self-reported which opens up the possibility to evade. In line with these concerns, the results from a survey aimed at eliciting perceptions of tax cheating in Sweden in 2006 indicate that individuals perceiving tax evasion as common believe the wealth tax to be the tax most likely to be subject to evasion (Hammar et al., 2006).

2.3.1 Institutional Setup

While the Swedish estate tax was repealed in 1953, the annual wealth tax was in place until 2007.[22] Taxable assets consisted primarily of shares in publicly traded companies, bonds, bank-account holdings, real estate, cars, boats and capital insurance. The wealth tax base was defined as the total value of these taxable assets net of liabilities, like real-estate mortgages and consumption loans, i.e., net worth. In 1991, a system with three marginal tax brackets was converted into a two-bracket system, with a zero marginal tax rate for net wealth below SEK $900,000$ and a marginal tax rate of 1.5 percent for net wealth above this threshold.[23] The tax was filed jointly for couples with children below 18 years of age. As of 2000, the threshold was different for singles and for couples who were required to file jointly. During the period 1999-2006, the threshold was increased several times, as displayed in Figure 2.1. Approximately 8 percent of the population paid the wealth tax in 1999.

The filing of the wealth tax occurred in the spring of year $t+1$ for wealth holdings as of December 31 in year t. The Swedish Tax Agency (Skatteverket) sent out prepopulated tax forms which were based on third-party reports from banks, investment funds, brokers

[20] Formally, wealth was taxed from 1910, but before 1947 a fraction of the wealth was added to income and taxed through the income tax system.

[21] See Nationalekonomiska Föreningen [1948]. Heckscher criticized the proposed wealth taxes for cultivating a tax-avoidance norm, rather than inducing individuals to behave as if the price on transfers or savings was raised. Ohlsson [2011] gives a detailed summary of the events surrounding the reform.

[22] The tax on inheritance and *inter vivos* gifts was abolished in 2005.

[23] $\$ 1 \approx 6.5$ SEK, implying a threshold equivalent to USD $130,000$.

and other financial institutions. Figure 2.2 displays a prepopulated tax return. As the tax base included non-third party reported assets such as cars, boats, securities and liabilities held abroad, and debt within families, tax payers were required to self-report such holdings. The taxable net wealth equalled the sum of third-party reported net wealth and self-reported net wealth. The form in Figure 2.3, which was appended to the return sheet, explains how to calculate the tax liability. Upon receiving the tax form, the taxpayer was allowed to make adjustments and submit a final return by May 1 in year $t+1$. In the analysis, I investigate the effects of tax reforms on both third-party reported net wealth, and taxable net wealth.

The main purpose of the wealth tax was redistribution in recognition of the potentially distortive effects of the tax on investments. To avoid the displacement of firms abroad, various tax exemptions, which narrowed the tax base, were installed over time. Thus, stocks not listed or traded on organized exchanges were not subject to wealth taxes. Moreover, as the tax was considered to be a deterrent to stock enlistment of companies in publicly traded markets, company ownership above 25 percent was tax exempt. Although the taxable amount should, in principle, reflect market value, some stocks were taxed only at 80 percent of their market value, while other stocks remained completely tax exempt. Incentives for the placement of wealth in third-party reported assets that legally avoided the wealth tax were amplified by retirement savings being exempt from the tax.

While individuals may have responded to tax changes through strategic portfolio choices, the wealth tax was not associated with deduction opportunities. However, individuals with low income and high net wealth were tax exempt to some degree. Both a general law stipulating that total tax liability should never exceed 60 percent of income and a specific law entailing a reduction in the wealth tax for households with low income and high real-estate value were in place during the latter years of the study (2005-2006).[24] In case of excess tax liability, the wealth tax was lessened, but not by more than 50 percent of the pre-limit liabilities. The wealth tax could not be exempted in its entirety.

In addition to the wealth taxes, real estate was taxed annually at 1 percent of the (assessed) taxable value. Movements in the wealth tax bracket between years t and $t+1$ were, in practice, indexed against changes in the taxable value of real estate.[25] When the tax value for real estate was revalued, the wealth tax was also reformed to avoid implausible increases in tax liability. By the end of the 1990s, the government renewed its procedure for computing the taxable value of real estate, implying substantial increases in tax liabilities. These were accommodated by movements in the wealth tax bracket several times at the beginning of the 2000s. The indexation of the wealth tax bracket to the real estate tax suggests that the increase in the threshold was not driven by a

[24]The general tax reduction law was in place during the whole sample period.

[25]See the Swedish Government Official Reports [2004]. Henrekson and Jakobsson [2001] provide a survey of the impacts of wealth taxes on business ownerships in Sweden and recapitulate the political agenda behind various reforms.

powerful lobby of wealthy households.

2.3.2 Data and Sample Restrictions

The data I use in this paper come from the following administrative registers provided by Statistics Sweden: (i) The Income and Tax Register (Inkomst- och Taxeringsregistret); (ii) The Integrated Database for Labour Market Research (LISA); and (iii) Military Enlistment Data from the National Service Administration (Pliktverket).

From The Income and Tax Register, I retrieve all third-party reported asset items, taxable assets, taxable liabilities, liable wealth taxes as well as cheating penalties for each Swedish taxpayer above 15 years of age over the years 1999-2006. Using the third-party reports and tax records from the Swedish Tax Agency, I calculate two distinct measures of net wealth: third-party reported net wealth and taxable net wealth. The latter measure coincides with the former if no self-reported adjustments were made by the taxpayer. In the tax records, taxable net wealth is often not reported for households that do not pay wealth taxes. For the tax agency, keeping track of taxable net wealth for households that were not tax liable was of minor interest. In case the third-party reported net wealth was below the kink and the household did not pay any wealth tax, I assume that taxable net wealth was equal to third-party reported net wealth.[26]

For households with a prepopulated third-party reported net wealth above the kink that end up below the kink, taxable net wealth is also not reported in a number of cases. I deal with this in two ways. In the first approach, I do not include these households in the bunching estimation procedure at all. In the second approach, I assume that the distribution of reported taxable net wealth for those below the kink is representative for those where taxable net wealth is missing and extrapolate values to those with no reported taxable net wealth. I therefore refer to the second approach as the extrapolation method.[27] While both approaches imply a lower density to the right of the kink, the bunching estimate will be lower in the first approach.

Demographic information on individual characteristics such as age, education, occupation, wage earnings and family status, is collected from the LISA database, which includes both spouses' social-security numbers, enabling me to link couples filing the

[26]If the (imaginary) self-reported wealth for households below the kink behaved in the same way as for households above the kink, this procedure attenuates the bunching estimates. To the right of the kink, I find that households, in general, adjust their wealth downwards. If this were true to the left of the kink, such self-reports shift the distribution downwards. However, bunching at the kink reflects households coming from the right of the kink and does not shift. This implies that estimates are lower without knowing self-reports.

[27]This method may yield biased results if the sample with missing taxable net wealth values is endogenously selected. Running a regression on the sample of households with third-party reported net wealth above the kink who pay no wealth taxes, I find that an indicator for having a value is not significantly correlated with cognitive skills or wage earnings.

wealth tax jointly.[28] As in The Income and Tax Register, this database comprises individuals above 15 years of age. I am able to match 99.9 percent of the tax payers to the demographic database, yielding a matched dataset consisting of 58,015,897 observations over the period 1999-2006.

Since the wealth tax was filed jointly by households, with the sum of all household members' net wealth constituting the taxable net wealth, lack of data on individuals below 16 years of age implies that I am not able to assess taxable net wealth figures for households with children. Fortunately, the demographic dataset contains information about household status, including information about children below 18 years of age. I thus confine the sample to consist of single households and couples without children. This results in 20,773,835 observations of single households and 6,961,055 entries of couples filing the tax jointly. In sum, this represents 60 percent of the total number of observations. Out of these, 6 percent, or 1,668,465 observations include positive wealth tax payments. Self-employed individuals who used their assets in business activities were tax exempt. To avoid the results being driven by the self-employed, I restrict the sample by excluding individuals with assets in industrial property, agricultural property and rental property. Based on the aforementioned databases, couples in which one of the spouses possesses such assets are also removed. This implies dropping 2,011,649 observations.[29]

To obtain data on cognitive skills, I exploit psychological tests from the military draft. Before enlisting in the military, all men in Sweden were drafted and had to go through two days of various testing. The test procedure was in principle mandatory until 2010, but since the number of candidates fell in the 2000s, I confine the sample to include only cohorts of men born from 1951 until 1979. Approximately 90 percent of all men in my data who were born in the defined time period enlisted with the military.[30]

The enlistment usually took place the year in which the candidates turned 18. Apart from physical tests and a semi-structured interview with a psychologist that evaluated noncognitive skills, a cognitive skills test was taken.[31] This test consisted of four subtests with 40 questions each that evaluated logical, verbal and spatial capabilities and

[28]In addition to married couples, the database includes information on common-law spouses. The final dataset comprises both types of relationships.

[29]To investigate whether this sample confinement spuriously generates bunching, I estimate bunching with the self-employed in the sample. The results are stronger when including self-employed with a bunching estimate $b = 1.01$, compared with 0.53 for the sample retained.

[30]There is some variation in the availability of enlistment data over the years. For example, the enlistment data only includes about 40 percent of the men enlisting in the year 1985. There is no official reason for the varying data availability according to officials at the National Service Administration. In general, compliance is large and, according to E. Grönqvist et al. [2010], only the physically and mentally handicapped were exempted from enlistment. Consequences of refusal included fines and, even, imprisonment.

[31]A comprehensive overview of the test procedure is found in Lindqvist and Vestman [2011]. Carlstedt [2000] provides a review of the cognitive skills test, arguing that the test provides an accurate measure of general intelligence.

a test that evaluated the conscript's technical comprehension. Each subtest yielded a score between 4-36 points. As the test was subject to minor revisions in 1980 and 1994, I create a normalized measure of cognitive skills by ranking the sums of the subtest within enlistment years by percentile, and applying the inverse of the standard normal distribution to obtain a standard normal measure of cognitive skills. $4,103,044$ household-year observations of the remaining sample of tax payers were matched with enlistment data. No women appear in the resulting dataset and the unit of observation is either a single man or an enlisted man in a couple.

Table 2.1 summarizes the evolution of the number of taxpayers and government revenue over time. In 2001 and 2002, the threshold for paying wealth taxes was raised both for singles and couples filing jointly, thus reducing the revenue from the wealth tax, along with the number of taxpayers. Table 2.2 presents summary statistics for different subsamples. Consistent with a standard savings profile over the life-cycle, the table suggests that wealth tax payers are considerably older than the rest of the population. Since the oldest individuals in the sample with enlistment data were born in 1951, these households are, on average, younger and hold less assets than other households.

2.4 Empirical Analysis of Bunching

This section uses the data described in Section 2.3 to estimate the parameters needed for computing the tax elasticities implied by the framework in Section 2.2. These implied tax elasticities are presented in Section 2.5. I start by estimating bunching, i.e. the excess mass in the distribution of taxable net wealth at the kink point. In order to obtain an estimate, the counterfactual density, i.e. the mass at the kink point if the tax rate were zero, must be gauged. I do this in two ways. The first approach, described in Section 2.4.1, makes my results directly comparable to the literature. Here, I use the parametric method for estimating bunching employed in Chetty et al. [2011]. In Section 2.4.2, I instead take a non-parametric approach by estimating bunching from the two-dimensional data on third-party reported and taxable net wealth, thus capturing bunching occurring along the self-reported margin. In Section 2.4.3, I make the case that the self-reported margin may serve as a proxy for evasion. Finally, Section 2.4.4 calibrates the parameter p in the cost function (2.3), which is needed for identification of the parametrized models.

2.4.1 Parametric Estimation

Figure 2.4 plots the distribution of third-party reported net wealth for all Swedish households without children (below 18 years of age) for the years 1999-2006. In constructing the figure, I add up third-party reported assets, item by item, and subtract liabilities. The resulting number is cross-checked against the third-party reported measure of net wealth provided by Statistics Sweden. Figure 2.5 plots the corresponding distribution of taxable net wealth, obtained by adding the self-reported net wealth to the third-party reported net wealth. To obtain centralized measures of wealth while taking the changes in the threshold illustrated in Figure 2.1 into account, I calculate the difference between either wealth variable and the amount of wealth needed to reach the tax bracket with a positive marginal tax rate. Households are then grouped into SEK $5,000$ bins. Bin counts are plotted around the kink point, which is demarcated by the vertical line at zero.

Figure 2.4 reveals no bunching in the distribution of third-party reported net wealth around the kink point, but there is a marked spike in the otherwise smooth distribution of taxable net wealth at the kink point in Figure 2.5. To estimate bunching at the kink – a value needed to identify the parameters of interest – an estimate of the counterfactual density at the kink is required. However, in contrast to the theoretical predictions, the spike in the empirical distribution is diffused around the kink point. This presumably reflects households' inability to perfectly monitor savings to locate at the kink or, if households are bunching through evasion, their willingness to shroud tax evasion. When estimating bunching, I account for such noise. Specifically, to gauge the excess mass at the kink, I first estimate the distribution of taxable net wealth in the absence of taxation. Following Chetty et al. [2011], I fit a polynomial to the empirical distribution as follows:

$$N_j = \beta_0 + \beta_1 Z + \beta_2 Z^2 + \cdots + \beta_n Z^n + \sum_{i=-R}^{Q} \phi_i \mathbb{I}\left[Z_j = i\right] + \varepsilon_j, \tag{2.22}$$

where N_j denotes the number of households in bin j, Z_j is wealth relative to the kink in SEK $5,000$ intervals and n is the order of the polynomial. The sum of indicator variables on the right-hand side reflects the exclusion of observations close to the kink point, as these reflect bunching. R and Q define the lower and upper bounds of this interval, respectively. The counterfactual density is given by the predicted values, $\hat{N}_j$, excluding the contribution of the ϕ_i dummies around the kink. An estimate of the number of households bunching is thus $\hat{B} = \sum_{j=-R}^{0} N_j - \hat{N}_j$, i.e. the number of households in excess of the counterfactual density close to the kink point. The method overestimates the true value of $\hat{B}$, however, since the counterfactual density does not satisfy the integration constraint. To correct for this, I follow Chetty et al. [2011] and estimate a counterfactual density that shifts the counterfactual distribution to the right of the kink until the integration constraint is satisfied.[32] The relevant measure for the elasticity estimation is the estimated number of bunching households relative to the counterfactual density close to

[32]The counterfactual density is in this case given by: $N_j\left(1 + \mathbb{I}\left[j > Q\right] \frac{\hat{B}}{\sum_{j=Q+1}^{\infty} N_j}\right) = \beta_0 + \beta_1 Z + \beta_2 Z^2 + \cdots + \beta_n Z^n + \sum_{i=-R}^{Q} \phi_i \mathbb{I}\left[Z_j = i\right] + \varepsilon_j$.

the kink point, measured as

$$\hat{b} = \frac{\hat{B}}{\frac{\sum_{j=-R}^{Q} \hat{N}_j}{Q+R-1}}. \tag{2.23}$$

Figure 2.6 plots the distribution of taxable net wealth together with the counterfactual distribution, estimated as a seven-degree polynomial. The window of bunching is defined as SEK 40,000 below the kink, a value of -8 in the figure. Marginal changes in the window of bunching and the order of the polynomial do not affect the estimated excess mass to any considerable extent. I estimate bunching at $\hat{b} = 0.53$, meaning that there is 53 percent more mass relative to the counterfactual distribution within SEK 5,000 of the kink.[33] Figure 2.7 presents the bunching estimates using the extrapolation method, i.e including households with missing taxable net wealth values, which produces an estimate of $\hat{b} = 1.25$.

The standard error for $\hat{b}$ is estimated by a parametric bootstrap procedure, as in Chetty et al. [2011]. It reflects misspecification of the polynomial rather than sampling errors, as the estimate is constructed using the population distribution. The estimated standard error is 0.09, yielding a t-statistic on $\hat{b}$ of 5.89 and the null hypothesis of no bunching at the kink point is rejected with a p-value of 1.97×10^{-9}.

Figures 2.6 and 2.7 show a drop in the density to the right of the kink point, which is not explained by theory. Holes in the distribution arise in tax schemes with discontinuous jumps in average, not marginal, tax rates. One plausible explanation for this is that the time it takes to file the tax return and compute the value of non-third-party reported assets and liabilities constitutes a fixed cost. The existence of a fixed cost is analogous to a discontinuity in the average tax rate, i.e. a notch, predicting that there should be a hole in the distribution to the right of the kink.[34]

Figure 2.8 displays bunching in the distribution of third-party reported net wealth. Here, the estimated excess mass is only 0.10, but still statistically significant.

The identification of parameters in the theoretical framework relies on the assumption of the distribution of taxable net wealth being smooth in the absence of the wealth tax. This assumption can be relaxed by investigating movements in the tax bracket over time. The threshold defining the two tax brackets was different for singles and couples filing jointly. In addition, the threshold was changed several times and by separate amounts. Figures 2.9 and 2.10 display bunching around the kink in each year during the period

[33] One possibility that I am not able to address in the data, is that there is spurious bunching. Consider a household that is located to the right of the kink in third-party reported net wealth, but after truthful self-reporting would end up far to the left of the kink. If this household recognizes that the tax liabilities are equal just to the left of the kink and at the true point further to the left, it may choose to locate just below the threshold even though it is a tax complier. In ongoing work, I investigate this issue by comparing detailed administrative records on cars to the self-reported measure.

[34] Kleven and Waseem [Forthcoming] analyze bunching at thresholds in the income tax scheme when the average tax rate increases discontinually. This creates an even stronger incentive to bunch than do discontinuities in the marginal tax rate.

1999-2006 for singles and couples filing jointly, respectively. The estimated excess mass follows the kink closely over time. Figure 2.11 further investigates whether bunching follows the tax rate over time, or if alternative explanations account for the evolution of bunching. I compare the distribution of taxable net wealth for singles in 2001 to that of 2006, a time period marked by an increase in the threshold by SEK $500,000$. The excess mass in 2001 is located at the tax threshold and the figure presents three hypothetical scenarios for the location of the kink in 2006. The first placebo-kink denotes the threshold value that would be obtained had it followed inflation, the second location indicates the same value had it followed the riskfree interest rate and the final kink denotes the value had it followed the Stockholm Stock Exchange Index. The figure confirms that the kink does move to the 2006 tax threshold.

Turning to heterogeneity in bunching, Figures 2.9 and 2.10, in general, reveal no differences between singles and couples without children that file the tax jointly. The largest estimates of bunching for each subgroup are $\hat{b} = 1.4$ for singles and $\hat{b} = 1.54$ for couples.[35] I next turn to differences in excess mass across skill groups. To maintain statistical power, I employ the extrapolation method when performing this analysis. Figures 2.12 and 2.13 present taxable net wealth distributions by two skill groups defined by having obtained positive or negative z-scores on the cognitive military draft test.

The differences in bunching across cognitive skill groups are more pronounced over the years 2002-2006, a period during which the kink was held constant for single households, as indicated in Figures 2.14 and 2.15. The estimated bunching is 1.1 for the high-skilled and 0.42 for the low-skilled, i.e. almost three times as large for the high-skilled group. Comparing the top 25th percentile of the skill distribution to the bottom 25 percent in Figure 2.15, the coefficient of bunching is almost four times larger for high-skilled households than for low-skilled households. The difference in bunching across skill groups is a key finding, which translates into larger tax responses among high-skilled than low-skilled. Such differences will ultimately affect the wealth inequality across skill groups.

2.4.2 Nonparametric Estimation

The parametric estimates of bunching in Section 2.4.1 rely on the estimated counterfactual density accurately reflecting the distribution of taxable net wealth. Since a higher marginal tax rate does not only affect households close to the kink, but all households to the right of the kink point, it is not obvious that the estimated counterfactual density closely matches the true one. A biased estimate of the counterfactual distribution would not only bias the counterfactual density close to the kink, but also the estimated number of households bunching.

My second approach instead estimates bunching nonparametrically, by exploiting the

[35] Figure 2.9 shows a deviating pattern for 1999. The reasons for this deviation are difficult to address.

paired observations of third-party reported net wealth and taxable net wealth for each household. The idea is to gauge bunching by computing the number of households located above the kink in third-party reported net wealth but below the kink in taxable (self-reported) net wealth. Specifically, I use the following estimator of bunching:

$$\hat{B} = \sum_{i}^{N} \mathbb{I}\left[W^* - \delta < W_i < W^* \text{ and } T_i > W^*\right], \tag{2.24}$$

where W_i denotes taxable net wealth, T_i is third-party reported net wealth, W^* defines the kink and δ is a lower bound on the amount of bunching per household. Thus, the counterfactual density is given by the number of households to the left of the kink in terms of third-party reported net wealth. I group the third-party reported net wealth distribution into bins of SEK $5,000$ and use the number of households in the bin closest to the kink as a counterfactual.[36]

Table 2.3 presents the results using this method. The standard errors are computed using a bootstrap method, in which new distributions of third-party reported net wealth and taxable net wealth are drawn with replacement from the true distribution. The standard error of the bunching coefficient is represented by the standard deviation of the distribution of estimated bunching coefficients. Estimated excess mass is always significant, with the magnitude increasing in the bandwidth of allowed bunching. The coefficient is also larger when using the extrapolation method. The point estimates are of the same order of magnitude as those obtained using the parametric method. The heterogeneity in bunching across skills is validated. In contrast to the results discussed in Section 2.4.1, estimated bunching is always significant also for the low skilled.

The nonparametric approach provides a measure of bunching at the household level, a novel feature in the bunching-literature. The graphs presented in Figure 2.16 show how bunching follows the threshold. Panel (a) displays bunching of couples within a window of SEK $25,000$ below the kink of SEK $900,000$. In the years 1999 and 2000, this constituted the threshold for the marginal tax rate. However, in 2001 all bunching disappears. The same pattern is supported by the other graphs.

To summarize, the two estimation procedures give statistically significant estimates of bunching. The parametric approach has the advantage that it allows for the response to taxes to occur both through real and reporting responses. The nonparametric procedure, in contrast, presumes that the response occurs exclusively along the self-reporting margin. In the case of savings responses, the second approach, described by equation (2.24), thus underestimates the amount of bunching. On the other hand, bunching estimates from the first approach rely on the estimated counterfactual distribution being correct. In practice, the estimated excess mass will depend on the imposed interval width of bunching and

[36] Taking the average of the third-party reported net wealth in the interval of $[W^* - \delta, W^*]$ does not affect the results qualitatively, and the impact on estimated magnitudes is small.

functional form assumptions about the order of the counterfactual polynomial in (2.22). When associating the excess mass estimates with the theoretical models, I explore a range of estimated values of bunching as a robustness exercise.

2.4.3 A Proxy for Tax Evasion

The bunching documented above begs the question of whether the observed behavior is due to evasion or real savings responses. To address these questions, an empirical proxy for evasion is needed. I argue that a natural candidate for such a proxy is self-reported net wealth.

First, a simple comparison of the distributions of third-party reported and taxable net wealth in Figures 2.4 and 2.5, respectively, suggests that responses occur within the category of self-reported assets and liabilities. However, the absence of bunching in third-party reported net wealth does not automatically imply that savings are unaffected by the tax. If agents' possibility to evade were to be eliminated, bunching through real responses might arise.

Another way of assessing bunching through evasion is to study dynamic responses over time. Since wealth is a stock and not a flow variable, any bunching through real intertemporal responses would only manifest gradually over time. The estimated bunching upon shifting the kink should thus be different immediately after shifting the kink, as compared to, say, five years after. However, Figures 2.9 and 2.10 lend no support to the hypothesis that the excess mass estimates change over time. The observed pattern could be consistent with bunching through real responses if self-reported assets were more liquid than assets included in third-party reported wealth. However, liquid assets, such as bank account holdings and funds, are typical third-party reported assets while self-reported assets mainly comprise more illiquid assets such as cars and boats.

Second, if self-reported wealth indeed constitutes a proxy for tax evasion, it should be mirrored in monetary penalties for tax sheltering. Regressing imposed penalties on self-reported wealth and the positive difference between third-party reported and taxable net wealth, I do find that larger differences are associated with larger fines. To account for the endogenous selection of observed monetary fines – both in audit probabilities and likelihood of detection, conditional on an audit – I estimate the relation between evasion penalties and self-reported downward adjustments taking the endogenous selection into account. Following Heckman [1979], Table 2.4 presents the estimates from both the main and selection equations. A one SEK increase in positive downward adjustment gives an increase in the imposed fine in the range of SEK 0.02, depending on the specification.

Third, I cross-check self-reported assets against a register covering car holdings for all Swedes. Statistics Sweden administers register data on car holdings of the population. I link these records to the tax data. Since cars are part of the tax base and a self-reported asset, I compare the car holdings in the register data to the self-reported assets. If self-reported assets are lower than the value of car holdings, this is suggestive of tax evasion. However, since self-reported assets include other assets than just cars, self-reported assets being larger than car holdings should not be considered as proof of no evasion.

The car register data include information on car brand, model and vintage. To cross-check self-reported assets against car holdings, I collect price data on cars produced from 1990-2006 from the Swedish Tax Agency.[37] To assess the value of the cars I use a devaluation model from *Bilpriser*, a company that collects data on purchases of used cars.

Figure 2.20 presents the mean value of cars against taxable income around the threshold. Tax payers may have responded to the wealth tax by buying cheaper cars. Such distortionary effects would manifest themselves as a flatter slope of car values to the right of the kink point. The figure, however, suggests that people do not adjust their car purchases, consistent with the evidence of no responses in third-party reported wealth.

Figure 2.21, on the other hand, shows the fraction of car owners who report assets of at least the same value as their car holdings, against third-party reported net wealth to the right of the kink point.[38] Close to the kink, only 15–20 percent of the car owners actually report their cars, consistent with evasion being behind the responses. Even if car owners make imprecise predictions of what their cars are worth, it seems unlikely that noise in the valuation of the car explains this striking pattern.

These findings suggest that the difference between third-party reported and taxable net wealth may serve as a proxy for tax evasion. In the analysis that follows, I therefore treat it as such and refer to this difference as my measure of tax evasion or tax sheltering.

2.4.4 Estimating p

In addition to bunching estimates, parametrizing the models of Section 2.2 and computing the implied tax elasticities requires knowledge of the unit cost of tax sheltering. The cost function, $C(e,s) = \left(\frac{e}{s}\right)^{\gamma} \frac{pe}{1+\gamma}$, features a unit cost of evasion, denoted p. To measure p, I use data on the total value of sentenced fines per person. In practice, this cost should be viewed as a lower bound on the actual unit cost of evasion. First, it does not include transfers across non-government agents such as payments to accountants and tax planners. Second, according to officials at the Tax Authority, the detection of wealth tax sheltering increased the risk of being caught evading other taxes. Third, a history of tax fraud was a key parameter of the Tax Authority's auditing strategy, implying that

[37]The price data only stretches back until 1990. I believe that this is not problematic for two reasons. Veteran cars were tax exempt and therefore not present in the self-reported assets. Second, the value of cars dating from before 1990 that are not of veteran status are likely to be of low value and if anything, this should bias the evasion estimates downwards.

[38]Plotting this graph for third-party reported wealth levels below the kink is not meaningful, as the sample of tax payers to the left of the kink who actually self-report assets is selected.

detection was likely to lead to more frequent audits in the future. Finally, the measure does not include the (monetary) cost of going to court.[39] Nevertheless, sentenced fines should be able to serve as a proxy for the cost of tax sheltering.

Ignoring functional form assumptions regarding the relation between evasion and penalties, the mean amount of imposed fines, as a fraction of tax evasion, among the subset of individuals who do evade and are detected, is 0.86. In my derivation of the tax elasticities in the next section, $p = 0.86$ is thus chosen as the benchmark estimate.

2.5 Estimating Elasticities

In this section, I use the bunching and evasion cost estimates from Section 2.4 to compute tax elasticities according to the theories in Section 2.2. In Section 2.5.1, I use the framework for small-kink analysis, presented in Section 2.2.1. The sheltering parameter in the static model, layed out in Section 2.2.2, is identified in Section 2.5.2. In Section 2.5.3, I extend the analysis to include a savings response in accordance with the dynamic model of Section 2.2.3. Finally, in Section 2.5.4, I relate my results to the bounded-rationality model of Section 2.2.4 and analyze heterogenous bunching on the basis of cognitive ability.

2.5.1 Small Kink Analysis

Recall that equation (2.1), $B/h\left(z^{*}\right)z^{*} = -\varepsilon^{K}_{W,\tau}\mathrm{d}\tau$, identifies the compensated tax semi-elasticity of taxable net wealth, $\varepsilon^{K}_{W,\tau}$, without making any parametric assumptions about the preferences generating observed behavior. The approach is valid, as long as the kink is small and the associated income effects are negligible. The first column of Table 2.5 presents estimated elasticities based on different estimates of bunching obtained using the parametric and nonparametric methods of Section 2.4. The elasticities are in the range $[0.12, 0.32]$ with higher elasticities reflecting larger bunching estimates.

An important issue when analyzing the impact of wealth taxes is the time horizon considered. How much of the effect occurs already in the short run as compared to the long run? I use the small kink analysis to assess the evolution of the tax elasticities over time. Incorporating shifts in the threshold, the results are indexed by the time that the kink has remained constant. In 2002, the kink was shifted from SEK $1,000,000$ to $1,500,000$ for single households and was then held fixed at the latter level, until the tax was repealed in 2006. The year-by-year plots of taxable net wealth in Figure 2.9 reveal that the extent of bunching is fairly stable during the period 2002-2006, which yields steady elasticity estimates over time. Analyzing third-party reported net wealth each

[39] In an interview with representatives from the Tax Authority (Skatteverket), the evaded amount is only one parameter that determines whether the case is brought to court. Small amounts of tax sheltering may also lead to court if the tax authority can verify a history of cheating.

year, as in Figure 2.17, suggests no dynamic effects on bunching in accumulated savings over time.[40]

Next, I estimate taxable net wealth elasticities using variation over time, i.e., shifts in the threshold. Changes in the kink generate different changes in tax rates across households depending on their net wealth. This motivates a difference-in-difference research design.

Let $\Delta \log W_{i,t}^{x} = \log W_{i,t} - \log W_{i,t-x}$ represent the log change in taxable net wealth between year t and $t-x$, (for $x = 1, 2, 3$) and let $\Delta \log NTR_{i,t}^{x}$ denote the log change in net-of-tax rates over the same time period. Following Gruber and Saez [2002], I then estimate:

$$\Delta \log W_{i,t}^{x} = \alpha_i + \theta \Delta \log NTR_{i,t}^{x} + f\left(W_{i,t-x}\right) + \mu X_{i,t-x} + \varepsilon_{i,t} \tag{2.25}$$

using two-stage-least-squares (2SLS). I instrument for the log change in the net-of-tax rate, $\Delta \log NTR_{i,t}^{x}$, by $\Delta \log NTR_{i,t}^{x,sim}$, the simulated change in the net-of-tax rate holding the household's wealth equal to its level in the base year. $f\left(W_{i,t-x}\right)$ denotes a 10-piece linear spline in taxable net wealth and $X_{i,t-x}$, a vector of household characteristics. Both are measured at the base year. The regression includes a household fixed effect α_i, capturing household-specific time trends.

When estimating (2.25) by 2SLS, the first-stage regressions of $\Delta \log NTR_{i,t}^{x}$ on $\Delta \log NTR_{i,t}^{x,sim}$ yield coefficients around 0.75 with implied t-statistics above 50. To avoid mean reversion in wealth driving the results, I exclude households with base-year taxable net wealth values below SEK $800,000$. Columns (5) in Table 2.5 present estimated tax elasticities of taxable net wealth. All specifications include year fixed effects and a spline in base-year taxable net wealth. The estimated elasticities are approximately semi-elasticities, given that $\Delta \log (1-\tau) \approx \Delta \tau$ for low taxes. Exploiting shifts in the threshold yields larger estimates of the tax elasticity, in the range $[0.45, 0.80]$, depending on the chosen time interval.

2.5.2 Static Model

The static model in Section 2.2.2 assumes that the response to the wealth tax occurs along the evasion margin. In the parametrized model, bunching at the kink reflects the fraction of sheltered wealth. Larger estimates of excess mass thus imply larger fractions of evasion. Using the benchmark estimated value of the unit cost of evasion, $\hat{p} = 0.86$, column (2) of Table 2.5 presents tax semi-elasticities of taxable net wealth for a range of bunching estimates. The elasticities are slightly larger compared to the small-kink analysis and take values in the range $[0.18, 0.42]$.

[40] Figures 2.10 and 2.18 display bunching of taxable net wealth and third-party reported net wealth, respectively, over time, for couples. These graphs corroborate the finding that there are no dynamic responses of savings.

The semi-elasticities of evasion, presented in column (1) of Table 2.6, are substantial, ranging between $[87.8, 105.7]$. The large magnitudes arise as the estimated fraction of evasion is small. At low levels of evasion, the percentage change in evasion arising from a one percentage point increase in the tax rate is large. However, the tax elasticities of evasion are smaller and the estimates are slightly larger than one, indicated by column (2) of Table 2.6. At higher tax rates, for example, $\tau = 0.35$, the response is lower partly due to higher fractions evaded, with semi-elasticities around 2.5. Such a hypothetical tax rate is high in the context of a wealth tax, but when applying the framework to tax sheltering of, for instance, labor income taxes, such an example is realistic. On the other hand, the evasion cost technology is different for higher tax rates and other tax rates, in particular labor-income taxes, so out-of-sample predictions should be interpreted with caution.[41] The implied semi-elasticities of around 2.5 are in line with those in Fisman and Wei [2004], that estimate a semi-elasticity of evasion of 3 using tax rates in the range of 30 to 50 percent, and within the same range as the estimates between 0.5 and 3 reported by Clotfelter [1983].

Using data on self-reported downward-adjustments of taxable net wealth, I next study how much money is saved by bunchers. The estimates vary slightly depending on whether extrapolation is used. A conservative estimate not using the extrapolation method reveals that the median shelters SEK $220,000$ from the government and therefore gains SEK $3,300$ from not paying the tax. The average values correspond to SEK $400,000$ and SEK $6,000$.[42]

The parameterized model predicts a linear relation between sheltered resources and wealth. Figure 2.19 tests this by plotting mean evasion against third-party reported net wealth. When estimating this relationship by means of regression, higher order terms are statistically significant but the magnitude of the estimated coefficient is practically zero.

2.5.3 Dynamic Model

When I allow bunching to arise from adjustment on two margins, equation (2.14) states that bunching comprises the sum of the savings responses and reporting responses. In the parameterization of the model, I use standard values from the macro literature: $\beta = 0.96$ and $\sigma = 0.25$.[43] The results are not sensitive to how these parameters are chosen and I also present results for $\sigma = 1.1$. With $\sigma < 1$, the fraction of income saved is larger on the right-side of the kink than on the left-hand side. To make estimated bunching consistent with this fact, the estimated evaded fraction has to be larger here compared

[41] Kleven et al. [2011], Slemrod and Kopczuk [2002] argue that the responsiveness of evasion to tax changes depends on various parameters including the enforcement regime and the definition of the tax base.

[42] These computations allow for bunching in an interval of SEK $100,000$ below the kink.

[43] $\beta = 0.96$ reflects that the period length corresponds to one year.

to the case when $\sigma > 1$. Moreover, such a parameter value implies that a tax increase triggers evasion both directly and indirectly due to increased savings. In contrast, if $\sigma > 1$, agents respond to a tax increase by both reducing savings and increasing evasion. The response of evasion is mitigated by a decrease in savings.

Columns (3) and (4) of Table 2.5 present uncompensated tax semi-elasticities of taxable net wealth using different excess mass estimates.[44] The uncompensated elasticities – now reflecting both intertemporal substitution and tax sheltering responses – are close to the estimates obtained in the static framework. A lower σ implies more evasion and a larger tax elasticity of taxable net wealth. For $\sigma = 0.25$, the elasticities range in $[-0.20, -0.40]$.

The bunching estimates and the parameter choices imply that the real responses to tax changes are almost zero and the responses to tax changes occur mainly along the tax-sheltering margin. This is manifested in columns (3) - (6) in Table 2.6, displaying uncompensated and income-compensated semi-elasticities and regular elasticities of evasion. Evasion responses are slightly larger in the dynamic case. Even though the savings response is an outcome of the parameterization, varying the parameter values persistently yields this result.

2.5.4 Bounded Rationality

In the theory presented in Section 2.2.4, the cognitively less able find it too costly to take the tax rate into account when optimizing. Figure 2.12 presents the distribution of taxable net wealth for high- and low-skilled households over the full sample period, while Figures 2.13 and 2.14 present the corresponding distributions and an associated bunching estimate for the years 2002-2006. To maintain statistical power, I retain the extrapolation method described in Section 2.3 for dealing with missing taxable wealth values throughout the analysis of cognitive skills. The analysis in this section is primarily conducted on the stable years 2002-2006, i.e., the period during which the kink was constant for single households and only shifted once for couples.

The underlying cognitive ability variable is normally distributed with zero mean and unit variance. The samples in the figures are based on dividing the sample into two groups with positive and negative z-scores, respectively. Figure 2.15 compares the top 25th percentile of the skill distribution to the bottom 25th. The observed elasticities, estimated using the small kink analysis summarized by equation (2.1), $\frac{B}{h(z^*)z^*} = -\varepsilon^K_{W,\tau} \mathrm{d}\tau$ and the parametric estimates in Figures 2.14 and 2.15, are presented in Table 2.5, in the rows denoted *Parametric* and *High* and *Low*.

[44]The compensated tax elasticities of taxable net wealth are positive. Neither savings nor taxable net wealth are arguments in the agents' utility function. A tax increase implies lowered consumption in the second period, also if the tax increase is compensated for, but the impact on taxable net wealth is positive.

The nonparametric estimates of bunching based on cognitive skills are displayed in Table 2.3, denoted by b_{HIGH} and b_{LOW}. The associated elasticities are found in the rows labeled *Nonparametric* and *High* and *Low*, respectively, of Table 2.5. For the high-skilled, the elasticities range between $[0.20, 0.50]$, whereas the corresponding interval for low-skilled is $[0.05, 0.32]$. These elasticities are not structural parameters, as they also reflect the optimization errors incurred by bounded rationality.

Exploiting my empirical proxy for tax sheltering, I test the theoretical prediction that high-skilled households evade more. This is done in two steps. First, I investigate self-reports that involve bunching, i.e., households locating above the kink in third-party reports and below the kink in taxable net wealth. Second, I study the intensive margin, namely the amount of self-reported adjustment in SEK.

To investigate extensive-margin responses, I regress a binary indicator of bunching on a 10-piece linear spline in third-party reported net wealth, wage earnings and education, on the sample of men with enlistment data located between SEK 1.5 and 2 million in terms of third-party reported net wealth. Figure 2.22 plots the residuals from this regression against cognitive skills. This captures the partial relation between skills and bunching while controlling for wealth, income and education. The results suggest that, conditional on important factors, cognitive skills and bunching are positively related.

The intensive-margin response is investigated in a similar way. I regress the evaded amount on a 10-piece linear spline in third-party reported net wealth, wage earnings and education. Figure 2.23 plots the residuals from this regression against cognitive skills. Although less obvious, there is still a positive relation between self-reported wealth and cognitive skills also along this margin.

I quantify this positive relation in the regressions presented in Table 2.8. Controlling for education, third-party reported net wealth and wage earnings, a one-standard-deviation increase in cognitive ability is associated with an increase in the probability of bunching by around 1 percent. The same increase in skills increases the percentage evaded by about 0.14 percent.

These results provide consistent support for the prediction that low-ability households respond less to tax changes than households with higher ability. I discuss alternative mechanisms behind this pattern in the next section.

2.6 Heterogeneous Responses to Wealth Taxes

The results in Section 2.5.4 provide support for the prediction that households with a low cognitive ability respond less to tax changes than households with a higher ability. The bounded-rationality model presumes that these findings are driven by different abilities to understand the tax system. However, alternative channels could also explain the correlation between cognitive ability and evasion. I next explore the two alternative mechanisms that are arguably the strongest contenders: heterogenous costs of tax sheltering and heterogenous preferences.

To address the possibility that agents have access to different technologies for tax evasion, I regress sentenced fines on cognitive skills, controlling for evasion, wealth and income levels. The results, presented in columns (1) and (2) in Table 2.9, confirm that once evasion, wealth and income are taken into account, cognitive skills are of no importance for the amount of sentenced penalties. In rare instances, it is marginally statistically significant but with the wrong sign, i.e., suggesting a positive correlation between skills and sentenced fines. These findings are consistent with the view that the same technologies are available to all.

To investigate whether different preferences are driving the results, I use a unique dataset on risk and time preferences obtained from a survey in Mollerstrom and Seim [2012]. The survey elicits these preferences by survey questions developed in lab experiments and proven to correlate closely with results from actual experiments.[45] To compute a risk-aversion index, the subjects were asked to choose between a riskfree alternative and a gamble with two outcomes, holding the expected value of the gamble fixed but ranging the riskfree amount over a number of questions. The number of times the subject chose the riskfree alternative comprises a measure of risk aversion.

Similarly, a time-preference index is based on a choice between receiving money today and money in twelve months, where today's value was held fixed and the value in the future was varied. The impatience index measures the number of times the subject opted for receiving money today.

Matching these survey data at the individual level to enlistment data on cognitive ability yields a sample of 153 individuals. The results from regressing risk and time preferences on cognitive skills, displayed in columns (3) and (4) in Table 2.9, suggest no statistically significant correlation between them. In contrast to the previous literature, the point estimates would suggest that high-skilled individuals are more risk averse while, in line with earlier studies, the point estimates on impatience are negative.[46]

To summarize, I find support neither for heterogeneity in evasion technology across skill groups nor for heterogeneity in risk- and time preferences across skill groups.

2.7 Conclusion

In this paper, I address the behavioral effects of a wealth tax from different perspectives. I merge detailed administrative data with Swedish tax records over the period 1999-2006

[45]The questions build directly on the work by Falk and Becker [2010], which aims at finding survey methods for eliciting risk and time preferences that are akin to lab experiments.

[46]Benjamin et al. [2013], Frederick [2005] find that individuals with a high cognitive ability are less risk averse and less impatient.

and exploit features of the institutional setup to identify tax elasticities of taxable net wealth. In addition, I address potential mechanisms for the observed behavior and study whether effects are heterogeneous across skill groups.

The Swedish wealth tax comprised two tax brackets: one with a zero marginal tax rate and one with a marginal tax rate of 1.5 percent. I find strong evidence of bunching at the threshold separating the two brackets and report tax semi-elasticities of taxable net wealth in the absolute range $[0.12, 0.33]$. When I instead exploit the variation induced by movements of the threshold over time, I find larger semi-elasticities of taxable net wealth: in the range of 0.45 to 0.80.

Importantly, the tax base comprised both net wealth reported by third-parties, as well as self-reported assets and liabilities. This institutional feature allows me to disentangle savings responses from reporting responses. My results suggest that the observed bunching occurs exclusively along the self-reported margin. I conjecture that the tax triggers evasion and corroborate this by two pieces of evidence. First, I find that self-reported downward adjustments of taxable net wealth are positively correlated with monetary penalties for evasion. Second, following movements of the threshold over time, I find no evidence of dynamic adjustments in bunching along third-party reported net wealth and therefore no signs of intertemporal adjustments in terms of savings.

To learn more about the underlying mechanisms behind the observed effects, I explore different theoretical models and take them to the data using the bunching estimates. When translating my findings into tax semi-elasticities of taxable net wealth, my results suggest that they are in the range $[0.18, 0.42]$ when I employ a simple static setting, and in the range $[0.20, 0.40]$ when I explore a dynamic framework.

Since the tax system often is perceived as difficult to comprehend, I conjecture that individuals with high cognitive ability are more likely to understand the workings of the tax scheme. I test this by exploiting results from cognitive-skills tests from military enlistments. My findings suggest that the cognitively able respond more. The mass of households who locate just below the threshold for paying the positive marginal tax rate, in relation to an estimated counterfactual distribution, is greater for the subsample of high-skilled households than for low-skilled households. Also, on the intensive margin, I find that self-reported downward adjustments are larger for high-skilled households compared to low-skilled, even when controlling for income, wealth levels, type of wealth held and education. To address whether cognitive ability indeed lies behind the differences across groups, I explore key alternative explanations: having access to different technologies for evasion or different preferences with respect to risk and time. None of these alternatives are supported by the data.

Due to the scarce empirical evidence on taxes on wealth and intergenerational transfers, future research should aim at identifying the long-term effects of such taxes. The data used in the paper at hand only comprise nine years of observations and the horizon may be too short for long-run effects on accumulated assets to be observable. Additionally, the wealth-accumulation responses in the present paper arise from parameter values chosen from the macro literature. Future research should estimate, simultaneously, parameters that determine real and reporting responses. Obtaining accurate measures of the long-term impact of wealth and capital taxes on wealth accumulation may provide guidelines in the design of public policy.

Another interesting avenue of future research would be to explore taxes on bequests. Understanding whether households circumvent estate and inheritance taxes, or if these taxes induce real responses, is crucial for the design of future tax schemes related to inter-generational transfers.

On a final note, my findings on heterogeneous effects across skill groups demonstrate the merits of salient and simple tax policies. If governments cannot reduce the costs of internalizing tax policies, for instance by providing information about the workings of different tax schemes, they should make transparent and easily understandable tax policies a priority. Exploring how household heterogeneity affects optimal public policy is a challenging but truly important task.

Bibliography

S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba. Taxes on wealth transfers. In *Tax by Design. The Mirrlees Review*, chapter 14, pages 347–367. Oxford University Press, 2011.

J. Agell, P. Englund, and J. Söodersten. Tax reform of the century – the swedish experiment. *National Tax Journal*, 49:643–664, 1996.

M. G. Allingham and A. Sandmo. Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1:323–338, 1972.

A. Atkinson and J. E. Stiglitz. The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 6:55–75, 1976.

J. Banks and P. Diamond. The base for direct taxation. In James A. Mirrlees, Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba, editors, *Dimensions of Tax Design: the Mirrlees Review*, chapter 6, pages 548–674. Oxford University Press, 2010.

D. J. Benjamin, S. A. Brown, and J. Shapiro. Who is "behavioral"? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 2013. URL `http://economics.cornell.edu/dbenjamin/iq.pdf`.

R. Boadway, E. Chamberlain, and C. Emmerson. Taxation of wealth and wealth transfers. In James A. Mirrlees, Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba, editors, *Dimensions of Tax Design: the Mirrlees Review*, chapter 8, pages 737–836. Oxford University Press, 2010.

R. D. Brown. A primer on the implentation of wealth taxes. *Canadian Public Policy / Analyse de Politiques*, 17:335–350, 1991.

B. Carlstedt. Cognitive abilities: Aspects of structural, process and measurement, 2000. Gothenburg University.

C. Chamley. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica*, 54:607–622, 1986.

R. Chetty. Is the taxable income elasticity sufficient to calculate deadweight loss? the implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1:32–52, 2009.

R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *NBER Working Paper 13330*, 2007.

R. Chetty, A. Looney, and K. Kroft. Salience and taxation: Theory and evidence. *American Economic Review*, 99:1145–1177, 2009.

R. Chetty, J. N. Friedman, T. Olsen, and L. Pistaferri. Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics*, 126:749–804, 2011.

C. T. Clotfelter. Tax evasion and tax rates: An analysis of individual returns. *Review of Economics and Statistics*, 65:363–373, 1983.

J. Conlisk. Why bouned rationality? *Journal of Economic Literature*, 34:669–700, 1996.

P. Diamond and E. Saez. The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives*, 25:165–190, 2011.

E. Grönqvist, B. Öckert, and J. Vlachos. The intergenerational transmission of cognitive and non-cognitive abilities. *IFAU Working Paper 2010:12*, 2010.

A. Falk and A. Becker. Economic preference module, 2010. University of Bonn.

N. E. Feldman and P. Katuščàk. Should the average tax rate be marginalized? *Center for Economic Research and Graduate Education Economics Institute Working Paper 304*, 2006.

M. Feldstein. The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy*, 103:551–572, 1995.

M. Feldstein. Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81:674–680, 1999.

R. Fisman and S. Wei. Tax Rates and Evasion: Evidence from "Missing Imports" in China. *Journal of Political Economy*, 112:471–496, 2004.

S. Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19:25–42, 2005.

J. Gruber and E. Saez. The elasticity of taxable income: Evidence and implications. *Journal of Public Economics*, 84:1–32, 2002.

H. Hammar, S. C. Jaegers, and K. Nordblom. Skatter och skattefusk, 2006.

J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, pages 153–161, 1979.

M. Henrekson and U. Jakobsson. Ägarpolitik och Ägarstruktur i Efterkrigstidens Sverige. In *Vem Ska Äga Sverige?* Stockholm: SNS Förlag, 2001.

K. Judd. Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics*, 28:59–83, 1985.

H. J. Kleven and M. Waseem. Tax notches in pakistan: Tax evasion, real responses, and income shifting. *Quarterly Journal of Economics*, Forthcoming.

H. J. Kleven, M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez. Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79: 651–692, 2011.

W. Kopczuk. Taxation on intergenerational transfers and wealth, 2013. Prepared for Handbook of Public Economics.

J. B. Liebman and R. J. Zeckhauser. Schmeduling, 2004. Harvard University.

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

J. Mayshar. Taxation with costly administration. *Scandinavian Journal of Economics*, 93:75–88, 1991.

J. Mollerstrom and D. Seim. Does the demand for redistribution rise or fall with cognitive ability, 2012. Harvard University.

Nationalekonomiska Föreningen. De Nya Skatteförslagen, 1948. Stockholm: Häggströms.

H. Ohlsson. The legacy of the swedish gift and inheritance tax, 1884-2004. *European Review of Economic History*, 15:539–569, 2011.

T. Piketty and E. Saez. A theory of optimal capital taxation. *NBER Working Paper 17989*, 2012.

R. Reis. Inattentive consumers. *Journal of Monetary Economics*, 53:1761–1800, 2006.

E. Saez. Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2:180–212, 2010.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50: 665–690, 2003.

C. A. Sims. Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96:158–163, 2006.

J. Slemrod. A general model of the behavioral response to taxation. *International Tax and Finance*, 8:119–128, 2001.

J. Slemrod and W. Kopczuk. The optimal elasticity of taxable income. *Journal of Public Economics*, 84:91–112, 2002.

J. Slemrod and S. Yitzhaki. Tax avoidance, evasion, and administration, 2002. Handbook of Public Economics.

Swedish Government Official Reports. Reformerade egendomsskatter, 2004.

S. Yitzhaki. A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3:475–480, 1974.

A Appendix

A.1 Infinite Horizon Model

This section presents the infinite-horizon version of the two-period model laid out in section 2.2.3. Utility is additively separable in consumption and has the common *CES* form. Formally, the optimization problem is:

$$U = \max_{A_{t+1},e_t} \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\frac{1}{\sigma}} - 1}{1 - \frac{1}{\sigma}} \tag{2.26}$$

subject to

$$
\begin{aligned}
c_t + s_t &= y_t \\
A_{t+1} &= s_t + (1-\tau)(A_t - e_t) + e_t - \left(\frac{e_t}{A_t}\right)^{\frac{1}{\gamma}} \frac{pe_t}{1+\frac{1}{\gamma}} \\
c_t &\geq 0 \\
A_0 &\text{ given.}
\end{aligned}
$$

In an interior solution, the problem has first-order conditions

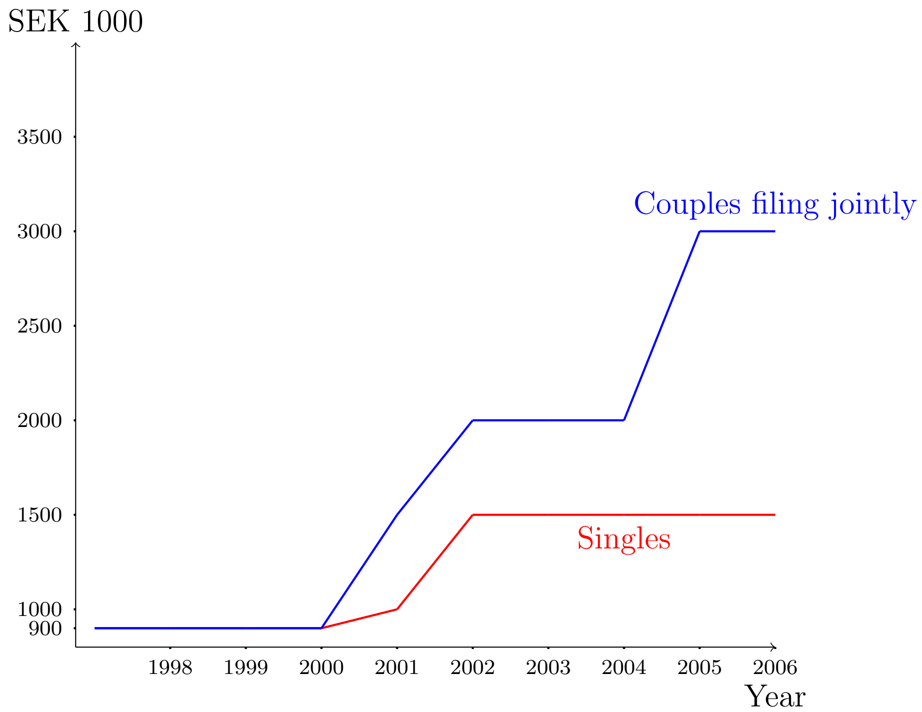
$$
\begin{aligned}
e_t &= \left(\frac{\tau}{p}\right)^{\gamma} A_t \\
c_t^{-\frac{1}{\sigma}} &= \beta \left(1 - \tau + \tau \left(\frac{\tau}{p}\right)^{\gamma} \frac{1}{1+\gamma}\right) c_{t+1}^{-\frac{1}{\sigma}}.
\end{aligned}
$$

These equations describe the dynamics of savings, wealth, evasion and consumption. Starting out with any tax rate and wealth distribution, it is possible to describe the evolution of the response to a tax reform. On a balanced growth rate, all variables grow at a constant rate given by:

$$
g = \left(\beta \left(1 - \tau + \tau \left(\frac{\tau}{p}\right)^{\gamma} \frac{1}{1+\gamma}\right)\right)^{\sigma} \tag{2.27}
$$

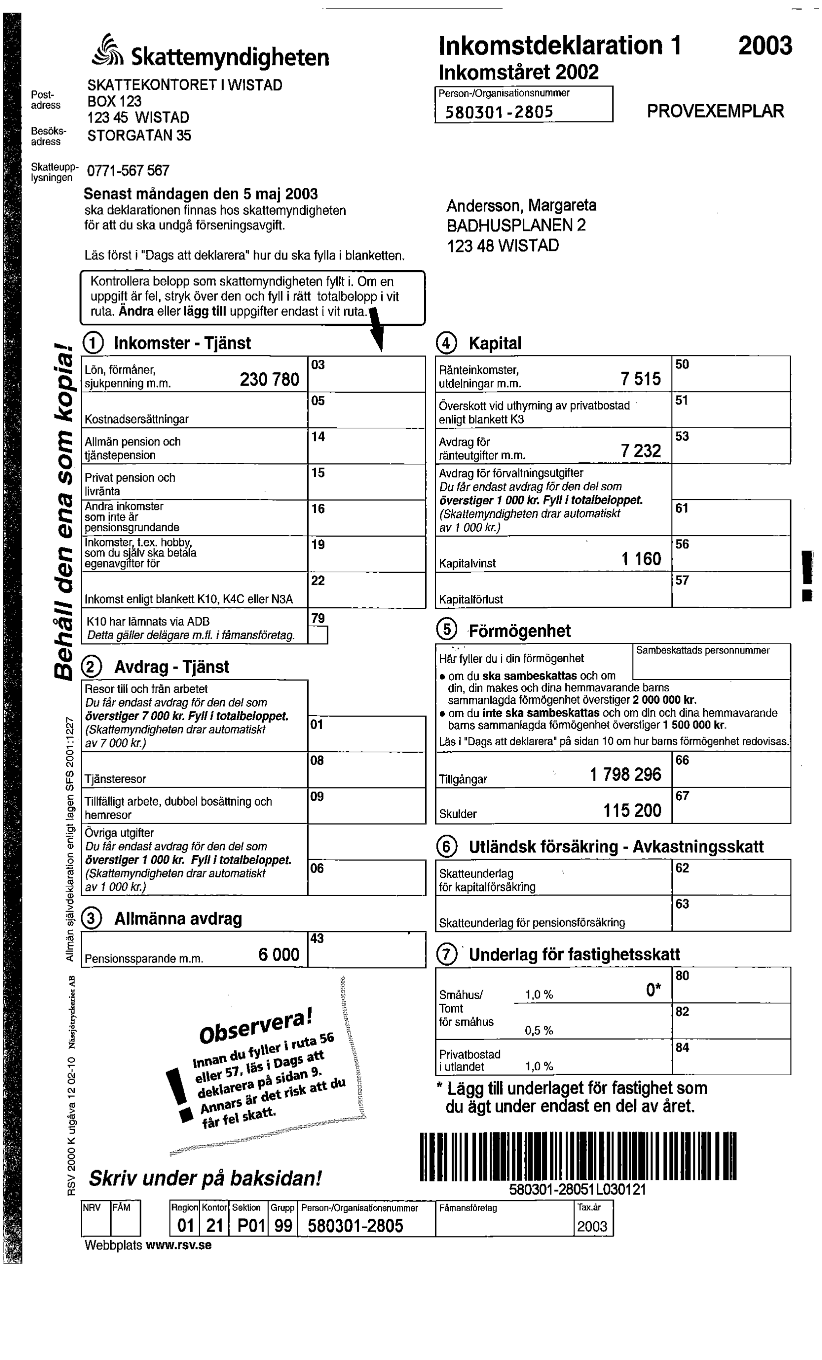
An increase in the tax rate has a distortionary effect on the growth rate. However, this effect is attenuated by agents' tendency to shelter money from the government. From a growth-enhancing point of view, sheltering behavior is not too bad.

Figure 2.1: Marginal tax rates over time.



Notes: The figure shows the evolution of the marginal tax rate bracket over time, for single households and couples without children who file the tax jointly.

Figure 2.2: Prepopulated tax return.


Skatteverket logo: Skattemyndigheten

Inkomstdeklaration 1 2003

Inkomståret 2002

Person-/Organisationsnummer 580301-2805

PROVEXEMPLAR

Post-adress: SKATTEKONTORET I WISTAD, BOX 123, 123 45 WISTAD

Besöks-adress: STORGATAN 35

Skatteupplysningen: 0771-567 567

Senast måndagen den 5 maj 2003 ska deklarationen finnas hos skattemyndigheten för att du ska undgå förseningsavgift.

Läs först i "Dags att deklarera" hur du ska fylla i blanketten.

Andersson, Margareta
BADHUSPLANEN 2
123 48 WISTAD

Kontrollera belopp som skattemyndigheten fyllt i. Om en uppgift är fel, stryk över den och fyll i rätt totalbelopp i vit ruta. **Ändra** eller **lägg till** uppgifter endast i vit ruta.

Behåll den ena som kopia!

① Inkomster - Tjänst

Lön, förmåner, sjukpenning m.m.	230 780	03	
Kostnadsersättningar		05	
Allmän pension och tjänstepension		14	
Privat pension och livränta		15	
Andra inkomster som inte är pensionsgrundande		16	
Inkomster, t.ex. hobby, som du själv ska betala egenavgifter för		19	
Inkomst enligt blankett K10, K4C eller N3A		22	
K10 har lämnats via ADB. *Detta gäller delägare m.fl. i fåmansföretag.*		79	

② Avdrag - Tjänst

Resor till och från arbetet. *Du får endast avdrag för den del som **överstiger 7 000 kr. Fyll i totalbeloppet.** (Skattemyndigheten drar automatiskt av 7 000 kr.)*	01	
Tjänsteresor	08	
Tillfälligt arbete, dubbel bosättning och hemresor	09	
Övriga utgifter. *Du får endast avdrag för den del som **överstiger 1 000 kr. Fyll i totalbeloppet.** (Skattemyndigheten drar automatiskt av 1 000 kr.)*	06	

③ Allmänna avdrag

Pensionssparande m.m.	6 000	43	

④ Kapital

Ränteinkomster, utdelningar m.m.	7 515	50	
Överskott vid uthyrning av privatbostad enligt blankett K3		51	
Avdrag för ränteutgifter m.m.	7 232	53	
Avdrag för förvaltningsutgifter. *Du får endast avdrag för den del som **överstiger 1 000 kr. Fyll i totalbeloppet.** (Skattemyndigheten drar automatiskt av 1 000 kr.)*		61	
Kapitalvinst	1 160	56	
Kapitalförlust		57	

⑤ Förmögenhet

Här fyller du i din förmögenhet — Sambeskattads personnummer
- om du **ska sambeskattas** och om din, din makes och dina hemmavarande barns sammanlagda förmögenhet överstiger **2 000 000 kr.**
- om du **inte ska sambeskattas** och om din och dina hemmavarande barns sammanlagda förmögenhet överstiger **1 500 000 kr.**

Läs i "Dags att deklarera" på sidan 10 om hur barns förmögenhet redovisas.

Tillgångar	1 798 296	66	
Skulder	115 200	67	

⑥ Utländsk försäkring - Avkastningsskatt

Skatteunderlag för kapitalförsäkring	62	
Skatteunderlag för pensionsförsäkring	63	

⑦ Underlag för fastighetsskatt

Småhus/	1,0 %	0*	80	
Tomt för småhus	0,5 %		82	
Privatbostad i utlandet	1,0 %		84	

* Lägg till underlaget för fastighet som du ägt under endast en del av året.

Observera! Innan du fyller i ruta 56 eller 57, läs i Dags att deklarera på sidan 9. Annars är det risk att du får fel skatt.

Skriv under på baksidan!

580301-28051L030121

NRV	FÅM	Region	Kontor	Sektion	Grupp	Person-/Organisationsnummer	Fåmansföretag	Tax.år
		01	21	P01	99	580301-2805		2003

Webbplats www.rsv.se

RSV 2000 K utgåva 12 02-10 Allmän självdeklaration enligt lagen SFS 2001:1227

Notes: Section 5 in the figure displays the filing of the wealth tax. The section *Tillgångar* refers to assets. Taxpayers were supposed to fill in the total value of taxable assets in field 66 if their taxable net wealth exceeded the threshold. The section *Skulder* refers to liabilities. Taxpayers filled in the total value of liabilities in field 67.

Table 2.1: Wealth taxpayers and government revenue.

Year	Number of Wealth Tax Payers	Share of Population Paying the Tax	Government Revenue (Million SEK)	Share of Total Government Revenue
1999	707997	0.080	8380	0.012
2000	685929	0.077	7990	0.010
2001	467534	0.053	6270	0.008
2002	258724	0.023	3790	0.005
2003	312705	0.035	4720	0.007
2004	328162	0.036	5140	0.007
2005	223139	0.025	4760	0.006
2006	276373	0.030	5870	0.007

Sources: Wealth tax data from Income- and Tax Register, population data from Statistics Sweden and government revenue data form The Swedish National Financial Management Authority.

Table 2.2: Summary statistics for the Swedish population and different subsamples, 1999-2006.

	Population	Singles and couples without children			
		All	Excluding self-employed		
			All	Wealth Tax Payers	Enlistment Data
	(1)	(2)	(3)	(4)	(5)
Demographics					
Age	47.704	53.068	52.699	65.970	37.919
Children	0.570	0	0	0	0
Primary School (%)	0.302	0.332	0.328	0.283	0.152
Secondary School (%)	0.436	0.425	0.426	0.361	0.534
Higher Education (%)	0.261	0.244	0.246	0.356	0.314
Female (%)	0.509	0.500	0.506	0.507	0
Married (%)	0.429	0.362	0.341	0.642	0.135
Income and Taxes					
Wage Earnings	130,747	113,056	113,396	116,792	220,515
Wealth Tax Payed	794	951	798	13,079	351
Third-Party Reported:					
Assets	404,722	418,724	343,164	1,878,774	256,197
Debt	199,058	151,973	135,651	100,727	242,403
Portfolio					
Real Estate (%)	0.312	0.254	0.252	0.374	0.252
Bank Account (%)	0.183	0.219	0.211	0.207	0.183
Funds (%)	0.188	0.203	0.201	0.233	0.182
Stocks (%)	0.056	0.063	0.059	0.143	0.087
Bonds (%)	0.018	0.023	0.022	0.040	0.011
Skills					
Cognitive Skills					-0.069
Observations	58,015,897	37,220,782	33,752,654	2,059,074	4,122,141

Notes: Table entries are means unless otherwise stated. Monetary values are in SEK. The population in column 1 comprises tax payers aged 16 and older. Columns (2)-(5) describe the subsample of single households and couples without children below 18 years of age. Children refers to the number of children below 18 years of age, living with the individual. Entrepreneurs are individuals who possess industrial property. The variable Cognitive Skills denotes the standardized value of the sum of subscores from the cognitive skills test taken at enlistment.

Table 2.3: Nonparametric estimates of bunching, $\hat{b}$, 1999-2006.

	(1)	(2)	(3)	(4)
All	0.489***	0.950***	1.447***	2.837***
	(0.009)	(0.014)	(0.019)	(0.033)
Couples	0.533***	1.036***	1.443***	2.864***
(0.013)	(0.021)	(0.028)	(0.046)	
Singles	0.453***	0.880***	1.447***	2.805***
	(0.012)	(0.017)	(0.025)	(0.044)
Men with military	0.740***	1.455***	1.949***	3.794***
enlistment data	(0.046)	(0.120)	(0.097)	(0.176)
High-skilled	0.894***	1.710***	2.257***	4.359***
	(0.073)	(0.120)	(0.150)	(0.248)
Low-skilled	0.542***	1.127***	1.444***	2.996***
	(0.056)	(0.096)	(0.115)	(0.214)
δ	40,000	100,000	40,000	100,000
Extrapolation	No	No	Yes	Yes

Notes: The number of households that bunch are computed in accordance with equation (2.24) and the counterfactual density at the kink is obtained using third-party reported net wealth. δ denotes the width of the bunching interval. Extrapolation refers to the method described in Section 2.3.2. High-skilled and low-skilled households are defined as having, respectively, positive and negative z-scores in cognitive ability. Standard errors are computed using a nonparametric bootstrap procedure, in which new distributions of paired third-party reported net wealth and taxable net wealth are drawn with replacement from the true distribution. The standard error of each estimate is the standard deviation of the distribution of the $\hat{b}$':s. Significance codes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: Estimation results from regressing sentenced fines on downward self-adjustments using, Heckman selection model, 1999-2006. Dependent variable: penalties.

	(1)		(2)		(3)	
	Main	Selection	Main	Selection	Main	Selection
Self-reported	-0.006	$3.02 \times 10^{-9***}$	-0.005	7×10^{-11}	-0.005	4.1×10^{-10}
wealth	(0.004)	(6.40×10^{-10})	(0.004)	(2.18×10^{-9})	(0.005)	(2.29×10^{-9})
I (Downward-adjustment)	0.019^{**}	$3.01 \times 10^{-9***}$	0.016^{*}	$-6.84 \times 10^{-9*}$	0.016	7.07×10^{-9}
× Self-reported wealth	(0.010)	(6.4×10^{-10})	(0.010)	3.98×10^{-9}	(0.010)	(4.29×10^{-9})
Third party			0.001	$2.65 \times 10^{9***}$	0.001	$2.53 \times 10^{9***}$
wealth			(0.001)	(6.9×10^{-9})	(0.001)	(7.3×10^{-9})
Labor					0.040	$8.44 \times 10^{-6***}$
earnings					(2.152)	(1.397×10^{-6})
Fraction Caught		481.1^{***}		428.6^{***}		413.6^{***}
		(51.2)		(56.2)		(56.4)
Observations	1,096,067	1,096,067	1,096,067	1,096,067	1,096,067	1,096,067

Notes: The table shows Maximum likelihood estimates of a Heckman selection equation. Columns labelled *Main* represent main equation estimates and columns labelled *Selection*Â present results from the selection equation. The variable *Fraction Caught* is the instrument used in the selection equation and represents the fraction of tax payers caught for tax evasion in the municipality of individual i (not counting individual i). The variable self-reported wealth denotes self-reported wealth adjustment of taxable net wealth. I (Downward-adjustment) is an indicator of self-reported downward adjustment. All regressions include the following controls: I (Downward-adjustment). Addition control used in columns (2) and (3) is age. Standard errors clustered at the household level. Significance codes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5: Tax semi-elasticities of taxable net wealth, $\varepsilon_{W,\tau}$, 1999-2006.

			Small Kink Analysis	Static Model	Dynamic Model		IV
					$\sigma = 0.25$	$\sigma = 1.1$	
Method	Sample	$\hat{b}$	(1)	(2)	(3)	(4)	(5)
Parametric	All	0.53	-0.118	-0.18	-0.203	-0.134	
		1.25	-0.278	-0.377	-0.339		
	High	1.1	-0.244				
		0.97	-0.216				
	Low	0.42	-0.093				
		0.21	-0.047				
Nonparametric	All	0.49	-0.109	-0.172	-0.195	-0.121	
		0.95	-0.211	-0.300	-0.293	-0.259	
		1.45	-0.320	-0.420	-0.395	-0,389	
	High	0.89	-0.199				
		1.71	-0.380				
		2.26	-0.502				
	Low	0.54	-0.120				
		1.13	-0.250				
		1.44	-0.320				
IV							
$x = 1$							-0.799***
							(0.070)
$x = 2$							-0.539***
							(0.079)
$x = 3$							-0.452***
							(0.101)

Notes: The bunching estimates, $\hat{b}$, are obtained using either the parametric method described in Section 2.4.1 or the non-parametric method described in Section 2.4.2. The estimates are estimated on the entire sample (All) or the subsamples comprising individuals with high cognitive ability (High) and low cognitive ability (Low). The two entries for each subsample using the parametric approach denote the benchmark estimate and the estimate obtained using the extrapolation method, respectively. The three entries for each subsample using the non-parametric approach denote the benchmark estimate, the results for a broader bunching interval and the estimate obtained using the extrapolation method, respectively. Columns (1)-(4) display the estimate implied by the indicated method/model. Where applicable, $\beta = 0.95$ and $p = 0.86$. The IV-estimates in column (5) display the results from estimating equation (2.25) by 2SLS and x denotes lag length. In these estimations, the standard errors, displayed within parenthesis, are clustered at the household level. Significance codes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6: Tax semi-elasticities and elasticities of evasion, $\varepsilon_{e,\tau}$, 1999-2006.

		Static Model		Dynamic Model			
				Uncompensated		Compensated	
		Semi-elasticity	Elasticity	Semi-elasticity	Elasticity	Semi-elasticity	Elasticity
Method	$\hat{b}$	(1)	(2)	(3)	(4)	(5)	(6)
Parametric	0.53	104.370	1.566	81.269	1.219	81.765	1.227
	1.25	90.241	1.354	76.659	1.150	77.155	1.157
Non-parametric	0.49	105.66	1.585	81.567	1.224	82.064	1.231
	0.95	94.760	1.421	78.424	1.176	78.920	1.184
	1.45	87.797	1.317	75.582	1.134	76.077	1.141

Notes: The bunching estimates, $\hat{b}$, are obtained using either the parametric method described in Section 2.4.1 or the non-parametric method described in Section 2.4.2. The two entries using the parametric approach denote the benchmark estimate and the estimate obtained using the extrapolation method, respectively. The three entries using the non-parametric approach denote the estimate obtained when using a window of SEK 40,000 on the 1999-2006 sample, the estimate obtained when using a window of SEK 100,000 on the 1999-2006 sample and the estimate obtained when using the extrapolation method, respectively. Columns (1)-(6) display the estimate implied by the indicated model. Where applicable, $\beta = 0.96$, $\sigma = 0.25$ and $p = 0.86$.

Table 2.7: Estimation results from regressing downward adjustments of taxable net wealth on third-party reported net wealth using OLS, 1999-2006. Dependent variable: Downward adjustment.

	(1)	(2)
Third-party Net Wealth	0.837*** (0.051)	0.522*** (0.109)
Third-party Net Wealth Squared	3.47×10^{-12}** (1.49×10^{-12})	2.60×10^{-11}*** (6.84×10^{-11})
Third-party Net Wealth Cubed		-3.65×10^{-22}*** (1.01×10^{-22})
Year FE	Yes	Yes
Observations	4,835,932	4,835,932

Notes: Downward adjustment is defined as downward-adjusted self-reported taxable net wealth. The standard errors, displayed within parenthesis, are clustered at the household level. Significance codes: ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.8: Estimation results from regressing bunching and downward adjustments of taxable net wealth on cognitive skills, 1999-2006.

Dependent variable:	Bunching				Log Downward Adjustment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cognitive Skills	0.007*** (0.002)	0.006*** (0.003)	0.014*** (0.003)	0.011*** (0.004)	0.149*** (0.032)	0.123*** (0.037)	0.225*** (0.039)	0.167*** (0.045)
Education controls	No	Yes	No	Yes	No	Yes	No	Yes
Top percentile excluded	No	No	No	No	Yes	Yes	Yes	Yes
Sample period	99-06	99-06	02-06	02-06	99-06	99-06	02-06	02-06
Observations	36,072	36,072	18,376	18,376	70,347	70,347	38,996	38,996

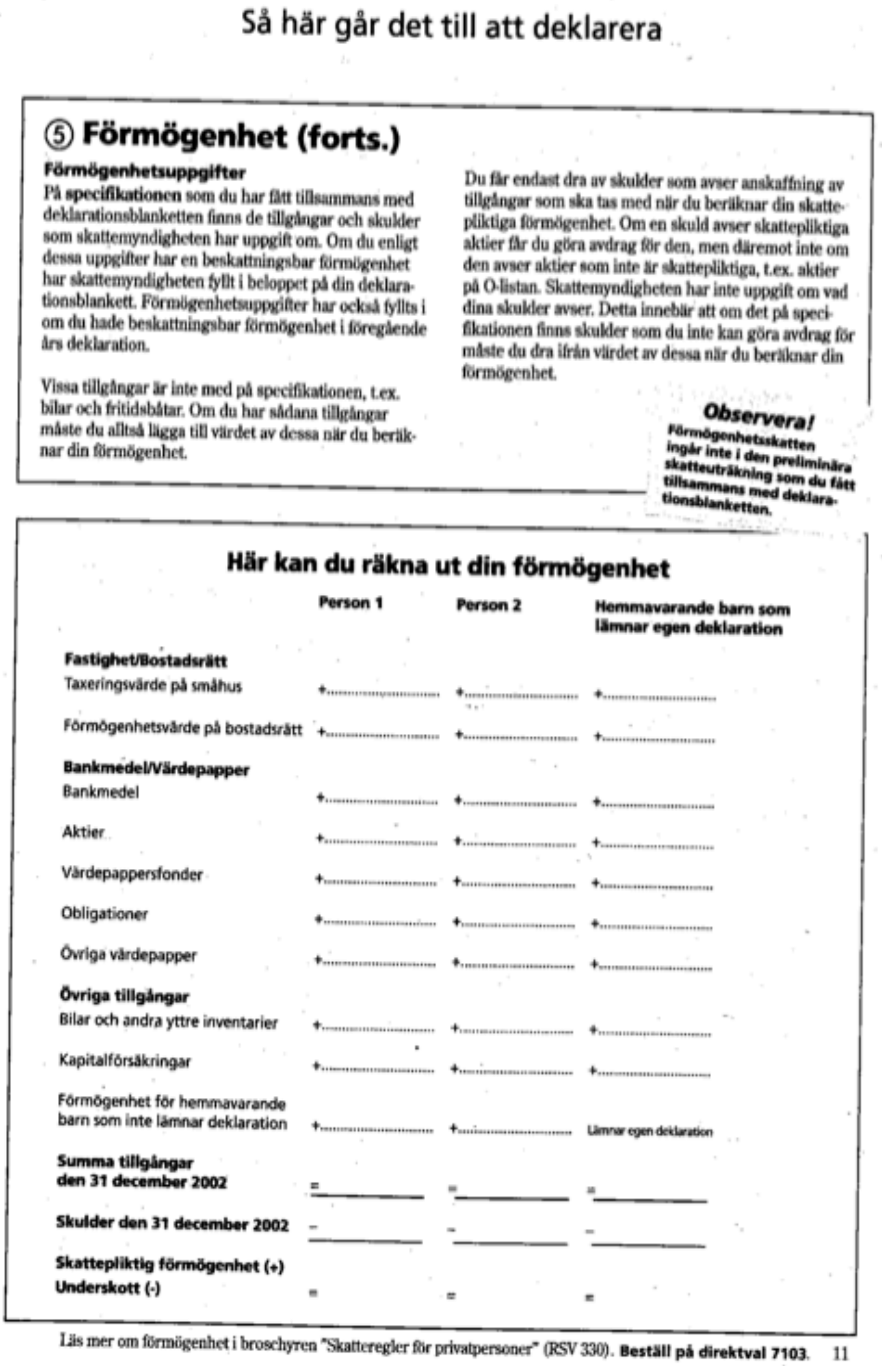
Notes: The bunching variable assumes the value 1 if the household is bunching, zero otherwise. Columns (1)-(4) thus represents the results from estimating a linear probability model on the subsample of households within SEK 500,000 above the kink in third-party reported net wealth. Downward adjustment is defined as downward-adjusted self-reported wealth. Both dependent variables are computed using the extrapolation method. All regressions include controls for wage earnings, a 10-piece linear spline in third-party reported net wealth and type of wealth held in the portfolio. The education controls comprise four dummies representing the highest education obtained: primary school, secondary school, tertiary education and PhD degree. Type of wealth refers to the share of real estate and financial property in the portfolio. Top percentile excluded refers to the exclusion of the 99th percentile of the wealth distribution. Significance codes: ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.9: Estimation results from regressing sentenced fines, risk aversion and impatience on cognitive skills.

Dependent var:	Sentenced Fines		Risk Aversion			Impatience		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cogn. Skills	0.112	0.110	0.120	0.124	0.149	-0.295	-0.291	-0.255
	(0.079)	(0.080)	(0.184)	(0.178)	(0.184)	(0.208)	(0.202)	(0.204)
Log Income	-0.1647***	-0.1653***	-0.122	-0.109	-0.081	-0.207	-0.192	-0.151
	(0.0243)	(0.0244)	(0.078)	(0.079)	(0.086)	(0.211)	(0.201)	(0.222)
Age	0.004	0.004		0.036			0.048**	
	(0.009)	(0.009)		(0.024)			(0.024)	
Divorced					0.653			0.915**
					(0.487)			(0.553)
Year FE	yes	yes	no	no	no	no	no	no
Spline in net wealth	yes	yes	no	no	no	no	no	no
Observations	576	576	153	153	153	153	153	153

Notes: The models in columns (1) and (2) are based on administrative data on sentenced fines on tax evasion and include controls for wealth. The variables risk aversion, impatience and the dummy variable divorced are based on self-reported survey data from Mollerstrom and Seim (2012). Significance codes: ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2.3: Formula for computation of tax liability.



Så här går det till att deklarera

⑤ Förmögenhet (forts.)

Förmögenhetsuppgifter
På **specifikationen** som du har fått tillsammans med deklarationsblanketten finns de tillgångar och skulder som skattemyndigheten har uppgift om. Om du enligt dessa uppgifter har en beskattningsbar förmögenhet har skattemyndigheten fyllt i beloppet på din deklarationsblankett. Förmögenhetsuppgifter har också fyllts i om du hade beskattningsbar förmögenhet i föregående års deklaration.

Vissa tillgångar är inte med på specifikationen, t.ex. bilar och fritidsbåtar. Om du har sådana tillgångar måste du alltså lägga till värdet av dessa när du beräknar din förmögenhet.

Du får endast dra av skulder som avser anskaffning av tillgångar som ska tas med när du beräknar din skattepliktiga förmögenhet. Om en skuld avser skattepliktiga aktier får du göra avdrag för den, men däremot inte om den avser aktier som inte är skattepliktiga, t.ex. aktier på O-listan. Skattemyndigheten har inte uppgift om vad dina skulder avser. Detta innebär att om det på specifikationen finns skulder som du inte kan göra avdrag för måste du dra ifrån värdet av dessa när du beräknar din förmögenhet.

Observera!
Förmögenhetsskatten ingår inte i den preliminära skatteuträkning som du fått tillsammans med deklarationsblanketten.

Här kan du räkna ut din förmögenhet

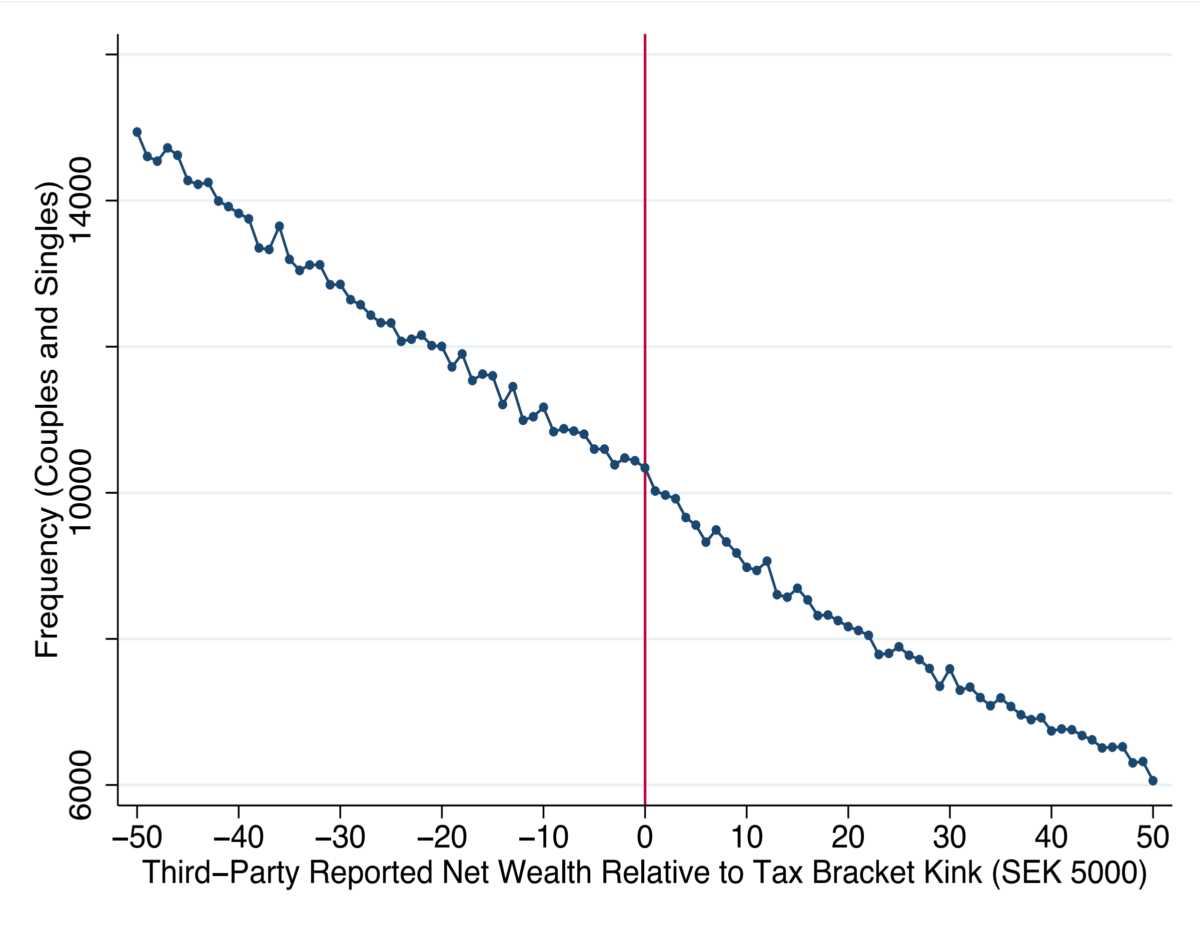
	Person 1	Person 2	Hemmavarande barn som lämnar egen deklaration
Fastighet/Bostadsrätt			
Taxeringsvärde på småhus	+	+	+
Förmögenhetsvärde på bostadsrätt	+	+	+
Bankmedel/Värdepapper			
Bankmedel	+	+	+
Aktier	+	+	+
Värdepappersfonder	+	+	+
Obligationer	+	+	+
Övriga värdepapper	+	+	+
Övriga tillgångar			
Bilar och andra yttre inventarier	+	+	+
Kapitalförsäkringar	+	+	+
Förmögenhet för hemmavarande barn som inte lämnar deklaration	+	+	Lämnar egen deklaration
Summa tillgångar den 31 december 2002	=	=	=
Skulder den 31 december 2002	–	–	–
Skattepliktig förmögenhet (+) Underskott (-)	=	=	=

Läs mer om förmögenhet i broschyren "Skatteregler för privatpersoner" (RSV 330). **Beställ på direktval 7103.** 11

Notes: This formula was appended to the prepopulated tax return. Households were supposed to use this to compute wealth tax liabilities.

Figure 2.4: Third-party reported net wealth distribution around the threshold.

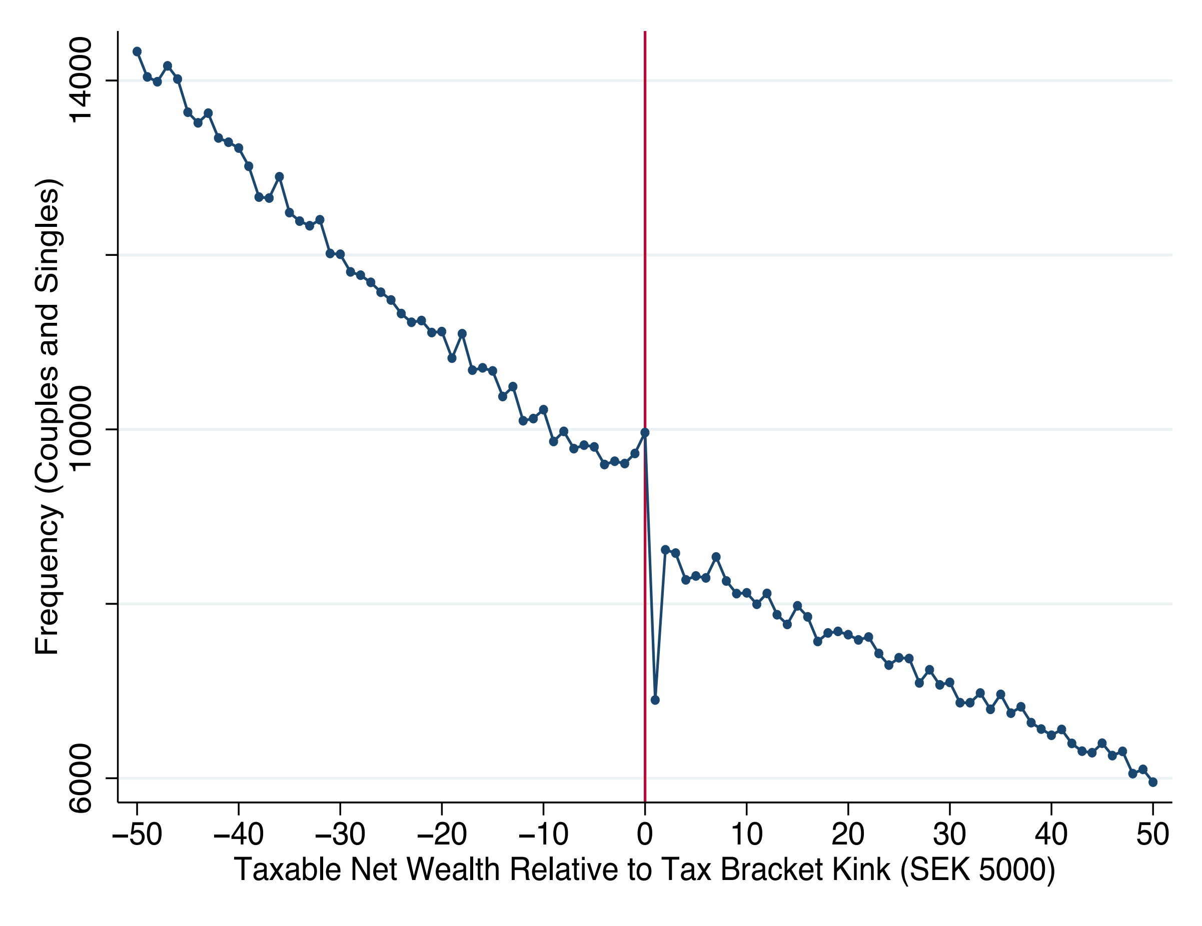
(a)



Notes: The figure shows the distribution of third-party reported net wealth around the shift in the tax brackets, demarcated by the vertical at 0, for the years 1999-2006. The dotted series consist of a histogram relative to the normalized kink point. Each bin corresponds to the number of households within SEK 5,000.

Figure 2.5: Taxable net wealth distribution around the threshold.

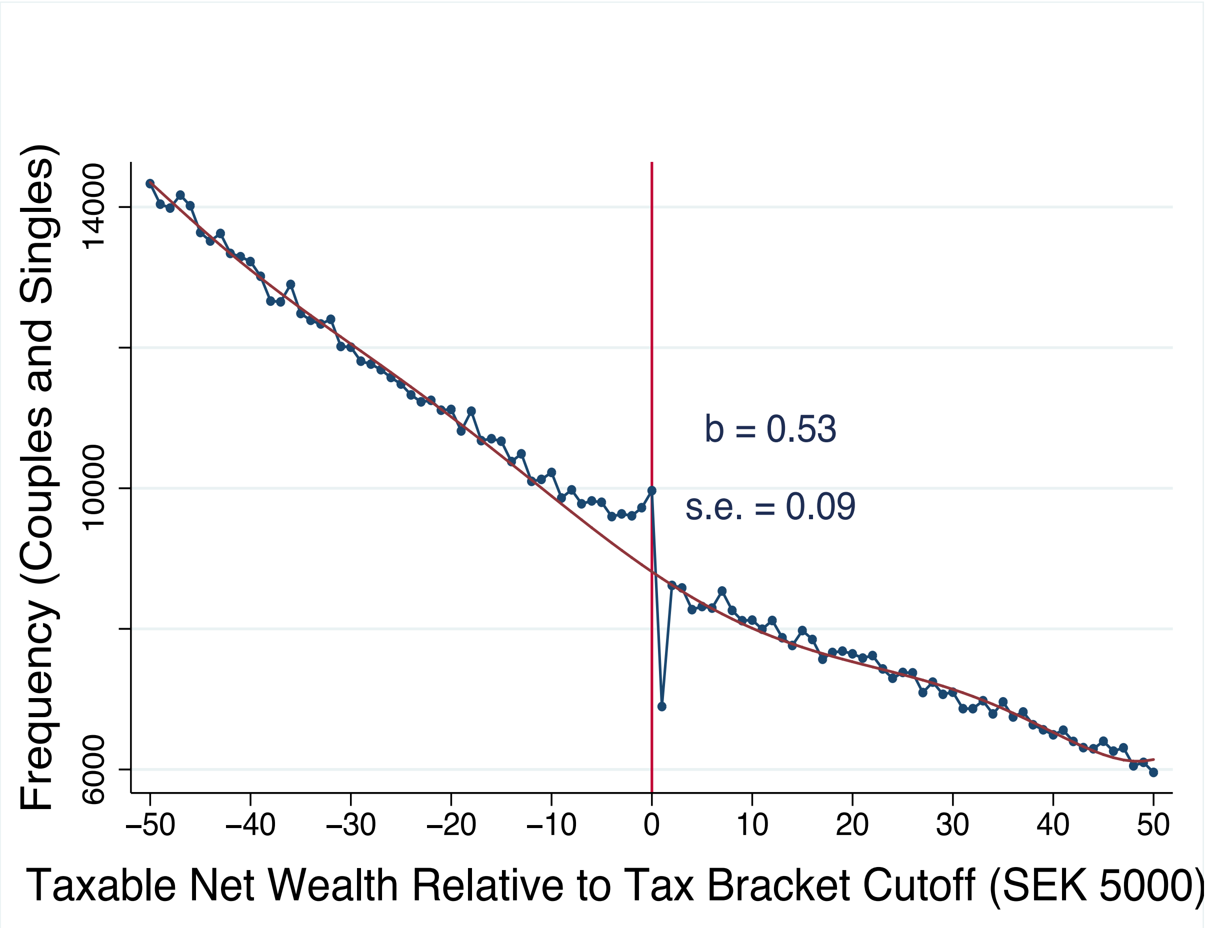
(a)



Notes: The figure shows the distribution of taxable net wealth around the shift in the tax brackets, demarcated by the vertical at 0, for the years 1999-2006. The dotted series consist of a histogram relative to the normalized kink point. Each bin corresponds to the number of households within SEK 5,000.

Figure 2.6: Estimated bunching of taxable net wealth at the threshold.

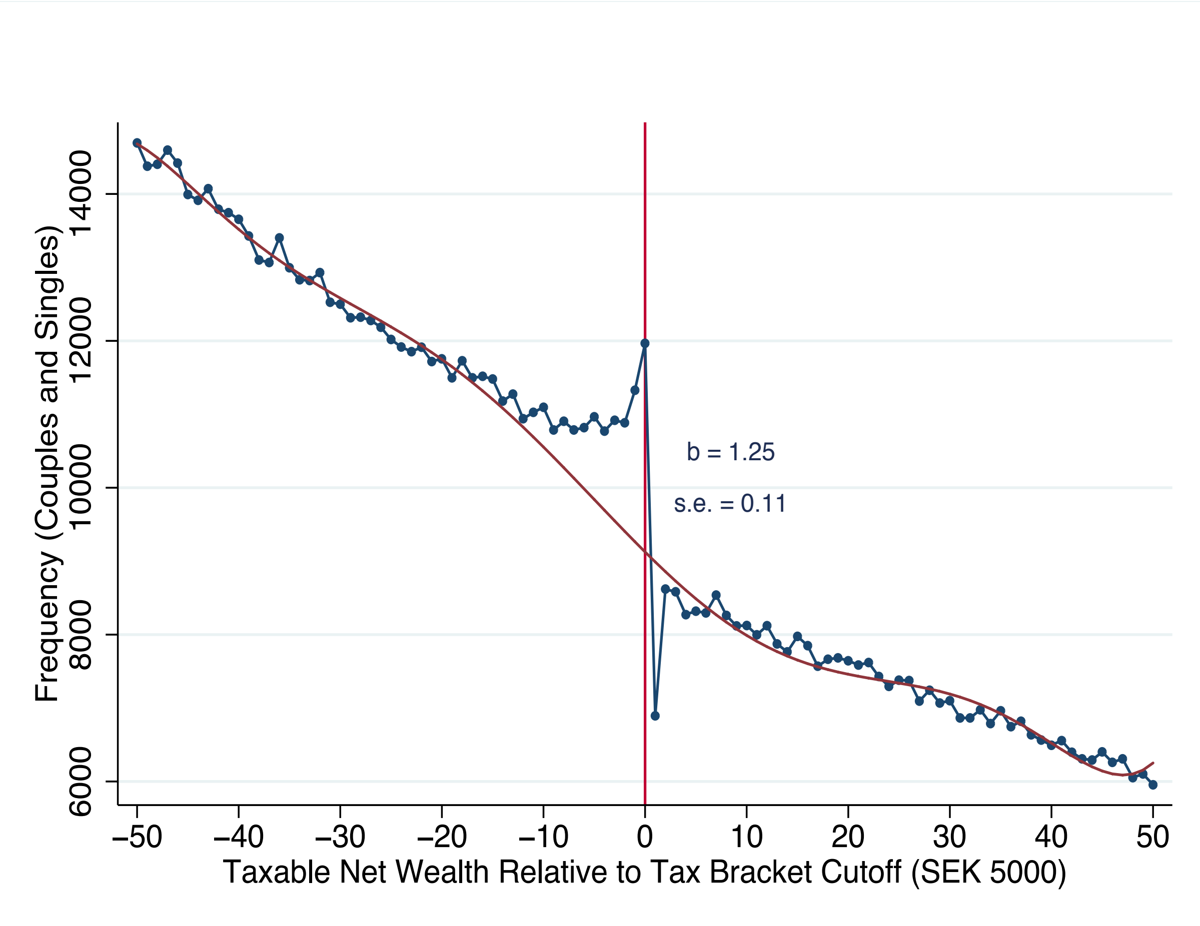
(a)



Notes: The figure replicates Figure 2.5 and adds the estimated counterfactual density, displayed by the solid line in red. The counterfactual density was obtained by fitting a seven-degree polynomial to the density, excluding points within SEK 40,000 below the kink. b denotes the estimated excess mass.

Figure 2.7: Estimated bunching of taxable net wealth at the threshold, extrapolation method.

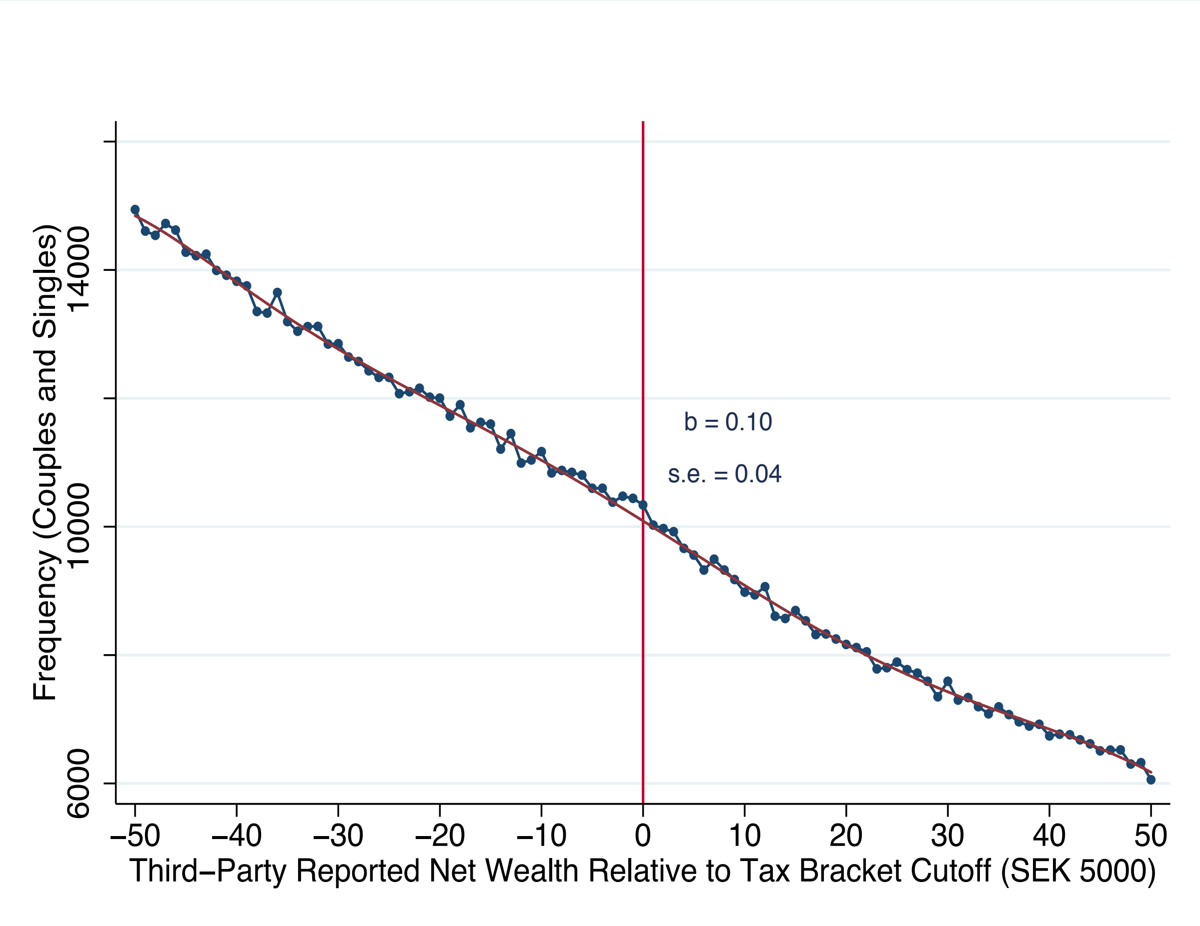
(a)



Notes: This figure replicates Figure 2.6 but uses the extrapolation method – which is described in the text – to deal with missing taxable net wealth values.

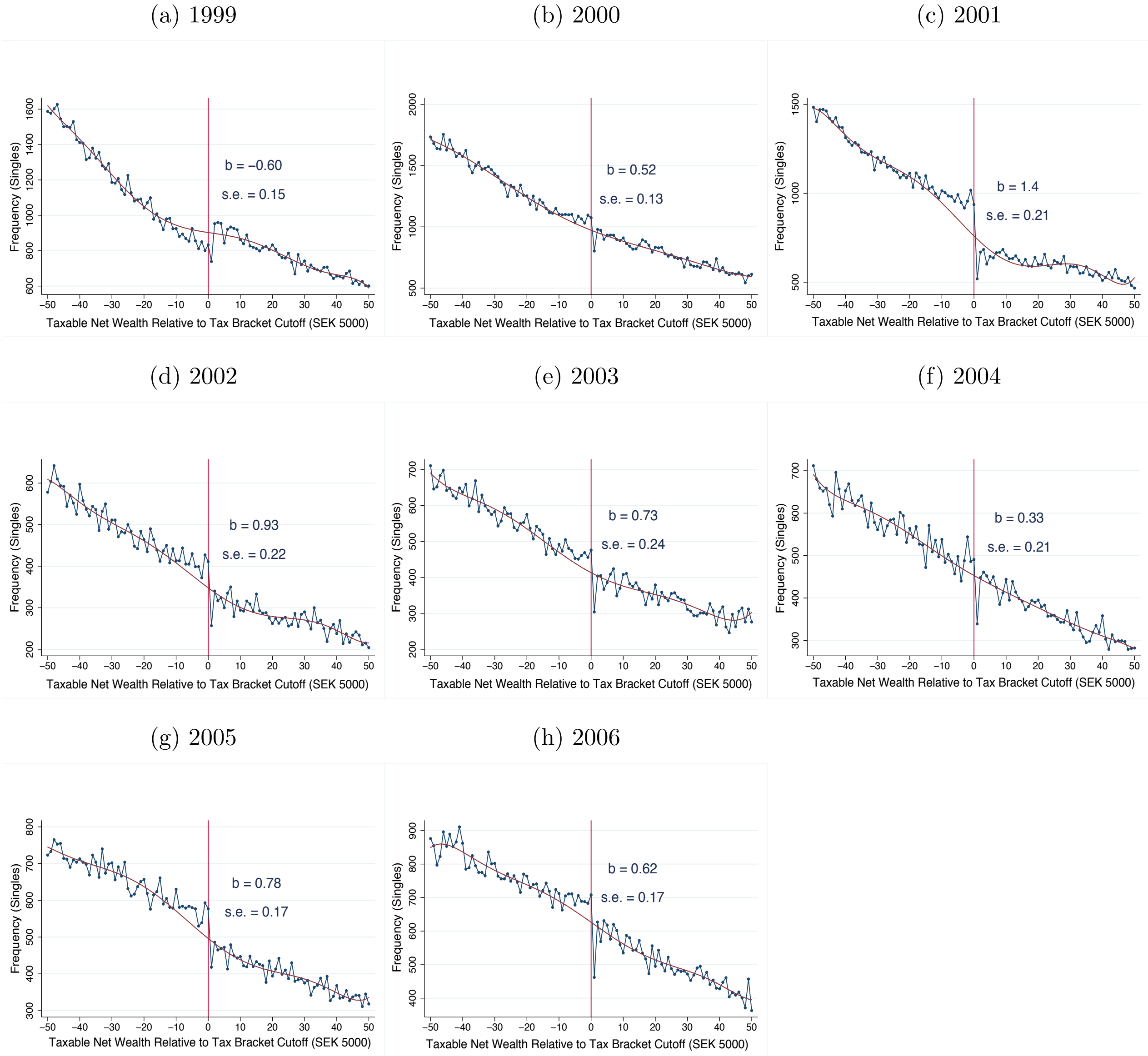
Figure 2.8: Estimated bunching of third-party reported net wealth at the threshold.

(a)



Notes: The figure replicates Figure 2.4 and adds the estimated counterfactual density, displayed by the solid line in red. The counterfactual density was obtained by fitting a seven-degree polynomial to the density, excluding points within SEK 40,000 below the kink. b denotes the estimated excess mass.

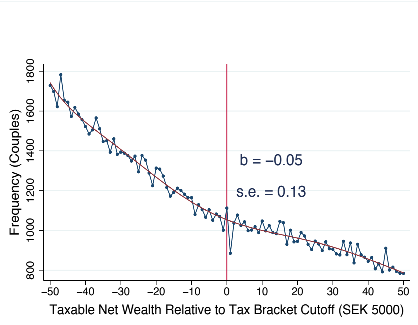
Figure 2.9: Taxable net wealth around the threshold 1999-2006, singles.



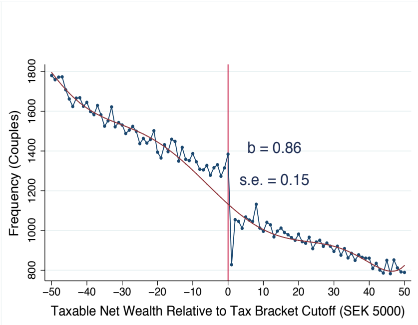
Notes: These figures plot the empirical distribution of taxable net wealth for single households around the kink point in each year from 1999-2006. The vertical line denotes the location of the threshold. The counterfactual density, graphed by the solid curve in red, was obtained as in Figure 2.6.

Figure 2.10: Taxable net wealth around the threshold 1999 – 2006, couples.

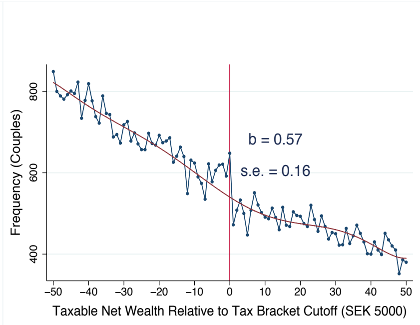
(a) 1999



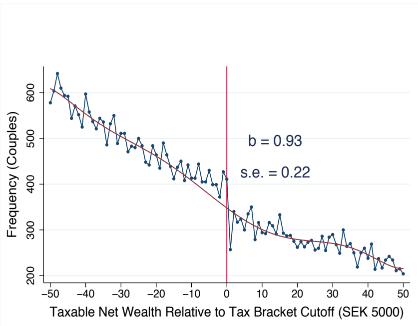
(b) 2000



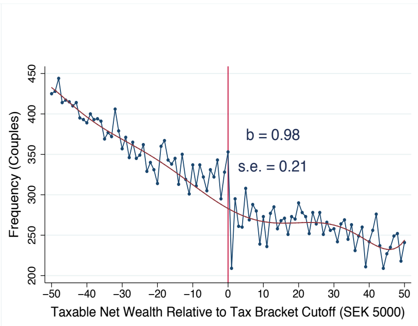
(c) 2001



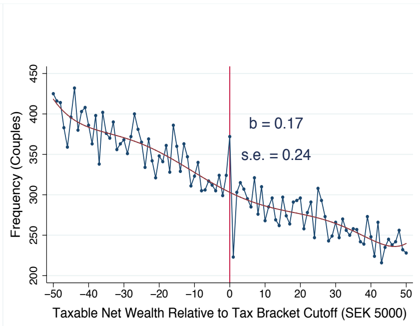
(d) 2002



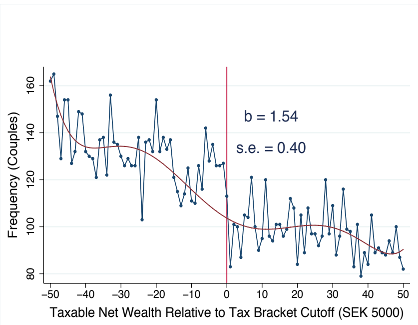
(e) 2003



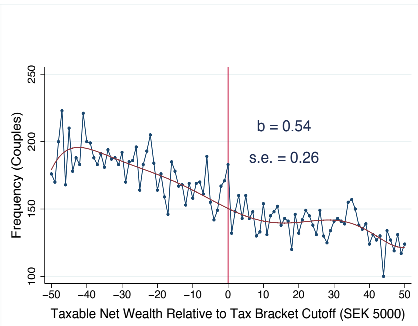
(f) 2004



(g) 2005



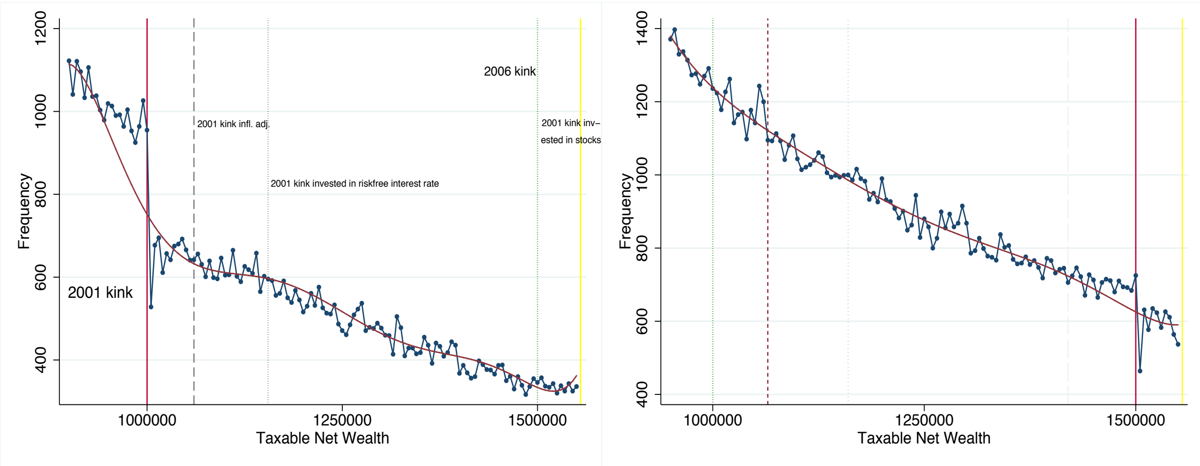
(h) 2006



Notes: These figures plot the empirical distribution of taxable net wealth for couples without children, who file jointly, around the kink point in each year from 1999-2006. The vertical line denotes the location of the threshold. The counterfactual density, graphed by the solid curve in red, was obtained as in Figure 2.6.

Figure 2.11: Does bunching track the tax? Bunching in 2001 and 2006.

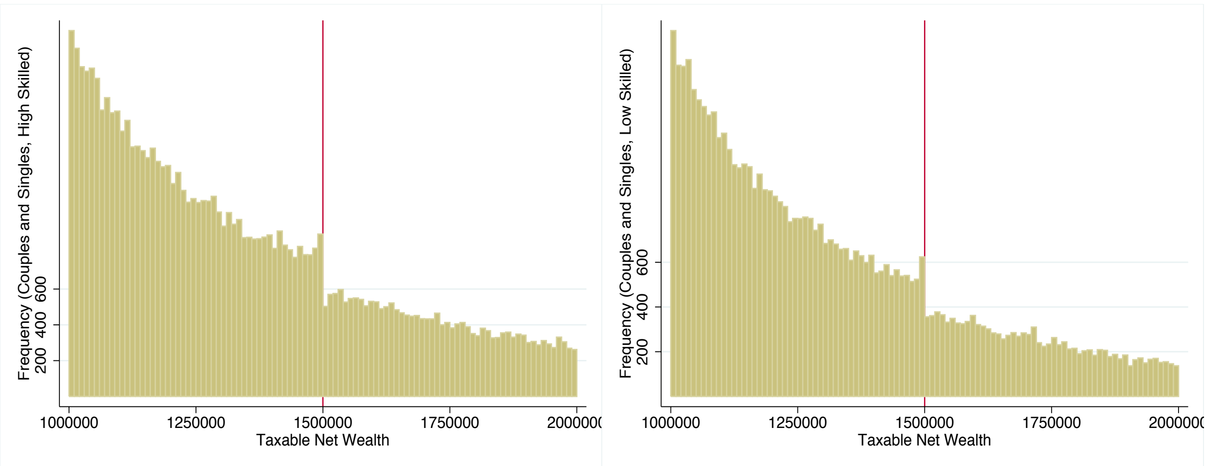
(a) 2001 (b) 2006



Notes: These figures present the taxable net wealth distribution for singles in 2001 and 2006. The figure shows the kinks in 2001 and 2006, located at SEK 1 million and SEK 1.5 million, respectively. The additional vertical lines represent the position of the 2001-kink if it followed - from the left to the right - inflation, the riskfree interest rate or a stock market index return, respectively. The inflation data was obtained from Statistics Sweden, the riskfree interest rate and the stock market index return from Sveriges Riksbank.

Figure 2.12: Taxable net wealth distributions, by cognitive skills.

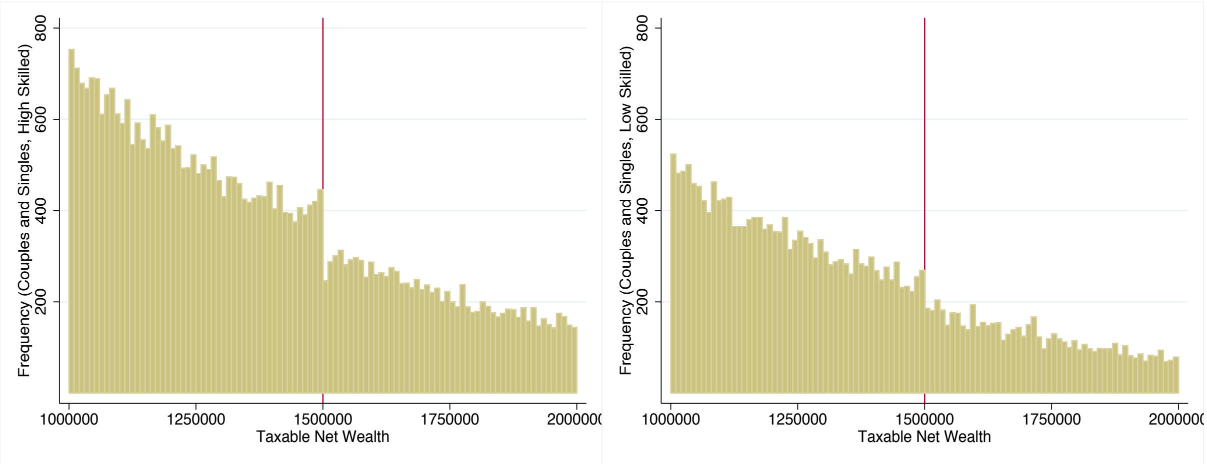
(a) High Skilled (b) Low Skilled



Notes: These figures show the distribution of taxable net wealth around the kink point, demarcated by the vertical at SEK $1,500,000$, for the years 1999-2006. Wealth distributions are recentered each year by family status to obtain the recentered kink. Each bin corresponds to the number of households within SEK $10,000$. High skilled is defined as having a positive z-score. For couples, the skill variable measures the z-score of the male.

Figure 2.13: Taxable net wealth distributions, by cognitive skills, 2002-2006.

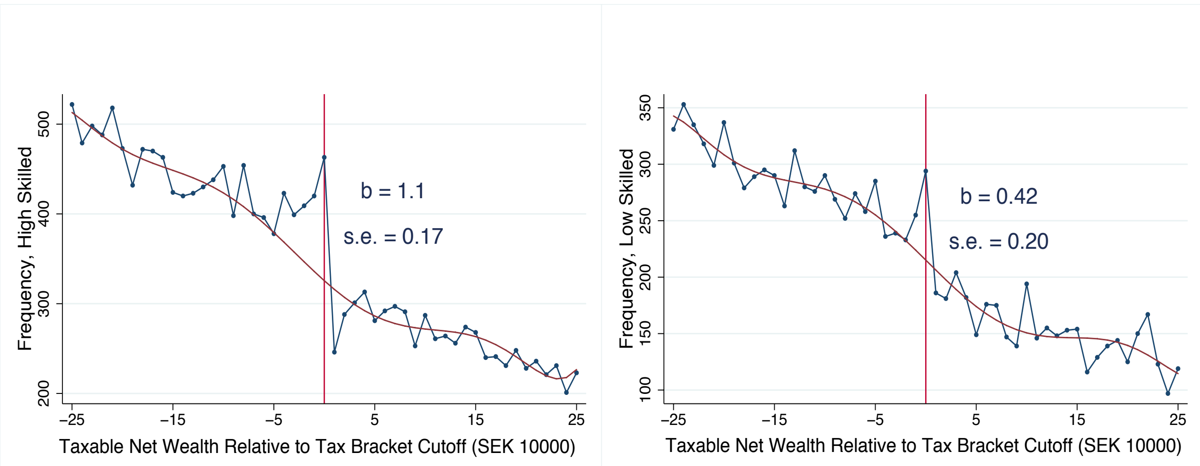
(a) High Skilled (b) Low Skilled



Notes: These figures replicate Figure 2.12 for the years 2002-2006.

Figure 2.14: Estimated bunching of taxable net wealth at the threshold, by skill groups.

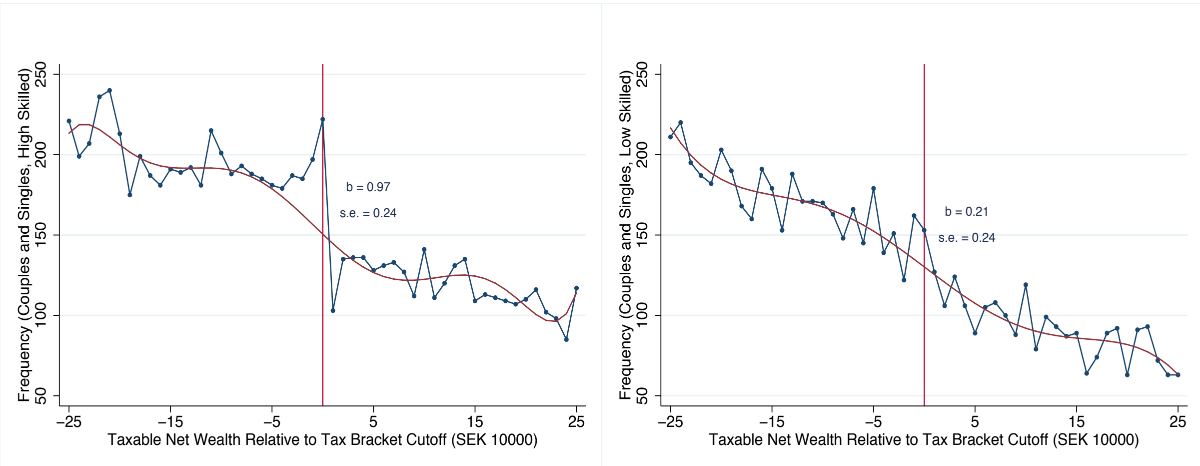
(a) High Skilled (b) Low Skilled



Notes: These figures show the distribution of taxable net wealth close to the kink point in the years 2002-2006, by skill groups. They add the estimated counterfactual density, displayed by the solid lines in red. The counterfactual densities were obtained by fitting a seven-degree polynomial to the densities, excluding points within SEK $40,000$ below the kink. b denotes the estimated excess mass. High skilled are defined as having a positive z-score. For couples, the skill variable measures the z-score of the male.

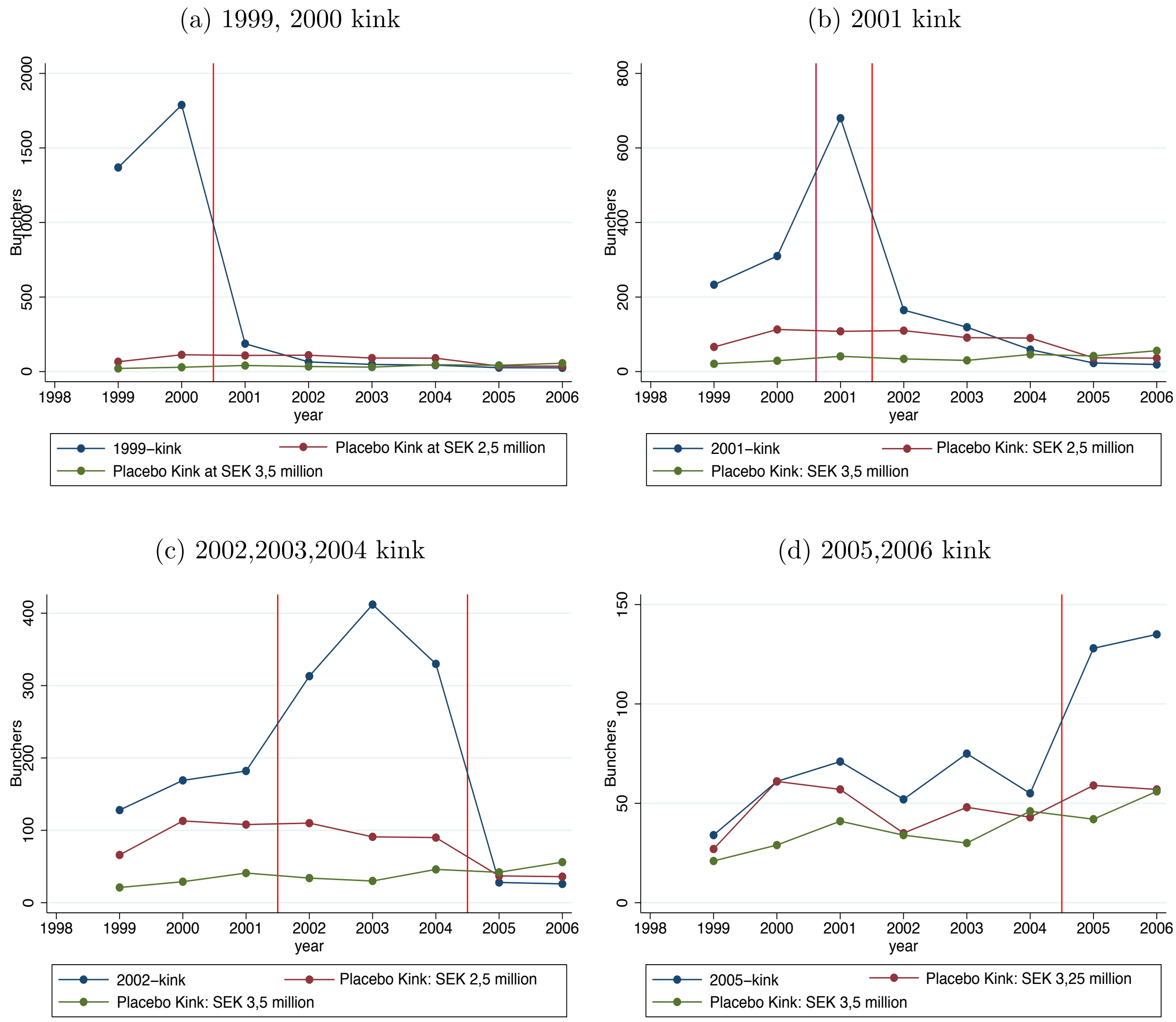
Figure 2.15: Estimated bunching of taxable net wealth at the threshold, by skill groups, alternative measure.

(a) High Skilled (b) Low Skilled



Notes: These figures replicate the plots in Figure 2.14 but change the skill group definition. High skilled is defined as having a score above the 75th percentile in the skill distribution. Low skilled below 25th. For couples, the skill variable measures the z-score of the male.

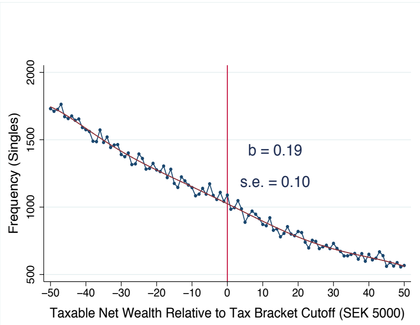
Figure 2.16: Bunching over time - couples.



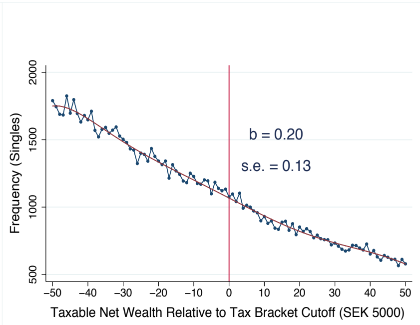
Notes: These graphs show bunching – computed by the nonparametric method – over time at different kinks for couples. In (a), bunching at the threshold of SEK $900,000$, which was the threshold in 1999 and 2000, with a window of SEK $25,000$ is displayed. (b), (c) and (d) present similar graphs for bunching at thresholds corresponding to the years 2001, 2002-2004 and 2005-2006, respectively. All graphs additionally plot bunching at placebo kinks of SEK 2.5 million and SEK 3.5 million. All bunching estimates are computed using the extrapolation method.

Figure 2.17: Third-party reported net wealth around threshold 1999-2006, singles.

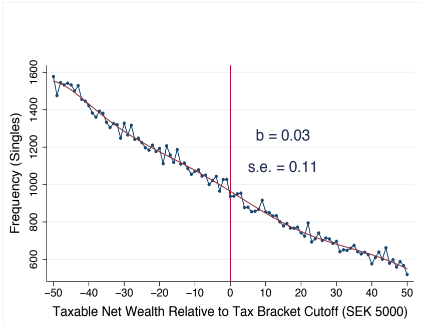
(a) 1999



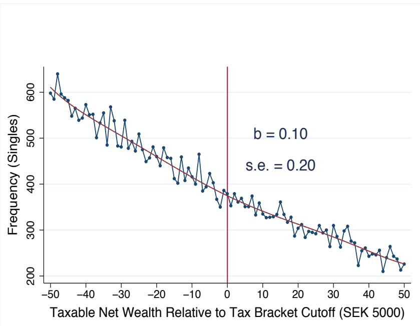
(b) 2000



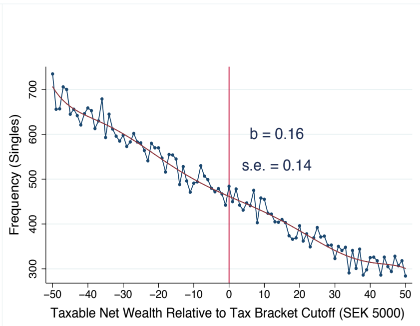
(c) 2001



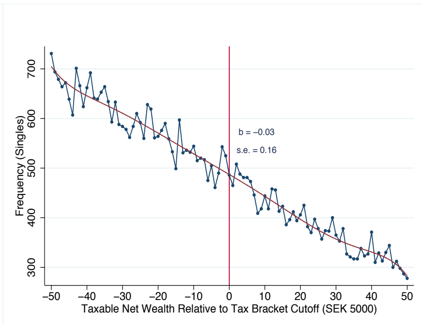
(d) 2002



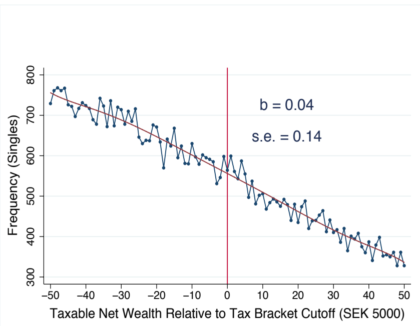
(e) 2003



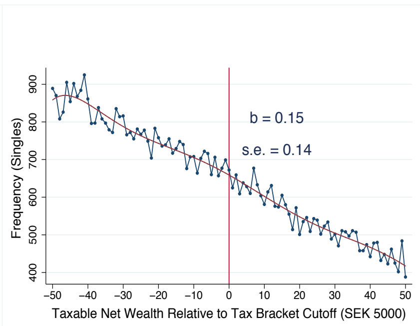
(f) 2004



(g) 2005



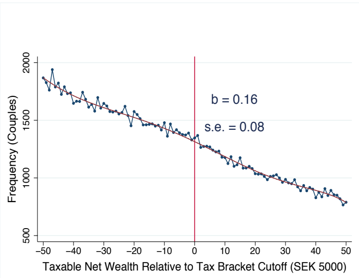
(h) 2006



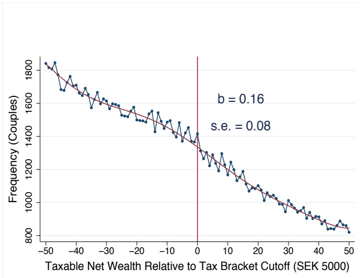
Notes: These figures plot the empirical distribution of third-party reported net wealth for single households around the kink point in each year from 1999-2006. The vertical line denotes the location of the threshold. The counterfactual density, graphed by the solid curve in red, was obtained as in Figure 2.6.

Figure 2.18: Third-party reported net wealth around the threshold 1999-2006, couples.

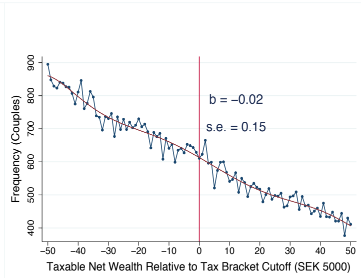
(a) 1999



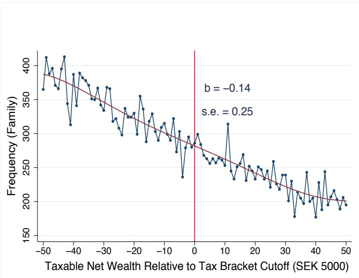
(b) 2000



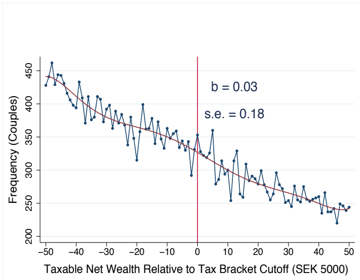
(c) 2001



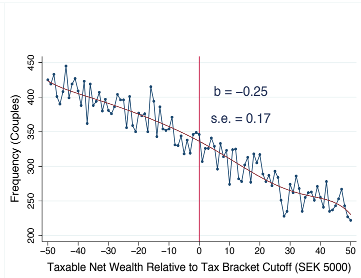
(d) 2002



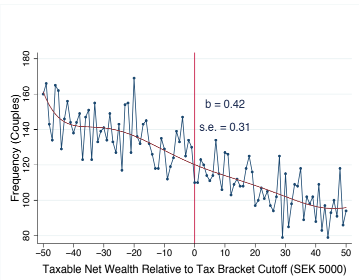
(e) 2003



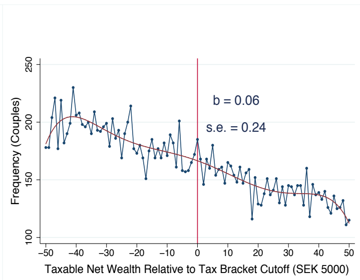
(f) 2004



(g) 2005

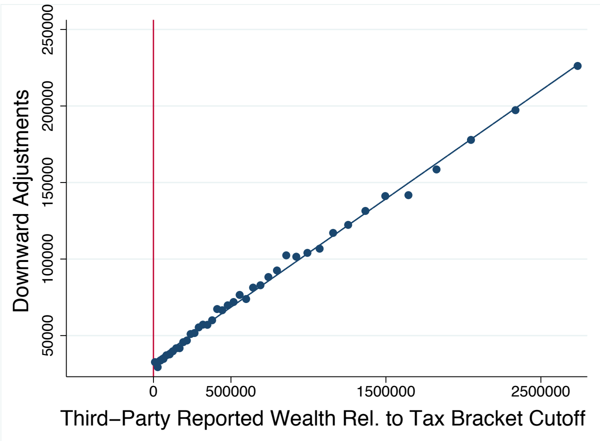


(h) 2006



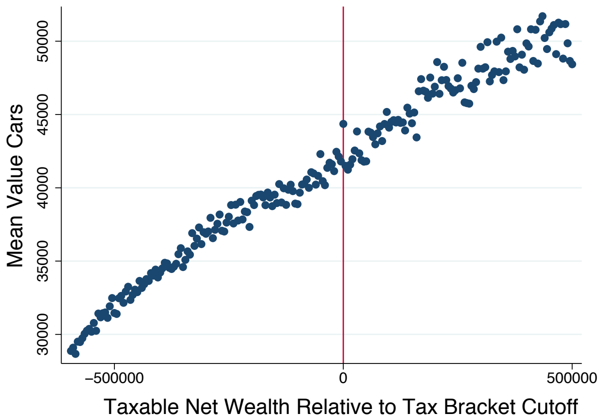
Notes: These figures plot the empirical distribution of third-party reported net wealth for couples without children who file the tax jointly, around the kink point in each year from 1999-2006. The vertical line denotes the location of the threshold. The counterfactual density, graphed by the solid curve in red, was obtained as in Figure 2.6.

Figure 2.19: Downward adjustment and third-party reported net wealth.



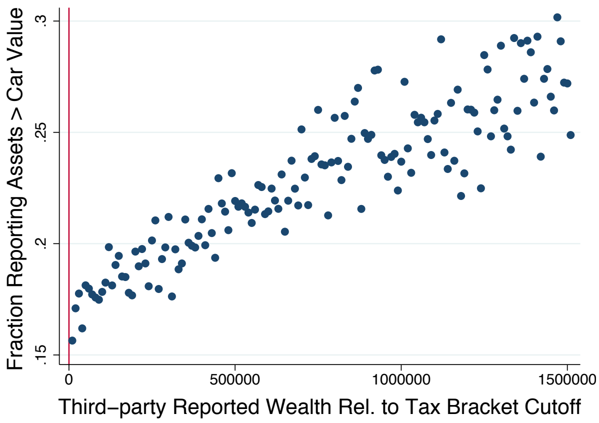
Notes: The figure shows mean of self-reported downward adjustments of taxable net wealth against third-party reported wealth. Third-party reported wealth is normalized with zero denoting the exemption threshold.

Figure 2.20: Mean value of cars around threshold.



Notes: The figure shows mean value of car holdings around the threshold (normalized to zero). Values of cars were assigned using information on the car brand, model and vintage, together with prices on new cars from the Swedish Tax Authorities and a devaluation model which is based on actual purchases of used cars, collected by http://www.bilpriser.nu.

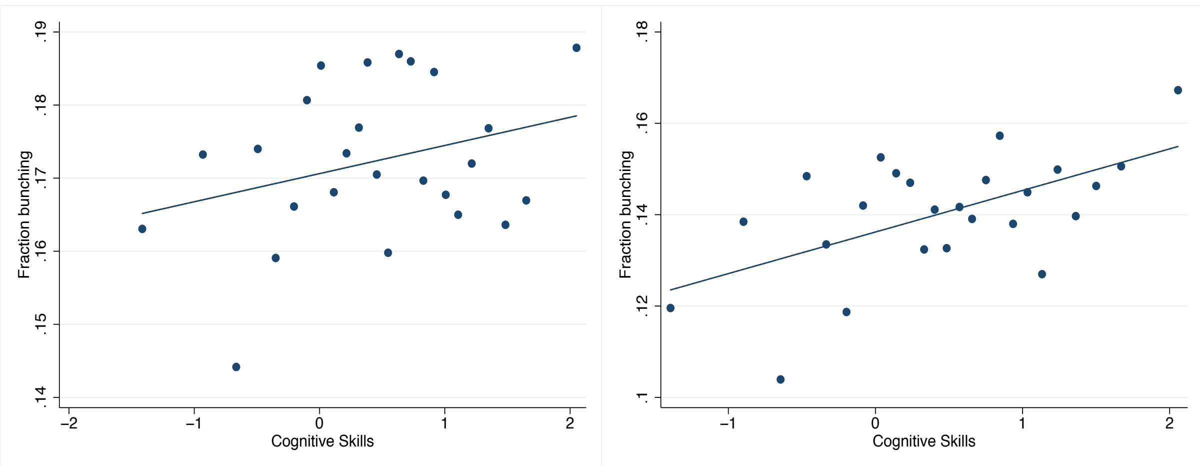
Figure 2.21: Fraction of cars that are self-reported.



Notes: The graph shows the fraction of car owners whose self-reported assets are at least as large as the car holdings – computed using register data on car holdings and a pricing scheme discussed in Figure **??** – to the right of the kink, which is normalized at zero.

Figure 2.22: Bunching and cognitive skills by quantile.

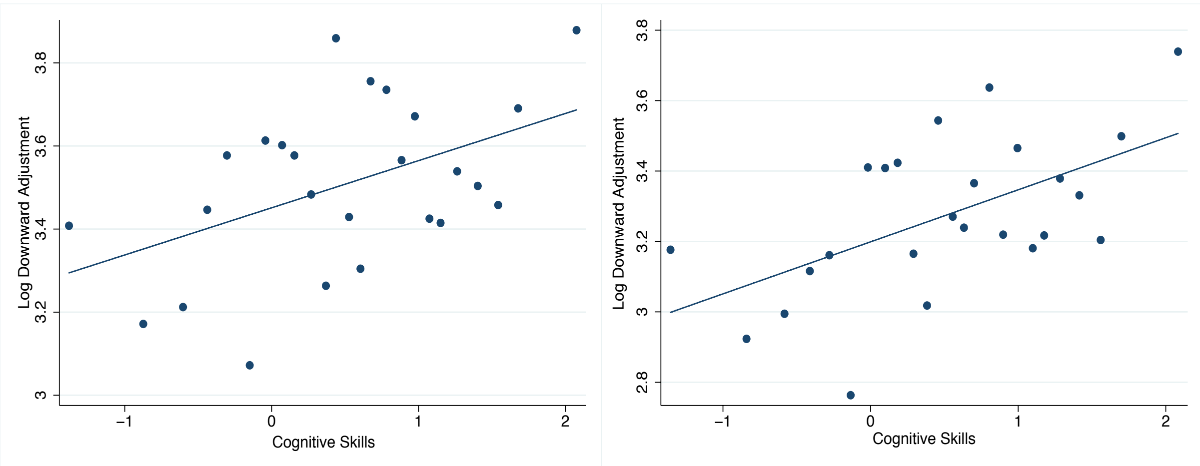
(a) 1999-2006 (b) 2002-2006



Notes: The scatter plots present residuals, after regressing an indicator variable of bunching on a 10-piece spline in net wealth, wage earnings, education, share real estate and share financial property of the portfolio, on the y-axis, against cognitive skills on the x-axis. The sample consists of households above the kink but within SEK $500,000$ above the kink. Each quantile comprises 4 percent of the sample.

Figure 2.23: Log downward adjustment and cognitive skills by quantile.

(a) 1999-2006 (b) 2002-2006



Notes: The scatter plots present residuals, after regressing self-reported downward adjustments of taxable net wealth on a 10-piece spline in net wealth, wage earnings, education, share real estate and share financial property of the portfolio, on the y-axis, against cognitive skills on the x-axis. The sample consists of households above the kink but within SEK 500, 000 above the kink. Each quantile comprises 4 percent of the sample.

Chapter 3

Job Displacement and Labor Market Outcomes by Skill Level*

3.1 Introduction

A large body of literature has established beyond reasonable doubt that the individual costs of job displacement are large and long-lasting. Earlier analyses cover the impact on labor market outcomes such as earnings and unemployment (Ruhm, 1991, Jacobson et al., 1996, Stevens, 1997, Kuhn, 2002, Farber, 2003a, Eliason and Storrie, 2006, Couch and Placzek, 2010), health and mortality (Sullivan and von Wachter, 2009, Eliason and Storrie, 2009), and the school grades of displaced workers' children (Rege et al., 2010). Jacobson et al. [1996] show that annual earnings fall dramatically upon job loss, and that the earnings losses converge to 25 percent of the predisplacement earnings after five years. Sullivan and von Wachter [2009] present similar patterns for mortality, with striking increases in death rates upon displacement that fall over time but never fade away entirely. Despite extensive research into the consequences of job loss, our understanding of who becomes displaced and how different individuals deal with displacement is far from complete. To identify adequate policy measures for assisting displaced workers, it is of utmost importance to understand which groups are most vulnerable.

That ability is crucial for labor market outcomes is well recognized within labor economics.[1] Measuring ability, however, is not straightforward. While proxies for acquired skills include IQ-tests, education and formal training, more general skills are difficult to gauge. A recent literature trying to capture skills distinguishes between cognitive and

*I would like to thank Ann-Christin Jans, at the Swedish Public Employment Service, who made the displacement data available. I have benefited from useful comments from Peter Fredriksson, Anna Larsson Seim, Jaanika Meriküll, Torsten Persson, Tairi Rõõm, David Strömberg and seminar participants at the Bank of Estonia and the European Association of Labour Economists meeting in 2013. I am grateful to Christina Lönnblad for editorial assistance.

[1]See, for instance, Cawley et al. [2001], Herrnstein and Murray [1994] and the summary provided by Bowles et al. [2001].

noncognitive abilities. Heckman et al. [2006] and Lindqvist and Vestman [2011] show that noncognitive abilities, such as persistence, motivation and trustworthiness, are as important as cognitive abilities in explaining labor-market outcomes, such as earnings and wages.

It has also been suggested that the long-term effects of displacements are different for workers with high and low ability. Neal [1998] argues that high-skilled workers recover faster due to their ability to re-accumulate job-specific skills. Previous empirical studies have assessed the importance of education, occupation, age and industry for the impact of displacement, and also for the selection into displacement.[2] But, to the best of my knowledge, nobody has investigated how the impact of displacement varies with cognitive and noncognitive skills, let alone how the incidence of job loss varies with such traits.

In this paper, I attempt to add to our knowledge about displacements and labor market outcomes by using a large dataset of Swedish workers. The main building block of the dataset comprises longitudinal employer-employee matched data on all Swedish individuals in the labor force during any of the years 1999-2009. This dataset, which contains information on basic characteristics such as age and education, is merged with administrative individual level data on nearly all displacements that occurred in the same years. In addition, the dataset contains information on cognitive and noncognitive skills for almost all men, measured at conscription, i.e., when the men were about 18 years of age.

My contribution to the literature is twofold. The first novelty comprises the use of administrative data on job displacements. Previous studies of job displacement either use survey data or matched employer-employee data. The first approach has the advantage of directly asking about displacements but the drawback of self-reports. For example, one caveat with survey data, reported by von Wachter et al. [2009], is that a large fraction of actually displaced workers do not recall that they were indeed displaced. The latter approach, on the other hand, has the advantage of no measurement error (earnings are directly observed), but the disadvantage that displacements cannot be directly observed. Instead, papers in this literature label separations of employer-employee matches as displacements if they occur in times when the firm is downsizing or closing down. However, in Sweden, employers wishing to lay off five or more workers simultaneously, or 20 or more workers within a 90 day period, must report this to the Public Employment Service (henceforth PES). By linking these data to employer-employee matched records at the individual level, it is possible to separate voluntary quits from layoffs, thereby resolving problems in previous studies that use either survey or matched employer-employee data.

Second, Sweden had mandatory conscription until 2010 with the implication that all men around the age of 18 had to undergo two days of physical and mental tests.[3]

[2]For differential effects of job loss, see e.g. von Wachter and Handwerker [2009].

[3]Importantly, military enlistment was mandatory and could only be avoided under very special cir-

The draft procedure included a test of the conscript's cognitive skills with sub-tests of logical, spatial, verbal and technical ability. Moreover, in a semi-structured interview, a psychologist evaluated the noncognitive skills of each conscript. Personality traits that yielded high test scores include persistence, an outgoing character and the willingness to assume responsibility.[4] Merging these data with the matched employer-employee data and the PES records on displacements allows me to estimate both how the incidence of job loss depends on cognitive and noncognitive skills and how the impact of job loss varies with abilities.

With these data, I first investigate how the selection into displacement depends on skills and age. It turns out that workers with high cognitive and noncognitive skills are significantly less likely to experience job loss than workers with low skill levels, also controlling for differences in tenure. An increase of one standard deviation in either skill measure decreases the probability of the worker being displaced by one percentage point. In the same analysis, age is shown to matter for the displacement decision. Younger workers are significantly more likely be displaced when plants downsize, again, controlling for tenure.

In the analysis of displacement effects on economic outcomes, I first estimate the impact on total income. Income is then disaggregated into labor earnings and dependence on social and unemployment insurance. I find that the estimated initial drop in income is not as large as the impact on labor earnings, which is estimated at 23.2 percent of the pre-displacement earnings. Turning to the long-term effects, the estimated labor earnings losses amount to 15 percent of the predisplacement earnings seven years after the displacement.

When studying how ability and age matter for post-displacement patterns, the heterogeneous effects on labor earnings are investigated and I find that the patterns for labor earnings are equal in percentage terms for high-skilled and low-skilled workers. The strongest differential impact for labor earnings is found across age groups. Young workers recover at a much quicker pace than older workers, regardless of the outcome variable under study. To study how skills matter for the willingness to acquire new skills, I look at the heterogeneous effects on student transfers received from the government for adult secondary and tertiary studies as well as participation in job-training programs. I find that cognitively able and young workers are more likely to start studying after displacement, but there is no such pattern for noncognitive skills. Less able workers are more likely to participate in job-training programs, irrespective of the skill measure used.

The final part of the paper addresses the potential bias that might plague previous

---

cumstances. Test scores are therefore not only observed for a selected sample. Moreover, it was not possible to avoid military service by scoring low on the tests.

[4]Lindqvist and Vestman [2011] argue that although the purpose of the interview was to assess conscripts' ability to fit in with the military environment, the traits that were rewarded are also prized by the labor market. See Carlstedt, 2000 for a review of the tests.

studies of effects of job loss that use matched employer-employee data and thus are unable to distinguish quits from layoffs. I use the Swedish employer-employee records and define separations of matches as job displacements according to the literature. Estimates from regressions using conventional approaches are compared with the estimated effects when using the PES data, where displacements are directly observed. I do not only assess the potential bias in previous studies, but also propose how to define separations as displacements in the absence of direct data on job displacements. First, separations in periods of mass separations are defined as displaced, following e.g. Jacobson et al. [1996] and Sullivan and von Wachter [2009]. The mass layoff estimates turn out to significantly understate the effects of displacement. Instead, defining separations in three-year time windows prior to plant closure as displacements leads to an overestimation of the effects. The most accurate estimates are obtained by defining displacements as separations occurring during a one-year window preceding plant closure.

The paper is organized in the following way. Section 3.2 presents the institutional background to the administrative data on displacements and lays out the properties of the data. Section 3.3 studies the selection into displacement, exposing sample restrictions, the econometric framework and results. Section 3.4 goes on to the effects of displacement, including sample restrictions, the econometric framework and the empirical results. Section 3.5 estimates the effects of job loss using the conventional approaches and presents a comparison between the different approaches. Section 3.6 concludes.

3.2 Institutional Background and Data

The first generation of papers studying the effects of job loss used survey data.[5] Eventually, the caveats with survey analyses – e.g. measurement error in self-reported income and in the cause for the separation (Kuhn [2002] discusses the tendency of employers and employees to disagree on the cause in surveys), and attrition bias – lead researchers to address the question using administrative records. In their pioneering study using administrative data, Jacobson et al. [1996] (henceforth JLS) exploited matched employer-employee records from Pennsylvania that cover quarterly earnings histories for a panel of individuals over thirteen years. Approaching the question using employer-employee records is now the established practice in the job displacement literature.[6] However, em-

[5]The Displaced Worker Surveys (DWS) and the Panel Study of Income Dynamics have been extensively employed, see e.g. Topel [1990], Ruhm [1991], Stevens [1997] as well as Farber [1993, 1997, 2001, 2003b, 2007]. The studies that used panel data on income constituted an advancement in the literature by taking pre-displacement incomes into account and, even more importantly, by using methods from the program-evaluation literature with a well-defined control group.

[6]A noteable exception is von Wachter et al. [2009] who link data from the Displaced Worker Survey to administrative data on quarterly earnings. They find that workers who are defined as displaced according to the administrative records do not report being displaced in the survey, but the overall estimates of effects of job loss do not differ depending on the definition. Here, in contrast, I use administrative data

pirical studies that follow this methodology suffer from the inability to distinguish quits from layoffs. When only observing existing matches that are coming to an end, it is not possible to know whether the employer or the employee initiated the separation. In order to resolve this issue, the current convention is to define displacements as separations occurring within certain windows of mass separations. Or, an alternative method is to define workers as displaced if they separated in a time window preceding closures of entire plants. The advantage of the latter approach is that selection into displacement within the plant is smaller compared to mass layoffs, thus making displacement more exogenous with respect to individual characteristics.

I resolve this issue using Swedish administrative data on displacements. Under Sweden's employment-protection law, an employer intending to displace five or more workers simultaneously, or 20 or more workers within a 90-day period, must notify the PES in advance.[7] The notification occurs in two steps: (i) first, the intended number of displaced workers must be reported by the employer to the PES together with the cause for displacement and (ii) no later than one month after the first report, a list must be submitted with the displaced workers and their displacement dates.[8]

With these data, I address the selection into displacement in a given downsizing event. Such an analysis is not possible with matched employer-employee records along with the convention of defining displacements as separations of matches in times of firm distress. Furthermore, instead of estimating the effects of job loss on a subset of the actually displaced or a set of workers which includes voluntary quitters, these data enable me to analyze the costs of job loss on those truly displaced.

To estimate the effects of job loss on earnings and unemployment, I construct a matched dataset with information from various administrative sources. Earnings histories are obtained from employer-employee matched records (*Registerbaserad Arbetsmarknadsstatistik*) stretching from 1999 to 2009 for all Swedish citizens who were part of the labor force in any of the years 2002, 2003 or 2004. This dataset is merged with demographic information on education, occupation, age and gender from the Longitudinal Integration Database for Health Insurance and Labour Market Studies (*LISA*). Unemployment records are obtained from the Unemployment Register (*HÄNDEL*) and merged with the other data.

Finally, I merge these data with individual-level data from the Swedish military enlistment.[9] Officially, Sweden had conscription until 2010, with the implication that all men had to undergo two days of intensive tests of their medical status, physical capacity and

for nearly all job displacements that occurred in 2002, 2003 and 2004, and match it to employer-employee records.

[7]Displacements are reported separately for each plant.

[8]These lists are determined in negotiations with labor unions but are characterized by the "last-in-first-out" principle, which means that the workers who were employed last must leave the firm first when downsizing occurs. This principle is set out in the Swedish Employment Protection Act.

[9]See Carlstedt [2000] and Lindqvist and Vestman [2011] for overviews of the enlistment procedure.

cognitive and noncognitive skills.[10] The tests were usually taken at the age of 18 or 19. The cognitive test consisted of four sub-tests, each comprising 40 tasks presented in increasing order of difficulty. The four tests measured inductive skills, verbal skills, spatial ability and technical comprehension. After the cognitive test, the noncognitive ability of each conscript was evaluated by psychologists. The assessment looked at the conscripts' capacity to match the psychological requirements of the military and rewarded them for traits such as a willingness to assume responsibility, independence, an outgoing character, persistence, emotional stability and power of initiative (see Lindqvist and Vestman [2011]). Like the cognitive test, the noncognitive measure consisted of four sub-tests.

Importantly, all men did the test since it was mandatory by law. This means that there is no selection into enlistment of individuals on unobserved characteristics (gender is observed) which would bias our ability measures. Furthermore, it was not possible to avoid military service by scoring low on the tests.

As the tests changed several times over the years, I rank the individuals' scores by percentile, within each enlistment year. Under the conventional assumption of normally distributed skills, the variables measuring cognitive and noncognitive skills are obtained from the inverse of the normally distributed CDF, producing standard, normally distributed variables. By merging these data with that above, I obtain skill measures for almost all male workers.

3.3 Selection into Displacement

3.3.1 Sample Restrictions and Empirical Framework

It is important to uncover patterns in the incidence of job displacement. An investigation of circumstances surrounding the selection of job displacement is called for, both to make policy suggestions and to understand the labor market.[11] This paper investigates how important general skills are for the selection into displacement. In contrast to previous studies of the incidence patterns, the PES data allow me to identify the selection into displacement within a plant at a given displacement.

To construct an analysis sample, I first merge the PES data to matched employer-employee records at the plant-worker level. The PES data have three identifiers: displacement event id, plant id and worker id (the equivalent of a scrambled social security number). In some instances, the plant id is missing from the records. For these dis-

[10] In practice, mandatory enlistment was gradually abandoned, implying that the individuals taking the tests during the 2000s consisted increasingly of men and women who were motivated to do military service.

[11] For instance, Davis and von Wachter [2011] investigate how the incidence of a layoff when macroeconomic conditions are good compares to a contrasting situation. They find that estimated earnings losses are significantly larger when the unemployment rate in the displacement year is higher.

placement events, I first merge the displacement data with employer-employee records at the individual level. Since the employer-employee data contain plant id, I then identify the most common plant id within displacement id. In case of ambiguities, I drop those events altogether. I also drop events that only comprise the advance notification stage (with no lists submitted). Since the date at which the announcement was made is known, and since matched records include variables on when the match was formed and when it ended at the monthly level, I identify workers as nondisplaced if they worked at the displacing plant during the month of the announcement and their worker ids are not on the PES lists. Workers at the displacing plants whose spell ended before the announcement date or began after are dropped from the analysis. Finally, I drop events where the number of employees at the plant reported by the PES data is too different from the number of employees computed using the matched records.[12] This procedure results in 3,467 displacing plants and 62,171 workers (10,294 displaced and 51,877 nondisplaced) over the three-year period.

The sample retained for analysis was further restricted to workers with a match in the military enlistment data. Since the military enlistment data only cover men, this implies dropping women from the sample. Finally, to ensure that the displaced and nondisplaced workers are truly employed and not just possessing temporary contracts, I restrict the sample to those with at least 18 months of tenure with the current employer at the time of the displacement event.

To investigate how skills and age matter for the incidence of job displacement, I use the statistical model:

$$P_{ijt} = \alpha_j + \eta_t + \mathbf{z_i}\beta + \mathbf{x}'_{ijt}\gamma + \varepsilon_{ijt}, \tag{3.1}$$

as the baseline model. P_{ijt} is an indicator of worker i being displaced from plant j at time t, $\mathbf{z_i}$ denotes the independent variables of interest, including age categories and normalized measures of cognitive and noncognitive skills. Controlling for plant fixed effects, α_j takes time-invariant plant heterogeneity into account and the inclusion of year fixed effects, η_t, ensures that the results are not driven by macroeconomic developments. In addition, I control for tenure at the plant, denoted $\mathbf{x_{ijt}}$. When estimating the effect of age on displacement, I create three indicator variables denoting young (aged 25-34), middle aged (aged 35-44) and older workers (45-54). Equation (3.1) is estimated using Ordinary Least Squares.

A linear probability may not be appropriate for the study of effects of skills on displacement probability. Predicted probabilities may not be between zero and one and the assumption that effects are constant across the probabilities is strong. To complement my

[12]I have varied the definition of "too different". In the baseline, I take the difference between the two figures and divide it by the larger of the two numbers. If this fraction is larger than 25 %, I drop the event. The robustness of this procedure has been investigated and the results are quantitatively very similar to those reported.

analysis, I also estimate a logistic model, where the log of the odds of becoming displaced is assumed to be linear in the variables described above.

3.3.2 Results

Patterns in the incidence of displacement are presented in Table 3.1. Estimates of the linear probability model are presented in the columns labeled LP. While both skill measures appear important in predicting displacement, the slope is slightly steeper for noncognitive skills. Since the skill measures are positively correlated, the four left-most columns comprise upper bounds on the effect of the skill measures on the probability and the log-odds of displacement occurring, respectively.[13] When both ability measures are included, the magnitude of the estimated slopes goes down due to the positive correlation between the two. The addition of interaction and higher-order terms does not affect the results, which suggests that there is indeed a linear relationship between skill levels and the probability of displacement. The point estimates of the last column suggest that an increase of one standard deviation in either skill type decreases the probability of displacement occurring by one percentage point. The sample average probability of job displacement is about 16 percent.

Estimates of the logistic regression are presented in the columns labeled LM in Table 3.1.[14] The results are qualitatively similar to the linear probability model. When estimating a logistic regression with fixed effects, it is not possible to transform the coefficients to a general marginal effect on the probabilities as baseline probabilities differ across groups. However, assuming that the fixed effect is zero, the marginal effect of either skill measure on the probability of becoming displaced ranges between $[0.021, 0.028]$, depending on the specification.[15] This effect is computed using the predicted probability when evaluating the independent variables at their means. The effects are quantitatively twice as large as the estimates obtained by using the linear probability model.

The inclusion of plant fixed effects allows me to exploit within-plant variation and avoids the risk that plants with a high share of low ability workers may be more inclined to displace workers in the first place. For instance, industries such as construction and manufacturing employ relatively high numbers of low-skilled workers and exhibit a greater turnover due to their greater cyclical dependence. However, it may well be that the turnover rates within a firm are larger among low-skilled workers, implying that tenure

[13] The correlation between cognitive and noncognitive skills is 0.3501. Though the source of the covariance is ambiguous (see Lindqvist and Vestman, 2011), estimating equation 3.1 without each skill variable individually provides an upper bound for the effects while estimating the equation with both measures together yields a lower bound.

[14] The number of observations reported in the table, is lower than what was reported above. This is because some plants displace all workers which generates no within-workplace variation in the outcome variable.

[15] This assumption is imposed ex post to the within-plant regression. The logistic model coefficient is thus estimated using plant and year fixed effects.

is, on average, lower among low-skilled workers. This could imply differential layoff rates under the last-in-first-out rule (henceforth, LIFO) that governs displacement decisions. It could also be explained by the argument in Neal [1998]: as high-skilled workers are more capable of learning, they accumulate more job-specific skills and are therefore more highly valued by employers than low-skilled workers. This leads to lower turnover rates among high-skilled workers. However, all columns control for months of tenure at the plant.

Age is also important in explaining the incidence of displacement. According to Table 3.1, specifications (4) and (6), workers aged $35-44$ are less likely to be displaced than young workers aged $25-34$, although the difference is not significant.[16] Workers aged $45-54$, on the other hand, are significantly less likely to become displaced compared to young workers, with point estimates at 2.2 percentage points in the linear probability model.

3.4 Effects of Job Loss

3.4.1 Sample Restrictions and Empirical Framework

This section estimates the costs of job loss, but before starting the analysis, I impose certain restrictions on the data. The appropriate sample restrictions differ from the study of who becomes displaced. As in the study of the incidence of job loss, PES data are matched with employer-employee records at the plant-worker level using identifiers for the worker, the firm and the displacement event. After having added demographic variables to this dataset, I follow JLS by studying workers with stable employment relationships. In contrast to JLS, who focus on workers remaining with the same employer for six consecutive years, having a stable job is defined as having kept the same job for more than six quarters. I restrict the workers to those aged between 20 and 51 such that all workers have a potential attachment to the labor force during the sample period. Due to the high turnover in the construction sector, I exclude this sector.[17] When defining the control group, I also exclude the self-employed. As above, the sample is confined to men with enlistment data.

In the absence of randomly assigned displacement, I restrict the treatment group to only include displacement events where more than 80 percent of all workers who satisfy the above restrictions simultaneously received a displacement announcement. This is

[16] Due to the restrictions imposed on the data by the military enlistment records, the youngest workers in the sample are aged 25 in the displacement year and the oldest are aged 54.

[17] Eliason and Storrie [2006] impose this restriction because a displacement in the construction sector, that anyhow is characterized by a high turnover, is not comparable to displacements in other sectors. With the PES data, this issue is not as important, but I impose this restriction to make my results comparable to the previous literature.

done to make the selection into displacement as exogenous as possible within the workplace while maintaining statistical power. Alternatively, I also examine the robustness of my results by restricting the sample to include events where all workers were displaced simultaneously.[18] The potential control group consists of all workers in Sweden during this period who are not included in any list of displaced workers and satisfy the tenure-, age-, self-employment- and gender-restrictions imposed above. Tables 3.2 and 3.3 provide descriptive statistics of the sample.

The literature that estimates effects of job displacement typically uses the statistical approach taken in the program-evaluation literature by defining a treatment group (displaced) and a control group (nondisplaced).[19] When comparing outcomes before and after displacement across groups, I exploit the advantage of longitudinal panel data for workers in Sweden. I estimate the following model:

$$y_{ijt} = \alpha_i + \gamma_{jt} + \sum_{k \geq -5} \delta_j^k D_{ijt}^k + \mathbf{x}_{it}'\beta + \varepsilon_{ijt}, \tag{3.2}$$

where y_{ijt} denotes the outcome variable of individual i in group j at time t.[20] D_{ijt}^k is an indicator of worker i belonging to group j at time t and being displaced in year $t-k$. The formulation of the dummy variables implies that a worker displaced in 2002 faced the same situation in 2005 as a worker displaced in 2004 did in 2007. δ_j^k measures the effect on the outcome variable of displacement k years ago for a worker in group j. For simplicity, the displaced workers are sometimes referred to as treated and non-displaced workers as controls. By estimating the effects prior to actual displacement, I assess the validity of the treatment and control groups as the difference in their outcome variables should not differ significantly from zero in pre-treatment periods. One coefficient for each group and year is obtained and by including group-specific non-parametric time trends, γ_{jt}, the effect on the outcome variable is measured relative to nondisplaced workers within the same group and year. Because I include worker fixed effects, denoted α_i, the outcome differences refer to within variable variation. It also allows the selection into displacement to depend on time-invariant characteristics for the effects to be unbiased.

This approach yields the predicted effects for various characteristics, but the estimated parameters are too many to reveal whether the effects are significantly different from each other. In order to reduce the number of parameters to be estimated, I follow JLS and re-parametrize the above equation in the following way:

[18] These restrictions are imposed to minimize selection into displacement, in line with the plant closure literature. However, as discussed by Kuhn [2002], restricting the analysis to separations preceding plant closures has the caveat of falsely defining merging or simply restructuring plants as displacements. This is not a problem here, with lists of laid-off workers.

[19] Examples of studies taking such an approach include Jacobson et al. [1996], Couch and Placzek [2010] and Sullivan and von Wachter [2009].

[20] The only control variable included here, as captured by $\mathbf{x}_{it}'$, is age squared.

$$y_{ijt} = \alpha_i + \sigma_j t_j + \rho_j^{PRE} F_{ijt}^{PRE} + \rho_j^{IMPACT} F_{ijt}^{IMPACT} + \rho_j^{POST} F_{ijt}^{POST} + \mathbf{x}'_{it}\beta + \varepsilon_{ijt}, \quad (3.3)$$

where σ_j allows for differential group-specific linear time trends in the absence of displacement based on different characteristics, and the response to displacement is allowed to vary in the following way:

$$F_{ijt}^{PRE} = \begin{cases} t+k & \text{if } k \leq 0 \text{ and worker } i \text{ in group } j \text{ is displaced in year } k=0. \\ 0 & \text{otherwise} \end{cases}$$

$$F_{ijt}^{IMPACT} = \begin{cases} 1 & \text{if } k = 1,2 \text{ and worker } i \text{ in group } j \text{ is displaced in year } k=0. \\ 0 & \text{otherwise} \end{cases}$$

$$F_{ijt}^{POST} = \begin{cases} t+k & \text{if } k \geq 2 \text{ and worker } i \text{ in group } j \text{ is displaced in year } k=0. \\ 0 & \text{otherwise} \end{cases}$$

F_{ijt}^{PRE} represents a linear time trend before displacement, F_{ijt}^{IMPACT} captures a jump in the outcome variable immediately following displacement and F_{ijt}^{POST} is a linear time trend representing recovery. As the underlying trend accounts for heterogeneity, all estimates of the effect of job displacement will be relative to non-displaced workers in that same group.[21]

3.4.2 Results

Section 3.3 presents suggestive evidence that younger and less cognitively and noncognitively able workers are more likely to experience displacement. My ultimate purpose is to estimate whether the effects of displacement on different outcome variables are heterogeneous with respect to these characteristics. However, the effects are initially investigated for the full sample. Recall that the sample retained for an analysis of the job displacement effects differs from the sample used for analyzing the selection into displacement.

Using the extensive Swedish administrative data, I first analyze how disposable income evolves over time. This is a better measure than labor earnings for estimating the effect of displacement on a worker's welfare. Total income is then decomposed into labor earnings, capital income, and social and unemployment insurance in order to understand the mechanisms behind the effects on welfare.

Having studied the aggregate effects of displacement, I then study heterogeneous

[21] When individuals differ along a continuous dimension, I assume a linear relation in the heterogeneous impact. For cognitive and noncognitive skills which are normally distributed variables with mean zero, the baseline effect refers to workers with average skill levels.

responses based on cognitive and noncognitive skills and age. In order to estimate whether high-skilled workers are better at learning new skills and therefore experience faster rates of recovery, I focus on a subset of variables. In particular, I analyze the effects on labor earnings and the propensity to start studying, measured by the amount of student transfers received from the government, and by job training programs organized by the Public Employment Service.[22]

When estimating equation (3.2), the outcomes are measured annually and the data stretch from 1999 to 2009, thus encompassing a maximum of five pre-displacement years with seven post-displacement observations for each variable. I first estimate the equation without group-specific effects. The estimated coefficients and their associated 95 percent confidence intervals are displayed in Figure 3.1. The confidence intervals are derived from standard errors clustered at the level of the employer in the displacement year. Time is normalized so that year zero denotes the year that the displacement announcement was reported.

Looking at the figure, the first graph presents the effects on disposable income. As the sum of labor earnings, capital income, and social and unemployment insurance income at the individual level, this measure does not cover risk sharing and responses within the family, but the estimates provide an upper bound for the welfare impact. Disaggregating income into different components, the effects on labor earnings are presented in panel (b). In contrast to the results of the study by JLS, there is no estimated drop in labor earnings prior to displacement. If anything, displaced workers' earnings tend to increase slightly. One explanation for this may be that displacement is measured as the date at which the announcement was made, rather than the separation date. The existence of notice periods delays the effects of job loss.

The absence of effects prior to the displacement date ensures the validity of the parallel-trends assumption. However, one year after displacement, the labor earnings of displaced workers have dropped by SEK $67,747$, or 23.2 percent. There is some recovery over time but even seven years after displacement, the income is SEK $43,928$ lower than the counterfactual – a loss of 15.2 percent of the annual earnings. The post-displacement years (2003-2009) were macroeconomically stable apart from 2009. Unlike the analysis in JLS and Eliason and Storrie [2006], my results are thus not to be confounded by the additional impact of severe macroeconomic conditions on the effects of job loss.

To better understand the mechanisms behind this result and identify the consequences of displacement in more detail, Figure 3.1 also presents the effects on capital income (panel (c)). Interestingly, displaced workers experience significant capital income losses over time. With capital income comprising interest earnings on assets such as bank accounts, bonds and other securities, and income from rents, these losses are indicative of displaced

[22]Enrollment in adult secondary and tertiary education yields an entitlement to student transfers, while job training programs do not.

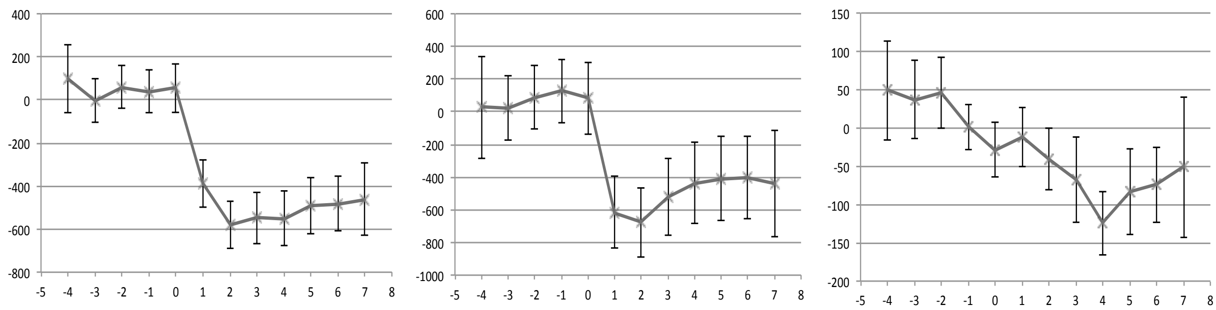
workers consuming their assets in order to smooth consumption. The magnitude of the effect grows over time after displacement, which may suggest that displaced workers use other means to smooth consumption before consuming their assets.

Figure 3.2 illustrates dependence on the welfare state by displaying the dynamic effects of displacement on unemployment-insurance benefits, student transfers received, labor-market training and participation in labor-market programs. Unemployment benefits provide income for displaced workers in the short run, but as the effect on UI benefits fades away, the dependence on other social insurance programs remains. In particular, student transfers increase upon displacement, implying that displaced workers aim at acquiring new skills. Student income comprises government transfers and loans issued by the government and students at universities, colleges, other tertiary educations and adult secondary schools are eligible. Four years after displacement, the increase in student transfers amounts to SEK 1,664, or 406 percent, relative to the counterfactual.

To validate the definition of displacement, I also present a more restrictive approach in which only those displacements where the entire workplace was displaced simultaneously are considered. The results from this exercise are similar in magnitude and are displayed in Table 3.6 in the Appendix. Furthermore, focusing on the effects for high-tenure workers, defined as those with more than 36 months of tenure with the same employer, yields similar effects.

Figure 3.1: Main Effects of Displacement, I

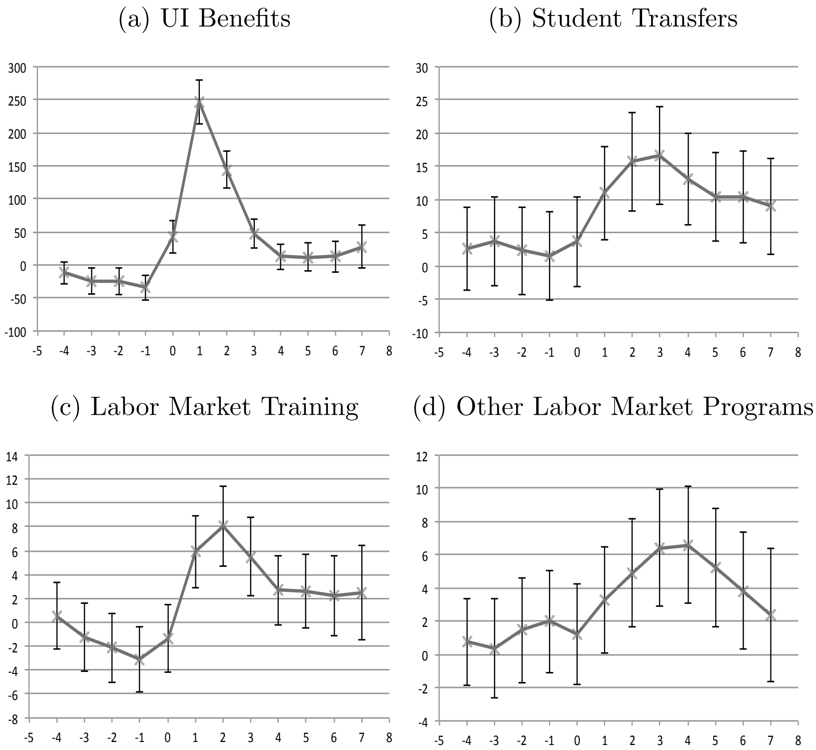
(a) Total Income (b) Labor Earnings (c) Capital Income



Notes: Estimates of the effects of job displacement on total income, labor earnings and capital income before and after displacement. Earnings and capital income are measured in 100 Swedish crowns (SEK). Earnings are expressed in real terms (denominated in 2002 SEK). The x-axis displays years since displacement. Confidence intervals are defined at the 95 % level and derived from standard errors clustered at the level of the employer in the displacement year.

To maintain clarity in the analysis of heterogenous effects of displacement based on cognitive and noncognitive skills as well as on age, I focus on a subset of the above outcome variables. Labor earnings and student transfers provide evidence of the actual adjustment to new circumstances and the willingness to acquire new skills. These variables also help me evaluate how skills and age matter for the ability to adapt to new labor-market conditions. Participation in education organized by the PES is separated

Figure 3.2: Main Effects of Displacement, II



Notes: Estimates of the effects of job displacement on UI benefits, student transfers, days in labor market training and days in other labor market programs before and after displacement. Income variables are measured in 100 Swedish crowns (SEK) and are expressed in real terms (denominated in 2002 SEK). The x-axis displays years since displacement. Confidence intervals are defined at the 95 % level and derived from standard errors clustered at the level of the employer in the displacement year.

from studies that qualify for student transfers to make clear the distinction between academic studies and labor-market oriented studies. The training provided by the PES is aimed at unemployed workers and participants qualify for unemployment benefits but not student transfers.

Figure 3.3 displays estimates of the impact of job displacement on labor earnings depending on cognitive and noncognitive skills, and age groups. For both cognitive and noncognitive skills, low-skilled refers to the bottom 25 percent of the skill distribution, high-skilled to the top 25 percent and average skills to those in between. For moderate changes in the definitions of low-, medium- and high-skilled workers, the results do not change. Focusing on small groups in the tails of the skill-distributions implies a severe loss of statistical power.[23]

Panels (a)-(f) in Figure 3.3 suggest that in absolute terms, the losses of high-skilled workers are larger than those of low-skilled workers, but that this is simply a consequence of high-skilled pre-displacement workers' earnings being higher.[24] More importantly, nei-

[23]The continuity of the skill variables is exploited in the regressions presented in Table 3.4.

[24]The average annual earnings one year before displacement among workers with low cognitive skills are 244, 960, implying earnings losses of about 23 % two years after displacement. For high-skilled

ther cognitive or noncognitive skills matter for the recovery pattern after displacement. The picture suggests that the abilities are equal in adapting to changing labor market conditions. However, confirming results established in earlier literature, age has a much stronger impact on the capacity to adapt after displacement than cognitive and noncognitive skills, as seen in panels (g)-(i).

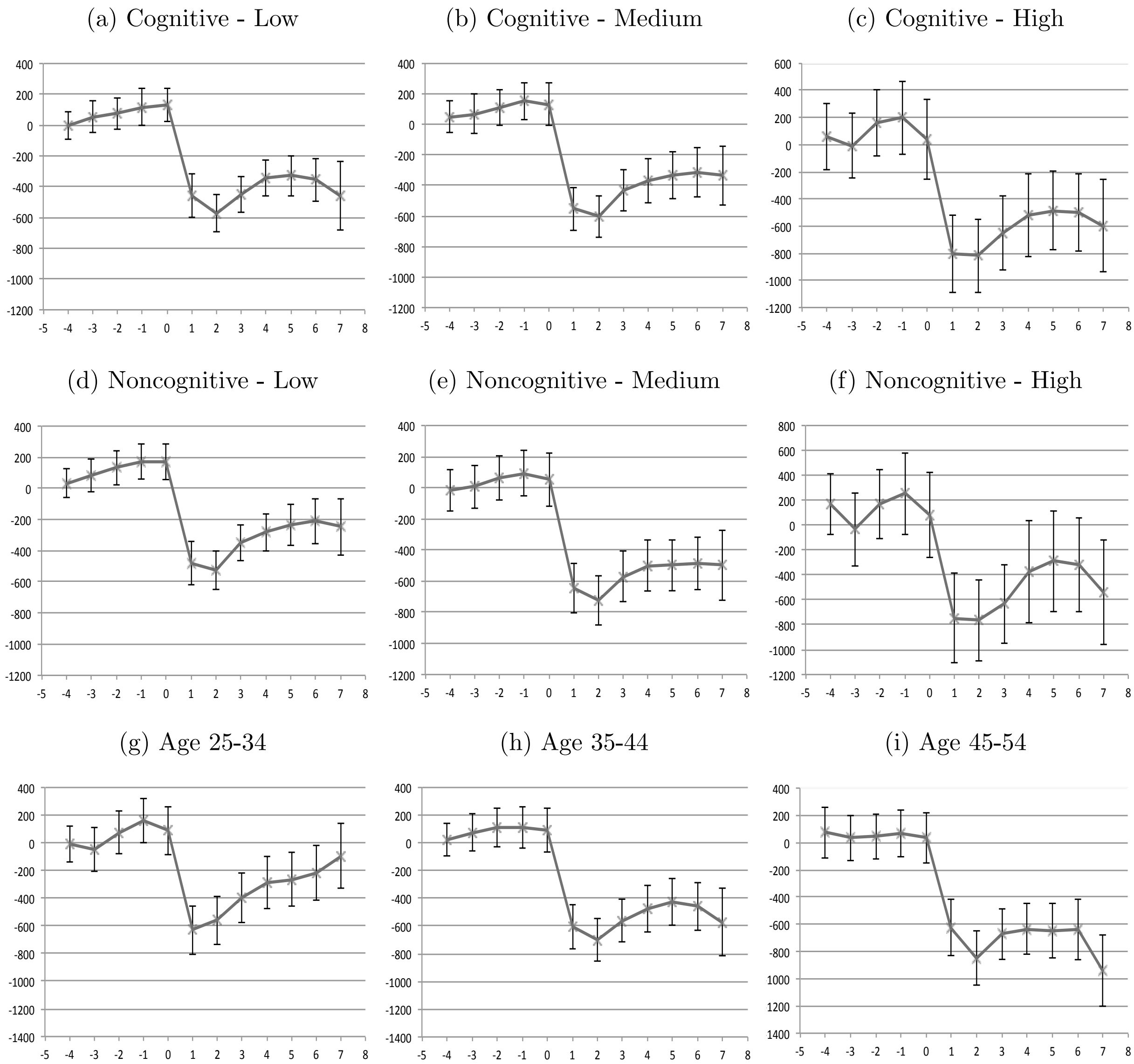
In another attempt to understand how displaced workers adapt to new labor market conditions by acquiring new skills, Figure 3.4 presents the effects of displacement on student transfers from the government. Panels (a)-(c) and (g)-(i) suggest that workers with high cognitive skills and young workers are more likely to start studying and receiving transfers. Eligibility for student transfers requires admittance to secondary or tertiary schools. Although acceptance to higher education often depends on cognitive ability, note that admission to public adult schools that do not select students based on skills also qualifies for student transfers. While the relatively larger impact of displacement on student transfers among workers with high cognitive skills may reflect a larger menu of education, this does not translate into a faster recovery of labor earnings. On the other hand, it may be that the larger effect on student transfers could partly, among young workers, explain the differential recovery of labor earnings between age groups.

Finally, Figure 3.5 presents the effects on labor-market training provided to unemployed workers. Low-skilled workers are more likely to receive labor-market training from the Public Employment Service than secondary or tertiary education. Even seven years after displacement, the average effect of displacement on training is positive and statistically different from zero at the 10 percent level, irrespective of skill measures. No differential effects across age groups emerge from the figures.

To analyze whether the earnings effects are significantly different across groups, I re-parametrize the estimation procedure along the lines of JLS and estimate (3.3) exploiting the continuous variables measuring skill and the indicator variables for age groups. Table 3.4 shows the results when earnings regressions are estimated for the pre, post and impact effects of job displacement. Workers with high skills experience larger initial drops in earnings but, once more, the difference is merely due to high skilled workers earning more when displaced. The recovery is not quicker among high-skill workers, however. The most pronounced differentials in the effects of displacement are found between age groups. Older workers fare significantly worse than younger workers and their recovery is significantly slower.

workers, the corresponding figures are $361,036$ and 22.7 %.

Figure 3.3: Effects on Labor Earnings by Skill and Age Groups



(a) Cognitive - Low (b) Cognitive - Medium (c) Cognitive - High

(d) Noncognitive - Low (e) Noncognitive - Medium (f) Noncognitive - High

(g) Age 25-34 (h) Age 35-44 (i) Age 45-54

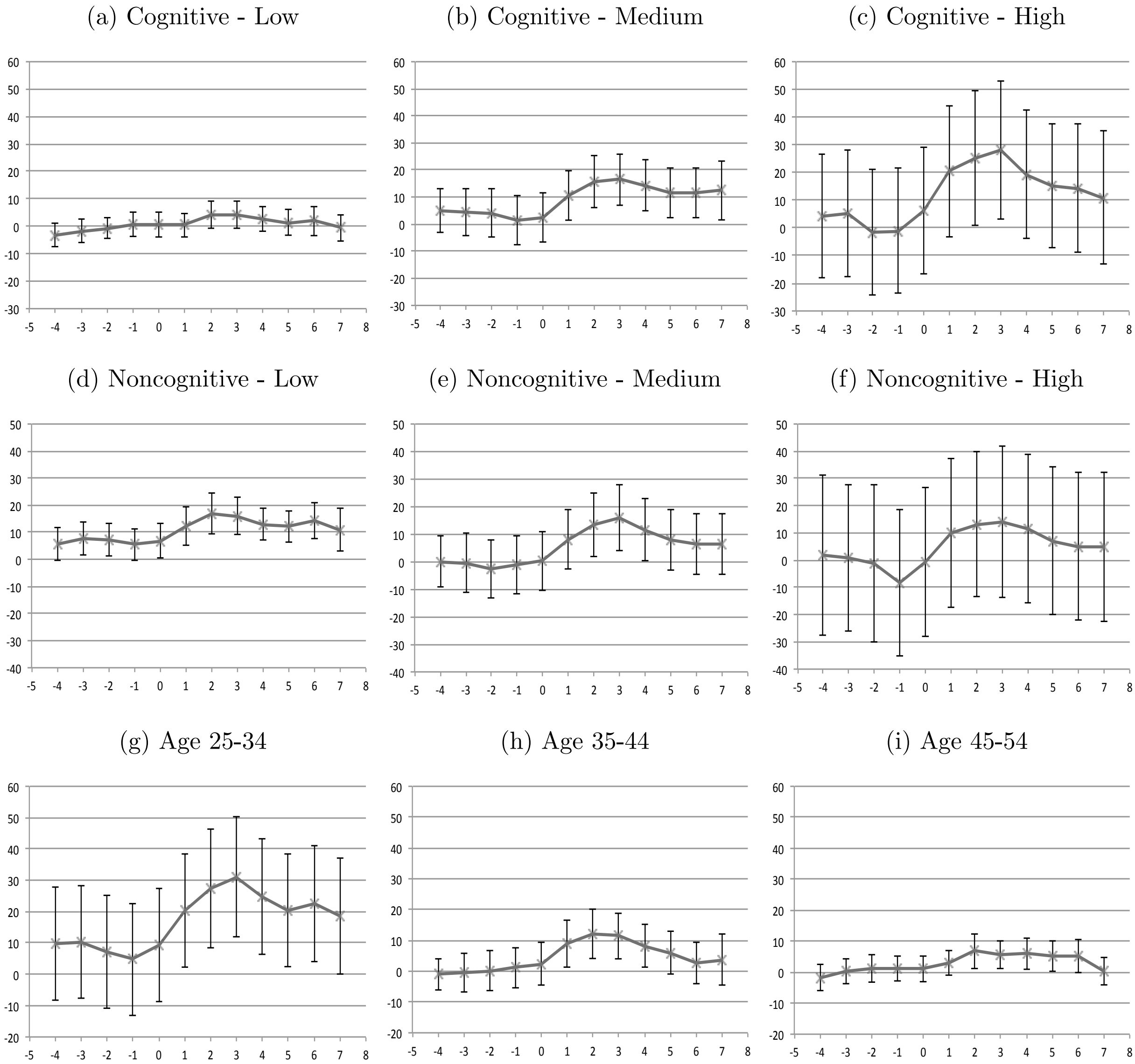
Notes: Earnings are measured in 100 Swedish crowns (SEK) and expressed in real terms (denominated in 2002 SEK). The x-axis displays years since displacement. Confidence intervals are defined at the 95 % level and derived from standard errors clustered at the level of the employer in the displacement year.

3.5 Bias Estimation

A weakness of earlier studies that use administrative data to study displacement effects on labor-market outcomes is the inability to distinguish quits from layoffs. To overcome this issue, separations in times of distress are regarded as involuntary quits initiated by the employer. Such time periods are either defined as windows around mass separations or time periods preceding plant closure.[25] However, some bias is inevitable without any information regarding actual displacements. First of all, workers have notice periods that

[25]The common problem of false firm deaths, or falsely labeling the disappearance of an organization identity number as a plant closure, is solved by following workers after a potential plant closure. If a majority continues with the same employer, the event is not coded as a plant closure.

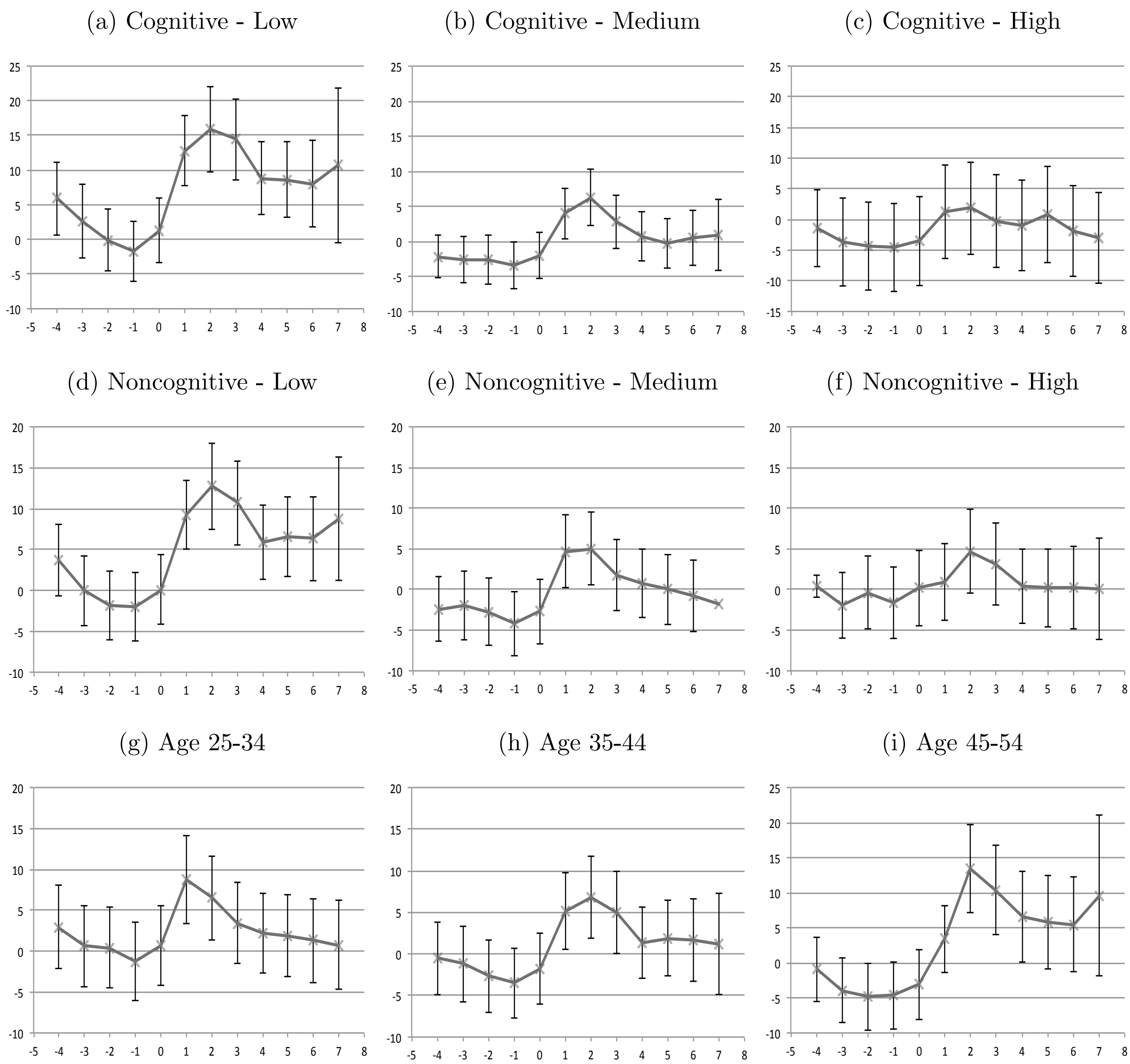
Figure 3.4: Effects on Student Transfers by Skill and Age Groups



Notes: Student transfers are measured in 100 Swedish crowns (SEK) and expressed in real terms (denominated in 2002 SEK). The x-axis displays years since displacement. Confidence intervals are defined at the 95 % level and derived from standard errors clustered at the level of the employer in the displacement year.

are negotiated in individual contracts and depend on variables such as tenure and age. An announcement of displacements for all workers at the same moment will thus lead to heterogenous layoff dates. An even more serious problem, pointed out by Kuhn [2002] is that the actual separation date is endogenous. The observed separation date in the data may be different from the reported layoff date since workers can start searching for new jobs, register as unemployed, or move to some other region before the prescribed displacement date. In fact, Addison and Portugal [1987, 1992] use data from the Current Population Survey and the Displaced Worker Survey, respectively, and show that longer notification times translate into shorter unemployment spells and quicker reemployment rates. This means that at the end of a particular time window, the observed separations

Figure 3.5: Effects on Days in Job Training by Skill and Age Groups



(a) Cognitive - Low (b) Cognitive - Medium (c) Cognitive - High

(d) Noncognitive - Low (e) Noncognitive - Medium (f) Noncognitive - High

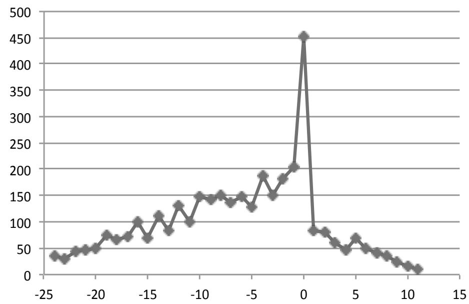
(g) Age 25-34 (h) Age 35-44 (i) Age 45-54

Notes: The x-axis displays years since displacement. Confidence intervals are defined at the 95 % level and derived from standard errors clustered at the level of the employer in the displacement year.

will consist of displaced workers with long notice periods who were not able to find new jobs during the notice period.

Expanding the time window will lead to more displaced workers being included, but it will also mislabel additional voluntary quits as displacements. If workers expect a displacement announcement, they may engage in an on-the-job-search before the announcement. Figure 3.6 shows the number of separations from the plant that occur around the displacement announcement, normalized such that zero corresponds to the announcement month. The figure is computed using all displacement announcements, not just those that satisfy the restrictions for an analysis of the incidence or costs of job loss, respectively. There is a gradual increase in the number of separations until the month of the announcement, when the number of separations spikes. The announcement is clearly anticipated and workers react before the actual announcement. Without information on who are actually displaced, the current convention prescribes labeling separations as displacements if they occur within a time window preceding plant closure or within time windows of mass-separations. The graph shows that this can potentially be problematic since quite a few separations around the announcement of a displacement are voluntary quits.

Figure 3.6: Total Number of Separations



This graph shows the total number of separations around the normalized month of the displacement announcement for displacing plants during 2002, 2003 and 2004. The x-axis denotes normalized difference in months between the announcement and the separations, with the announcement occurring in month 0. The y-axis shows the total number of separations. The workers have at least 18 months of tenure.

Kuhn [2002] further discusses the problem of falsely labeling separations as displacements. Distressed firms may find ways of disposing of workers, other than formally displacing them. Employers and employees may voluntarily sign a termination contract, older workers may enter into early retirement, acquisitions of parts of distressed firms may imply transitions of workers to the new employer, and workers may enter into disability insurance without having been displaced.[26]

The PES data permit an estimation of the consequences of defining displacements as separations occurring in time windows around mass separations or before plant closures, rather than employer-initiated layoffs.

In this paper, I define workers as treated if they were laid off in an event where 80 % of the workplace were simultaneously displaced, but have investigated the robustness of this definition in various ways (see Table 3.6). These definitions differ quite substantially from previous studies. In JLS and Sullivan and von Wachter [2009], separations are defined as displacements if the number of separations from a given firm in any calender year between 1980 and 1986 is larger than 30 percent of the maximum number of employees in any year during the 1970s. Eliason and Storrie [2006], instead, focus on time windows preceding firm deaths of the length of one, two and three years.

[26]Pfann and Hamermesh [2001] instead focus on the process of a plant closure and argue that the pool of workers who remain with closing plants until the end are a selected group of workers. They further show that the selection of workers out of closing plants will create a bias in the estimates of costs of job loss.

I now compare my baseline results with the results obtained using three different definitions that are common in the literature. First, I define workers as displaced if they separate from their plant in a year where the number of separations amounts to at least 30 percent of the number of employees in the previous years. JLS restrict the sample to firms with at least 50 workers, I drop firms with less than five workers. This procedure is pursued for separations in 2002, 2003 and 2004. Second, I identify separations occurring in time windows before plant closure. I follow Statistics Sweden's definition of plant closure, by defining events where at least 50 percent of the workforce separate within one year. Events in which a majority of these workers still work together one year after, either through the merger of two workplaces or through splits of one workplace, are further removed.[27] To qualify as a control worker for comparison with the treated in either year, a worker must hold a stable job during the year, have at least 18 months of tenure, be of working age throughout the sample period and not be displaced during that year.

Table 3.5 presents the results from estimating equation (3.2) using different definitions of displaced workers. Specification (1) of the table presents the baseline results when PES data have been used. Specification (2) shows the results of job loss when separations where more than 30 percent of the workforce leave the plant are defined as job displacements. The effects from the *mass-separation* definition are considerably lower compared to the baseline estimates. These workers experience labor earnings losses of a smaller magnitude than baseline estimates. Although pre-trend estimates indicate a violation of the parallel trends assumption (the effect of job loss is significantly different from zero in the years before displacement), the coefficients estimated using PES data and *mass-separations*, respectively, are significantly different from each other in the post-displacement period but the difference fades out over time. The fourth column from the left, labeled *T-stat*, shows t-statistics for the difference in coefficients using the baseline sample and the mass-separations sample.

Specification (3) of Table 3.5 shows estimated effects of job loss when separations in a three-year time window prior to plant closure are considered as job displacements. These estimates significantly overstate the effects of displacement. The results are surprising if we believe that workers would try to leave the firm preceding plant closure and the less able would remain until the end. On the other hand, the selection of workers who remain until the end could be those loyal to the firm and not necessarily those with the worst lookouts on the labor market. Finally, the estimates for workers separating within a one-year window preceding plant closure, presented in specification (4) of Table 3.5, are slightly larger than the coefficients for actual displacements. The difference is never

[27] Alternatively, I could just compare my results to alternative estimates in the literature. Because it is difficult to separate the impact of differences in information on displaced workers from other differences, e.g. macroeconomics conditions (see Davis and von Wachter, 2011), on diverging estimates of job loss, I estimate the impact using the same data.

significantly different from the baseline results, as indicated by the associated T-stat column.

The estimated differences are not purely due to the use of different definitions. The displacement data I use include information on the date at which the announcement was received by the PES. This enables me to restrict the sample of displaced workers to include only displaced workers who were part of an announcement in which a majority of the workers were displaced simultaneously, so as to minimize the selection into displacement. Such restrictions are not possible when one only relies on employer-employee data and, consequently, the estimated differences are also affected by the inability to distinguish quits from layoffs and the lack of information on the date at which the displacement was announced.

3.6 Concluding Remarks

This paper uses a unique dataset to analyze, first, who becomes displaced and, second, the effects of displacement. Because I use administrative data on displacements at the individual level, I can overcome the problem of distinguishing quits from layoffs in linked employer-employee data. I first analyze the selection into displacement and find that both cognitive and noncognitive skills reduce the probability of displacement. Moreover, older workers are less likely to be displaced than younger workers.

When I analyze displacement effects on different labor-market outcomes, I focus on displacement events where 80 percent or more of the workforce simultaneously receive an announcement of an impending displacement. The persistent earnings losses found in the earlier literature are confirmed. The estimates are large, especially since the period under study was macroeconomically stable.

Turning to heterogenous effects, the effects of displacement are not significantly different across groups with high or low cognitive or noncognitive skills and the largest differences are found in the effects of job loss across age groups. As expected, the labor earnings of younger workers recover at a faster pace than those of older workers.

These results have direct policy implications. The importance of age suggests that older workers require more assistance in adjusting to new labor market conditions after displacement. However, I find no justification for differential treatment with respect to skills.

Finally, I assess the validity of using employer-employee matched data when analyzing displacement effects. The previous literature has suffered from not being able to directly observe displacements and has therefore developed strategies to assess whether separating matches between employers and employees are displacements or not. I show that these strategies may not be accurate. Separations in times of mass-separations and separations in three-year time windows preceding plant closure significantly understate and overstate the costs of job loss, respectively. Labeling separations in a one-year time interval before a plant closes down generates estimates of job loss that are closest to those obtained by using actual displacement data.

Bibliography

J. T. Addison and P. Portugal. The effect of advance notification of plant closings on unemployment. *Industrial and Labor Relations Review*, 41:3–16, 1987.

J. T. Addison and P. Portugal. Advance notice and unemployment: New evidence from the 1988 displaced worker survey. *Industrial and Labor Relations View*, 45:645–664, 1992.

S. Bowles, H. Gintis, and M. Osborne. The determinants of earnings: A behavioral approach. *Journal of Economic Literature*, 39:1137–1176, 2001.

B. Carlstedt. Cognitive abilities: Aspects of structure, process and measurement. PhD Dissertation, University of Gothenburg, 2000.

J. Cawley, J. Heckman, and J. Vytlacil. Three observations on wages and measured cognitive ability. *Labour Economics*, 8:419–442, 2001.

K. A. Couch and D. W. Placzek. Earnings losses of displaced workers revisited. *American Economic Review*, 100:572–589, 2010.

S. J. Davis and T. von Wachter. Recessions and the costs of job loss, 2011.

M. Eliason and D. Storrie. Lasting or latent scars? swedish evidence on the long-term effects of job displacement. *Journal of Labor Economics*, 24:831–856, 2006.

M. Eliason and D. Storrie. Job loss is bad for your health – swedish evidence on cause–specific hospitalization following involuntary job loss. *Social Science and Medicine*, 68: 1396–1406, 2009.

H. S. Farber. The incidence and costs of job loss: 1982–91. *Brookings Papers on Economic Activity: Microeconomics*, 1:73–119, 1993.

H. S. Farber. The changing face of job loss in the united states, 1981–1995. *Brookings Papers on Economic Activity*, 28:55–142, 1997.

H. S. Farber. Job loss in the united states, 1981–1999, 2001.

H. S. Farber. Job loss in the united states. *NBER Working Paper 9707*, 2003a.

H. S. Farber. Job loss in the united states, 1981–2001, 2003b.

H. S. Farber. Job loss and the decline in job security in the united states. *Princeton University Industrial Relations Working Paper 520*, 2007.

J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24:411–482, 2006.

R. J. Herrnstein and C. Murray. *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press Paperbacks, 1994.

L. S. Jacobson, R. J. LaLonde, and D. G. Sullivan. Earnings losses of displaced workers. *American Economic Review*, 83:685–709, 1996.

P. J. Kuhn. *Loosing Work, Moving On: International Perspectives on Worker Displacement*. W. E. Upjohn Insitute for Employment Research, 2002.

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

D. Neal. The link between ability and specialization: An explanation for observed correlations between wages and mobility rates. *The Journal of Human Resources*, 33: 173–200, 1998.

G. A. Pfann and D. S. Hamermesh. Two-sided learning, labor turnover and worker displacement, 2001.

M. Rege, K. Telle, and M. Votruba. Parental job loss and children's school performance. *Review of Economic Studies*, 78:1462–1489, 2010.

C. J. Ruhm. Are workers permanently scarred by job displacements? *American Economic Review*, 81:319–324, 1991.

A. H. Stevens. Persistent effects of job displacement: The importance of multiple job losses. *Journal of Labor Economics*, 15:165–188, 1997.

D. Sullivan and T. von Wachter. Job displacement and mortality: An analysis using administrative data. *Quarterly Journal of Economics*, 124:1265–1306, 2009.

R. Topel. Specific capital and unemployment: Measuring the costs and consequences of job loss. *Carnegie-Rochester Conference Series on Public Policy*, 33:181–214, 1990.

T. von Wachter and E. Handwerker. Variation in the cost of job loss by worker skill: Evidence using matched data from california, 1991-2000, 2009.

T. von Wachter, E. Handwerker, and A. Hildreth. Estimating the 'true' cost of job loss: Evidence using matched data from california 1991-2000, 2009.

Table 3.1: **Selection into Displacement**

	(1)		(2)		(3)		(4)		(5)		(6)	
	LP	LM	LP	LM	LP	LM	LP	LM	LP	LM	LP	LM
Cognitive Skills	-0.014***	-0.177***			-0.011***	-0.137***			-0.011***	-0.144***	-0.011***	-0.141***
	(0.002)	(0.022)			(0.002)	(0.021)			(0.002)	(0.021)	(0.002)	(0.021)
Noncognitive Skills			-0.015***	-0.185***	-0.012***	-0.148***			-0.012***	-0.156***	0.013***	-0.162***
			(0.002)	(0.020)	(0.002)	(0.019)			(0.002)	(0.020)	(0.002)	(0.020)
Cognitive Squared									0.001	0.004	0.001	0.004
									(0.001)	(0.014)	(0.001)	(0.014)
Noncognitive Squared									-0.001	-0.023*	-0.001	-0.014
									(0.001)	(0.013)	(0.001)	(0.013)
Cogn. x Nonc.									-0.002	-0.024	-0.002	-0.029
									(0.002)	(0.019)	(0.002)	(0.019)
Age 35-44							-0.005	-0.107***			-0.005	-0.101***
							(0.003)	(0.039)			(0.003)	(0.039)
Age 45-54							-0.022***	-0.375***			-0.023***	-0.382***
							(0.004)	(0.053)			(0.004)	(0.054)
Observations	52,044	52,044	52,044	52,044	52,044	52,044	52,041	52,041	52,044	52,044	52,041	52,041

Notes: The dependent variable is an indicator of displacement in the columns labeled LP. In the columns labeled LM, the dependent variable is the log of the odds of becoming displaced. All columns control for tenure (in months), plant and year fixed effects. Standard errors are clustered at the plant level. *: p<0.10, **: p<0.05, ***: p<0.01.

Table 3.2: **Sample Characteristics, Effects of Displacement**

	All	Displaced Workers	Nondisplaced Workers
Labor Earnings, year -1	2802.091	2924.940	2801.424
	(1744.942)	(1544.177)	(1745.948)
Unemployment Inci, year −1	0.051	0.029	0.051
	(0.220)	(0.168)	(0.221)
Unemployment Days, year −1	4.550	2.2	4.563
	(26.44)	(16.937)	(26.486)
Age	37.917	36.857	37.923
	(6.820)	(6.715)	(6.821)
Tenure	87.906	66.286	88.024
	(58.386)	(49.531)	(58.409)
Skills:			
Cognitive Skills	0.056	-0.008	0.056
	(0.949)	(0.993)	(0.949)
Noncognitive Skills	0.005	-0.119	0.005
	(0.937)	(0.915)	(0.937)
Industry:			
Manufacturing	0.341	0.463	0.341
	(0.474)	(0.499)	(0.474)
Service Sector	0.146	0.261	0.145
	(0.353)	(0.439)	(0.352)
Wholesale and Retail trade	0.154	0.122	0.154
	(0.361)	(0.328)	(0.361)
Transport and Communications	0.122	0.113	0.122
	(0.327)	(0.316)	(0.327)
Public Sector	0.169	0.008	0.170
	(0.374)	(0.088)	(0.375)
Intensive Margin:			
Hourly Wage (SEK)	143.250	140.852	143.258
	(64.762)	(54.024)	(64.794)
Hours Worked as Share of Full-time	95.798	98.623	95.788
	(18.844)	(12.978)	(18.860)

Notes: Standard deviations in parentheses. The sample consists of men aged 20 to 51 with more than six quarters of tenure at the time of treatment. Workers qualify as displaced if the job loss event affects more than 80 % of the workers at the plant. Displaced workers who do not satisfy the restrictions are removed altogether.

Table 3.3: **Number of Observations, Effects of Displacement**

	Displaced Workers	Nondisplaced Workers
Person level	2,703	497,287
Person-Year level	29,733	5,470,157
Observations, selection	10,294	51,877

Notes: The analysis of the incidence of displacement is pursued using within-workplace variation. In the econometric analysis, workplaces where all workers are displaced simultaneously are thus dropped.

Table 3.4: **Labor Earnings Losses by Characteristics, Separate Regressions**

	Separate Regressions			One Regression		
	Pre	Drop	Post	Pre	Drop	Post
Main (mean cognitive and noncognitive, young)				135.215***	-240.934*	19.848
				(28.252)	(129.330)	(20.045)
Main (average cogn.)	76.136***	-489.538***	-19.651**			
	(17.409)	(75.054)	(8.480)			
Additional effect, cogn.	18.132	8.499	5.721	-23.296	-51.478	1.338
	(17.136)	(76.707)	(7.695)	(20.820)	(79.905)	(10.727)
Main (average ncog.)	72.142***	-491.728***	-18.096*			
	(19.070)	(83.353)	(9.584)			
Additional effect, ncog.	-6.722	-7.777	9.323	19.080	-17.141	-2.278
	(22.594)	(86.173)	(10.208)	(15.272)	(75.171)	(9.126)
Main (age 25-34)	126.661***	-280.885**	16.215			
	(29.769)	(135.686)	(20.398)			
Additional effect, age 35-44	-93.868***	-393.173**	-65.064**	-90.043***	-349.849**	-56.923**
	(33.079)	(170.335)	(25.523)	(33.806)	(170.609)	(25.813)
Additional effect, age 45-54	-99.775***	-474.131***	-84.425***	-91.233***	-398.492**	-66.007***
	(32.615)	(170.551)	(23.965)	(31.818)	(167.234)	(22.780)

Notes: Coefficients reveal differences across groups. For continuous variables, the main effect corresponds to the effect on an average worker and the impact is assumed to be linear. All specifications control for worker fixed effects and group-specific time trends. Standard errors are clustered on the level of the employer at the time of displacement. *: p<0.10, **: p<0.05, ***: p<0.01.

Table 3.5: **Labor Earnings Losses, Different Definitions**

	Main (1)	Mass Layoff (2)	T-stat	Plant Cl. - 3y (3)	T-stat	Plant Cl. - 1y (4)	T-stat
-4	26.865	72.209***	0.706	62.606**	0.464	-122.180	0.898
	(46.067)	(18.167)		(30.969)		(119.885)	
-3	22.286	152.912***	1.642	79.883**	0.619	7.642	0.092
	(53.173)	(26.385)		(39.887)		(106.224)	
-2	86.646*	209.108***	1.458	184.555***	1.04	181.884*	0.598
	(52.143)	(31.862)		(41.965)		(107.007)	
-1	125.878**	250.414***	1.427	249.145***	1.224	169.096	0.265
	(55.718)	(31.552)		(44.998)		(107.434)	
0	81.702	165.145***	0.913	9.021	0.653	-110.140	1.131
	(61.135)	(30.292)		(50.089)		(108.468)	
1	-615.854***	-105.948*	4.066***	-756.449***	1.179	-648.950***	0.188
	(64.113)	(61.294)		(55.164)		(112.016)	
2	-677.471***	-184.255**	3.57***	-917.426***	2.065**	-832.725***	0.905
	(59.573)	(78.565)		(56.634)		(111.978)	
3	-522.465***	-183.165***	2.83***	-831.744***	2.723***	-713.833***	1.12
	(58.584)	(61.302)		(55.006)		(112.317)	
4	-436.267***	-148.944***	2.529***	-774.376***	2.939***	-672.774***	1.324
	(62.313)	(51.296)		(52.715)		(116.332)	
5	-410.483***	-138.388**	2.185**	-693.841***	2.363**	-609.809***	1.078
	(64.512)	(60.038)		(55.416)		(120.308)	
6	-402.319***	-262.875***	1.189	-698.127***	2.377***	-629.788***	1.258
	(64.919)	(52.340)		(59.530)		(115.961)	
7	-439.282***	-277.464***	1.063	-719.237***	1.831*	-744.621***	1.515
	(79.622)	(72.607)		(73.251)		(121.896)	

Notes: The table presents the effects of displacement on labor earnings with different definitions of displacement. The column *Main* defines displaced workers as those on the list of displaced workers that the employers submit to the Public Employment Service. In *Mass Layoff*, displaced workers are defined as separations occurring in years when more than 30 % of the workforce separate and *Plant Cl. - 3y* and *Plant Cl. - 1y* provide evidence on the effects when workers who separate within a time window of three years and one years, respectively, are labeled as displaced. Standard errors are clustered at the level of the employer in displacement year. *: p<0.10, **: p<0.05, ***: p<0.01.

Table 3.6: **Effects on Displacement in Different samples**

Year	Baseline	Restrictive	Long Tenure
-4	26.865	-6.486	52.95
	(46.067)	(61.667)	(49.451)
-3	22.286	-29.288	118.606**
	(53.173)	(70.282)	(55.397)
-2	86.646*	74.72	82.525
	(52.143)	(67.442)	(53.571)
-1	125.878**	97.94	56.373
	(55.718)	(75.141)	(54.55)
0	81.702	7.437	67.509
	(61.135)	(84.798)	(60.395)
1	-615.854***	-722.086***	-551.861***
	(64.113)	(83.205)	(70.871)
2	-677.471***	-703.144***	-682.29***
	(59.573)	(76.004)	(60.587)
3	-522.465***	-540.683***	-565.695***
	(58.584)	(74.197)	(59.35)
4	-436.267***	-452.036***	-494.934***
	(62.313)	(81.882)	(62.103)
5	-410.483***	-423.49***	-470.502***
	(64.512)	(84.278)	(65.576)
6	-402.319***	-415.114***	-493.085***
	(64.919)	(84.94)	(66.169)
7	-439.282***	-539.986***	-519.298***
	(79.622)	(104.804)	(83.693)
Observations	5,466,696	5,134,501	273,616
R-squared	0.188	0.185	0.171

Notes: Effects of displacement depending on the definition of displacement. Column (1) shows the baseline estimations. In column (2), only displacements in which the entire workplace receive displacement at the same time are considered. Finally, column (3) requires workers to have more than 36 months of tenure before time 0. Effects are measured in 100 SEK and are expressed in real terms. The error terms are clustered at the level of the employer at the onset of displacement. Years are normalized so year 0 denotes the displacement year. All specifications control for year and worker fixed effects. *: p<0.10, **: p<0.05, ***: p<0.01.

Chapter 4

Complementary Roles of Connections and Performance in Political Selection in China*

4.1 Introduction

Who becomes a top politician in China? The spectacular economic performance of the Chinese economy in the past few decades suggests that the politicians selected to rule the country may have been conducive to growth, or at least not detrimental to economic development. However, the existing literature on political selection largely remains quiet about what determines the selection of politicians in a non-democratic country like China.[1]

In this paper, we attempt to fill this gap in the literature by examining what determines the promotion of provincial leaders in China, a pool of candidates for top posts in the central government.[2] In particular, we focus on two determinants of promotion often

*We thank Tim Besley, Antonio Ciccone, Claudio Ferraz, Lucie Gadenne, Nicola Gennaioli, Maitreesh Ghatak, Anna Larsson Seim, Shi Li, Rocco Macchiavello, Dirk Niepelt, Elena Nikolova, Mårten Palme, Albert Park, Torsten Persson, Per Pettersson-Lidbom, David Strömberg, Dan Tortorice, Shing-Yi Wang, Li-An Zhou, Fabrizio Zilibotti, and seminar participants at the CEPR/BREAD/AMID Conference on Development Economics in 2011, Barcelona Development Economics Workshop in 2012, Tartu, Gothenburg, Zurich, EBRD, Pennsylvania, Edinburgh, Strathclyde, Singapore Management University, National Graduate Institute of Policy Studies, Osaka, and Hong Kong University of Science and Technology. We also thank Christina Lönnblad for editorial assistance. Financial support from ERC grant 249580 (Kudamatsu) and Handelsbanken Research Foundation (Jia and Seim) is gratefully acknowledged.

[1] Most studies in the literature on political selection focus on democracy. See Cameron et al. [2008], Messner and Polborn [2004], Besley [2005], Mattozzi and Merlo [2008, 2011], Ferraz and Finan [2011], Galasso and Nannicini [2011] and Besley et al. [2012] among others. Acemoglu et al. [2010] and Besley and Reynal-Querol [2011] compare democracies and non-democracies in terms of political selection. Egorov and Sonin [2011] focus on the dictator's trade-off between choosing competent and loyal politicians as his subordinates. Garcia-Jimeno and Robinson [2010] provide empirical evidence on the type of politicians appointed to mayorship in 19th century Colombia. Francois et al. [2012] analyze how African dictators allocate cabinet minister positions across different ethnic groups.

[2] For example, the three most recent General Secretaries of the Chinese Communist Party (the highest ranked politician in China) all used to be provincial leaders: Shanghai (one of the four municipalities

discussed by academics and observers of Chinese politics: performance and connections.

Since China is ruled by the Communist Party, the political selection process is equivalent to the promotion of party officials. On the one hand, several empirical studies show that party officials in local governments are more likely to be promoted to higher levels of government if they achieve higher economic growth.[3] On the other hand, anecdotal evidence and a more systematic empirical analysis by Shih et al. [2012] suggest that social connections to top leaders determine the chance of promotion.[4] These two views also appear prominently in the mass media coverage of Chinese politics.[5] We argue that considering either or both of these two aspects in isolation fails to take into account the possibility of important interaction effects between connections and performance in the promotion process. To check this possibility, we conduct an empirical analysis of whether connections and performance are complements or substitutes as the determinants of promotion for provincial leaders.

As the interdependent roles of connections and performance in the promotion decision have never been discussed in the previous literature, we first present a simple model of promotion to illustrate why connections and performance may jointly affect the chance of promotion. In this model, promotion acts as a screening device.[6] Connections play either a loyalty-fostering role (increasing the top politicians' probability of survival) or an informational role (conveying information on the ability of candidates for promotion). When the loyalty-fostering role of connections dominates, connections and performance complement each other to increase the chance of promotion. If the informational role dominates, on the other hand, the two determinants of promotion are substitutes.

To investigate whether the correlation between promotion and performance is stronger or weaker for connected officials as compared to unconnected ones, we construct a sample of Chinese provincial leaders in office between 1993 and 2009. Using the curriculum vitae

---

with provincial status) for Jiang Zemin, Guizhou and Tibet for Hu Jintao, and Zhejiang and Fujian for Xi Jinping.

[3]Zhiyue [1996, 2002], Maskin et al. [2000], Li and Zhou [2005] and Chen et al. [2005] provide evidence for provincial leaders, Whiting [2000] and Edin [2003] for township leaders, and Landry [2008] for mayors. Huang [1998] appears to be the first to argue that the Chinese central government uses performance-based promotion to overcome the difficulty in controlling local government officials. Blanchard and Shleifer [2001] apply this argument to explain economic growth in China. Li [2012] provides evidence that the performance-based promotion scheme for provincial leaders drives business cycles in China. See also Xu [2011] for a thorough survey of this literature.

[4]Jiang Zemin, General Secretary from 1989 to 2002, is well-known for having promoted his former colleagues when he was the leader of Shanghai (those promoted are thus known as the Shanghai clique). Hu Jintao, who succeeded Jiang in 2002, is also widely known for having promoted his former colleagues while being the leader of the Communist Youth League, a youth organization of the Chinese Communist Party (such promoted officials are known as *tuanpai*).

[5]For example, in the lead-up of the 18th National Congress of the Communist Party held in November 2012, where many promotion decisions would be made, the *New York Times* published articles emphasizing both aspects: Zhang [2012] for the performance view and Wong [2012] for the connection view.

[6]In Appendix Section A.1, we also present a model of promotion where promotion acts as an incentive scheme for provincial leaders to boost economic growth.

of Communist Party officials, we measure connections between these provincial leaders and the top seven or nine party officials at the center by whether they used to work in the same branch of the party or the government in the same period.[7] To measure performance, we follow the literature and use the real GDP growth of the province that each leader rules.

Neither connections nor growth are exogenous, even if we control for province and year fixed effects. Therefore, we do not claim that our analysis establishes causal effects of connections and performance on promotion. Our aim is instead to present a robust empirical pattern that has previously been ignored in the literature, namely, the correlation between promotion and the *interaction* of connections with performance.[8]

We find that connected provincial leaders are, on average, significantly more likely to be promoted than unconnected ones. However, this difference is driven by a stronger positive correlation between promotion and economic growth for connected officials. Low-performing provincial leaders are unlikely to be promoted irrespective of their connections, while connections increase the likelihood of promotion for high-performing provincial leaders. In other words, connections and performance are indeed complements in the promotion of provincial leaders in China. In light of our theoretical framework, this evidence is consistent with the prediction that connections mainly play a role in fostering the loyalty of subordinates to top politicians. Moreover, we find that the complementarity is stronger for the connected pairs in which provincial leaders are substantially younger than the PSC members. Since Communist Party officials compete for high office with similar-aged peers, but not with senior peers, this finding provides further evidence in favor of the loyalty-fostering role of connections.

The documented pattern of correlations is remarkably robust to different definitions of promotion and to controlling for a wide range of covariates and their interaction terms with provincial economic growth that may confound the effect of connections. Finally, by looking at a subset of provincial leaders who hold provincial leadership positions more than once (so that we can measure their performance in terms of provincial economic growth after the promotion decision has been made), we find suggestive evidence that the promotion pattern we uncover does not distort the allocation of talent within the Communist Party.

In addition to the studies that have already been mentioned, this paper is also related to an emerging literature on the impact of social connections to policy-makers (Fisman, 2001, Khwaja and Mian, 2005 and i Vidal and Leaver, 2011, among others). While these

[7] We also measure connections based on education and birth place, but these measures do not have any significant correlations with promotion. See Section 4.5.4.

[8] For provincial economic growth, exploiting its exogenous variation does not really help us answer the question of whether performance affects promotion. An exogenous shock in economic growth is, by definition, beyond the control of provincial leaders, and its effect on promotion will reflect whether top leaders in China are fooled (see Bertrand and Mullainathan, 2001 for such a study in the context of CEO pay), a question that might be of less importance.

studies benefit from more convincing causal inference than our study, they do not look at the impact on political outcomes as we do.

In the context of China, Persson and Zhuravskaya [2012] find that provincial leaders who rule their native province invest less in infrastructure and spend more on education and health than those not born in the province they rule. They interpret this finding as the effect of stronger social connections between native provincial leaders and other provincial political elites. The focus of our paper, on the other hand, is on provincial leaders' connections to top leaders in the center. Shih [2004, 2008] gauges social connections of provincial leaders in a similar way to ours, but investigates the impact of connections on the loan-to-deposit ratio of each province, interpreted as the support from the central government (as major banks are under the control of the central government). We focus on the impact of connections on political outcomes instead of policy outcomes.

The next section briefly introduces relevant aspects of the Chinese political system. Section 4.3 then lays out a simple theoretical framework describing how connections and performance may interact with each other to affect the probability of promotion. Section 4.4 describes the data and the empirical strategy. Section 4.5 reports the empirical results and various robustness checks. Section 4.6 discusses our results and provides suggestive evidence on potential distortions in the allocation of talent. Section 4.7 concludes the paper.

4.2 Institutional Background

China's highest decision-making body is the Politburo Standing Committee (PSC) of the Communist Party of China. It consists of seven (or nine during 2002–2012) party officials including the General Secretary (the head of the Communist Party) and the Premier (the head of the Chinese government). Although the procedure of its decision-making is not public information, it is commonly believed that members of the PSC meet once a week and make decisions by consensus (Shirk, 1993, Miller, 2004 and McGregor, 2010). Among other things, the committee decides on which provincial leaders to promote.

Each of the 31 provinces of China (including four municipalities with provincial status and five autonomous regions) has two political leaders: provincial secretary and provincial governor. The former is the head of the provincial branch of the Communist Party, and the latter is the head of the provincial government. Provincial secretaries are ranked equally to ministers in the central government, and ranked higher than provincial governors. By "provincial leaders", we mean these two political leaders in each province.

Provincial leaders in China have a large discretion over economic policies at the provincial level while their career prospects are controlled by the central government.[9] Provincial

[9] Xu [2011] refers to this institutional feature as a *regionally decentralized authoritarian system*. Although it is interesting to study the coordination and competition between secretaries and governors, we

economic performance, specifically real GDP growth, is considered to be a performance measure highly relevant to the PSC's promotion decision (Maskin et al., 2000; Blanchard and Shleifer, 2001).

Promoted provincial leaders may join the Politburo of the Communist Party (the second most powerful decision-making body after the PSC) and/or become Vice-premier or State Councilor, the highest positions in the central government after the Premier. In other words, promoted provincial leaders become close colleagues of the PSC members. Promoting political enemies may thus threaten the power of the PSC members.

In the next section, we present a simple model that incorporates these institutional aspects of Chinese politics.

4.3 A Simple Model

Since the previous literature ignores the interplay of connections and performance in the promotion decision for Chinese provincial leaders, we propose a theoretical framework to demonstrate how such interplay may emerge. In this framework, the PSC acts as a unitary player. Social connections to a PSC member play either of two roles in promotion. First, connections foster loyalty of provincial leaders to PSC members, implying that the probability of the PSC to survive in office is higher if the PSC promotes a connected provincial leader. Second, connections provide PSC members with information on the ability of provincial leaders. We argue that the loyalty-fostering role of connections implies complementarity between connections and performance while the informational role implies substitutability.

Of course, other mechanisms may explain complementarity or substitutability of connections and performance. In the Appendix (Section A.1), we model promotion as an incentive scheme for provincial leaders to boost growth. That model predicts the interdependence of connections and performance as long as the PSC can commit to the promotion scheme, which may be a strong assumption. Bargaining among PSC members with conflicting interests may also yield complementarity or substitutability of connections and performance. However, we know very little about the actual bargaining process of the PSC. In addition, during the sample period of our data (1993–2009), the membership composition of the PSC only changed three times, not enough to test the implications of a bargaining model.

Although we do not claim that the framework in this section accurately portrays the mechanism at work for the promotion of Chinese provincial leaders, we think it is the simplest one to illustrate the interdependent roles of connections and performance.

---

leave this to future research.

4.3.1 Model

Consider a simplified version of the standard career-concern model (Holmstrom, 1982), where the PSC is the only strategic player and unitarily decides whether or not to promote a provincial leader i. Provincial leaders are assumed to be non-strategic: they simply produce provincial economic growth out of their talent. As we show in the Appendix (Section A.2), the theoretical results are mostly robust to a strategic provincial leader whose effort affects provincial growth as in the standard career-concern model.

The unitary PSC derives its utility from the rent obtained by being in office and the ability of the promoted provincial leader i, denoted by R and a_i, respectively. For simplicity, we assume the following functional form:

$$u(R, a_i) = R + \eta a_i$$

where parameter η measures the extent to which the PSC prefers promoting an official with higher ability (i.e. meritocracy).

Ability, a_i, is assumed to be unobservable to the PSC. However, provincial economic growth since i starts ruling the province, denoted by g_i, is observed and determined by:

$$g_i = a_i + \varepsilon_i, \tag{4.1}$$

where ε_i is normally distributed with mean zero and variance σ_ε.

We now introduce connections into this framework. Let $C_i \in \{0, 1\}$ be an indicator that takes the value of one if i is connected to the PSC and zero otherwise. The loyalty-fostering role of connections is represented by assuming that the probability for the PSC of staying in power depends on C_i, where this probability, denoted by $p(C_i)$, satisfies $p(1) > p(0)$. Unconnected, and thus disloyal, officials are more likely than connected ones to attempt to oust other PSC members, if they are promoted.

The informational role of connections is modeled as follows. The provincial leader's ability, a_i, is unobservable to the PSC but known to be normally distributed with mean $\bar{a}$ and variance $\sigma_a(C_i)$. If connections inform the PSC of the ability of provincial leaders, the precision is higher: $\sigma_a(1) < \sigma_a(0)$.[10]

The expected utility of PSC from promoting i with connection status C_i and growth performance g_i, which we denote by $W_i^{C_i}$, is then written as follows:

$$W_i^{C_i} \equiv E[u(R, a_i)|g_i] = p(C_i)[R + \eta E(a_i|g_i)], \tag{4.2}$$

where we normalize the payoff of being ousted from office to zero.

If the PSC does not promote i, its payoff is given by $\bar{u}$, which may represent the

[10] As discussed below, allowing connections to affect the mean ability does not change the nature of the interaction between connections and performance in determining the promotion probability.

payoff of promoting the most able official in the central government or leaving the high-office position vacant. Provincial leader i is promoted if $W_i^{C_i} \geq \bar{u}$. Assuming that $\bar{u}$ is distributed by its cumulative density function $F(\bar{u})$, the probability of promotion for i is given by $F(W_i^{C_i})$. For simplicity, we assume that $\bar{u}$ is uniformly distributed with the probability density μ.[11]

4.3.2 Analysis

From equation (4.1) and the distributional assumptions on a_i and ε_i, $E(a_i|g_i)$ is given by the weighted average of g_i and $\bar{a}$ with the weights being the relative precision of growth and ability:

$$E(a_i|g_i) = h(C_i)g_i + (1 - h(C_i))\bar{a}, \tag{4.3}$$

where

$$h(C_i) \equiv \sigma_a(C_i)/(\sigma_a(C_i) + \sigma_\varepsilon).$$

Note that $\sigma_a(1) < \sigma_a(0)$ implies $h(1) < h(0)$. That is, the informational role of connections makes the precision of the prior on i's ability higher.

Hence, the marginal increase in the promotion probability with respect to economic growth g_i is:

$$\frac{\partial F(W_i^{C_i})}{\partial g_i} = \mu\eta p(C_i)h(C_i). \tag{4.4}$$

If this expression is larger for $C_i = 1$ than for $C_i = 0$, connections and growth complement each other in increasing the promotion probability. If it is smaller for $C_i = 1$, connections and growth are substitutes. If it is the same irrespective of C_i, the effects of connections and growth on promotion are independent of each other.

Inspecting Equation (4.4) yields the following proposition.

Proposition 1. *The effects of connections and growth on promotion are:*

1. *Independent if*

   (a) $\eta = 0$ *or* $h(C_i) = 0$ *(i.e.* $\sigma_\varepsilon = \infty$*). In this case, provincial growth does not affect the promotion probability, and the PSC prefers promoting connected officials as long as* $p(1) > p(0)$*.*

   (b) $\eta > 0$*,* $h(C_i) > 0$*, but* $p(1)h(1) = p(0)h(0)$*. In this razors-edge case, the promotion probability increases with growth, but not with connections.*

[11] As shown in Appendix Section A.2, the functional form for $F(\bar{u})$ does not matter in deriving the interdependence of connections and growth in the promotion probability as long as one of the two roles of connections dominates the other. If the two roles of connections are similarly important, whether connections and growth are complementary or substitutes depends on the level of the observed growth rate, g_i.

2. Complementary if $\eta > 0$ and

$$\frac{p(1)}{p(0)} > \frac{h(0)}{h(1)}. \tag{4.5}$$

3. Substitutes if $\eta > 0$ and

$$\frac{p(1)}{p(0)} < \frac{h(0)}{h(1)}. \tag{4.6}$$

Proof. Substitution of parameter conditions into expression (4.4) and comparison between $C_i = 1$ and $C_i = 0$ trivially prove the statements in the proposition. □

Proposition 1 shows that the interplay of connections and growth in affecting the promotion probability requires three conditions: (1) the PSC cares about the ability of promoted officials, (2) provincial growth contains a signal on the ability of provincial leaders, and (3) connections play at least one of the two roles specified in this model.[12] The proposition also shows that if the loyalty-fostering role of connections (represented by $p(1)/p(0)$) is relatively larger (smaller) than the informational role of connections ($h(0)/h(1)$), the responsiveness of the promotion probability with respect to growth is higher (lower) for connected provincial leaders.

4.3.3 Discussion

The Appendix (Section A.2) discusses several extensions of the above model. First of all, the probability of staying in office, $p(C_i)$, may decrease in a_i if competent officials threaten the survival of top leaders.[13] On the other hand, connections may mitigate this threat by fostering loyalty. In this case, the result is robust as long as $p(1)/p(0)$ is sufficiently larger or smaller than $h(0)/h(1)$.

Second, provincial leader i may be able to boost the economic growth of his province by exerting effort (as in a standard career-concern model). As long as ability and effort affect growth additively, however, such strategic behavior of provincial leaders does not alter our result.

Finally, the role of connections in the above model may seem restrictive in several ways. Average ability, $\bar{a}$, for example, may depend on C_i. Connected provincial leaders may have a higher $\bar{a}$ if the PSC can screen out less able candidates for provincial leader-

[12] In the unlikely case in which the loyalty-fostering role of connections exactly cancels the informational role of connections to satisfy $p(1)h(1) = p(0)h(0)$, the interdependency of connections and growth disappears (case 1 (b) in the Proposition).

[13] Egorov and Sonin [2011] argue that the tradeoff between loyalty and ability creates a dilemma for the dictator when choosing high officials. See also Besley et al., 2012 in the context of the choice of electoral lists by political party leaders in Sweden.

ship positions among those connected. However, as expression (4.4) does not contain $\bar{a}$, allowing connections to influence $\bar{a}$ does not change the above result.[14]

The variance of the stochastic shock in the growth process, σ_ε, may also be smaller for connected provincial leaders. Because of their loyalty to the PSC, connected provincial leaders may have a larger incentive, or feel more obliged, to tell the PSC the truth about what has happened to the economy of their province beyond their control. As discussed in the Appendix, this role of connections makes condition (4.5) more likely to hold, even if $p(1) = p(0)$. We are unable to distinguish this mechanism from the political survival mechanism (i.e. through $p(C_i)$ in the model) in the empirical analysis below.

In summary, the above analysis shows that the effects of connections and growth on promotion can be intertwined, an insight that has been ignored in the previous literature. In the empirical analysis to follow, we investigate whether the promotion probability indeed responds to the interaction of connections and growth.

4.4 Data and Empirical Strategy

Our main data source is China Vitae [2012], a website run by a non-profit organization in the United States. It publishes curriculum vitae (CV) of Chinese Communist Party officials who have held important positions since late 1992. The CV includes the year of birth, the province of birth, colleges attended, and, most importantly, the list of positions held in the party or in the government (including state-owned enterprises) in the past, along with the period in which each position was held.

We first explain how our sample of provincial leaders is selected and then explain how we use their CV to measure promotion and connections. We also discuss the data on provincial economic growth and present summary statistics. After presenting the data, we explain our main empirical strategy.

4.4.1 Sample

We focus on provincial secretaries and governors, who hold office for at least twelve months in between June 1993 and June 2009.[15] There are 275 provincial leadership spells (137 secretaries and 138 governors) that satisfy these criteria. From this set of spells, we drop 17 (seven secretaries and ten governors) whose CV is not available in China Vitae

[14] For general distribution functions of $\bar{u}$, the result is robust as long as $p(1)/p(0)$ is sufficiently larger or smaller than $h(0)/h(1)$. See Appendix Section A.2.

[15] We start from 1993 because China Vitae [2012] does not cover officials in office in June 1992 or before. Our sample period ends in 2009 as annual growth data is available up to 2009 when the first draft of this paper was written. We look at June because, following Li and Zhou [2005], we measure the promotion outcome during the period between July of year t and June of year $t + 1$, which is to be matched with economic growth in year t. Finally, we drop provincial leaders whose tenure is less than twelve months because the promotion of such leaders is unlikely to be associated with annual provincial growth.

[2012].[16] Since some officials assume a provincial leadership position more than once, the total number of officials in the sample is 187, i.e. less than the 258 observed leadership spells.

4.4.2 Promotion

Following Li and Zhou [2005], we define the promotion of a provincial secretary as becoming a member of the Politburo (the second highest decision-making body in the Communist Party, consisting of 20 to 25 members that include all members of the PSC), a Vice-Premier or a State Councilor in the central government.[17] A provincial governor is promoted if he becomes a secretary of the same or a different province. There is no instance where a provincial governor joins the Politburo or becomes a Vice-Premier or a State Councilor, suggesting that the two sets of provincial leaders compete for separate higher offices.

Assuming other positions in the central government such as vice-chairmanship of the Chinese parliaments (National People's Congress and Chinese People's Political Consultative Conference) and, for governors, the head of a ministry could perhaps also be seen as promotion (see Tao et al., 2010). We investigate the robustness of our results to these wider definitions of promotion.

4.4.3 Connections

We measure the connection between a pair of party officials by whether they used to work in the same branch of the Party or of the government at the same time. In particular, we focus on links between each provincial leader and any current member of the PSC, given that the PSC is in charge of the decisions on the promotion of provincial leaders.[18] The connection indicator can thus change for the same provincial leader if his connected PSC member resigns or if his connected official joins the PSC.[19] In our sample period, we find

[16] There is little attrition bias in terms of provincial economic growth: the difference in the average annual provincial growth since assuming office is 0.7 percentage points (not statistically different from zero), 6% of the whole sample mean, between provincial leaders with and without an available CV, conditional on province and year fixed effects.

[17] Some provincial secretaries join the Politburo without leaving office. If this happens, we again follow Li and Zhou [2005] by recording this as their promotion and treating the rest of the leadership spell as a separate one. There are three such cases. When a provincial secretary with the Politburo membership leaves office, joining the PSC is defined as promotion.

[18] We also consider links between each provincial leader and any current members of the Politburo, to see if connections to the PSC may pick up the effect of being connected to the political elite more generally.

[19] Potentially we can exploit this within-individual variation to identify the impact of connections, as in Jia [2013], because membership changes in the PSC are plausibly exogenous. Out of 187 officials in the data, however, only 25 change their connection status during their tenure. When we estimate the stratified Cox proportional hazard model where the stratum is each individual (so we can control for unobserved heterogeneity across individuals), we obtain results with very large standard errors.

that connections between PSC members and provincial leaders are mostly formed at the provincial branches of the Party and of the government, but also at central bodies such as ministries, the National People's Congress and the Communist Youth League.[20]

The focus on workplace-based connections among Chinese politicians is motivated by the anecdotes mentioned in footnote 4: Jiang Zemin and Hu Jintao promoted their former colleagues in Shanghai and the Communist Youth League, respectively, once they became General Secretary. However, the literature on informal politics in China (see Dittmer, 1995, for example) also points out the importance of other sources of connections among politicians, such as graduating from the same college or hailing from the same province. We investigate whether these alternative sources of connections are also of importance.

An estimation of the effect of connections defined in this way should be seen as an "intention-to-treat" analysis, using the language of the program-evaluation literature. Having worked together in the past does not necessarily mean being loyal to each other or being well informed about each other's ability. However, it is plausible that having worked together increases the probability of being loyal to and/or familiar with each other. As a result, if we do not find connections significantly correlated with promotion, we should not interpret this to say that connections are unimportant.

4.4.4 Economic Growth

The data on provincial annual real GDP growth up to year 2009 is obtained from the National Bureau of Statistics of China [2009, 2011].

One may question the reliability of the provincial GDP growth data, given the possibility that higher growth increases the chance of promotion for top provincial politicians. The central government of China ensures the reliability of provincial GDP data in two ways.[21] First, each provincial government is required to submit the figures for various subcomponents of GDP. The National Bureau of Statistics (NBS) in the central government then double-checks the total GDP figure by aggregating these subcomponents on its own. Second, the NBS conducts its own survey to obtain its own estimates of provincial GDP. Therefore, even though provincial leaders may have an incentive to overreport the growth of their provincial economy, the provincial GDP data should reflect the actual performance of the economy to a large extent.[22]

[20] We also find that the difference in position ranks for connected pairs (each position of the Party and the government has an official rank) is usually no more than two when they worked together.

[21] We thank Li-An Zhou for providing us with this information.

[22] We corroborate the quality of the provincial growth data by checking if it reflects the growth in nighttime lights observed by satellites. Nighttime lights can be seen as an objective measure of living standards, and therefore its correlation with GDP is indicative of the quality of GDP data. Using data on nighttime lights from National Geophysical Data Center [2010] and on Chinese provincial boundaries form Natural Earth [2012], we follow Henderson et al. [2012] in measuring and aggregating nighttime lights to the provincial level. The correlation coefficient (conditional on province and year fixed effects) between annual GDP growth and annual light growth is about 0.1, significantly different from zero at the

4.4.5 Summary Statistics

Column 1 of Table 5.1 reports summary statistics for the variables used in the following analysis at the level of 258 leadership spells. Columns 2 and 3 restrict the sample to provincial secretaries and governors, respectively. About a quarter of the provincial leadership spells end with promotion. The promotion rate is lower for secretaries, consistent with the fact that secretaries are more highly ranked than governors in the Communist Party hierarchy. The share of spells with the provincial leader connected to PSC members for at least one year is about a quarter, with a slightly higher share for secretaries. The mean of average annual provincial real GDP growth since assuming office is around eleven percentage points. Each leadership spell lasts 4.3 years on average, consistent with the fact that the Communist Party makes major personnel decisions every five years when the Party Congress is held.

In the empirical analysis below, we construct a leader-year level sample in which each leadership spell is observed annually until the leader is transferred to another position, irrespective of whether it is a promotion or not. This process results in 966 observations. The summary statistics for this sample is reported in column 4 of Table 5.1.

Figure 4.1 shows the rate of promotion (the line graph) and the distribution of tenure length (the bar graph) by the number of years in office. It shows that the chance of promotion increases until the fifth year in office and declines thereafter. The peak at five years is expected, since many of the personnel decisions are made at the National Congress of the Communist Party which is held every five years. The figure also shows that the majority of leadership spells ends in five years or less.

4.4.6 Empirical Strategy

The structure of the leader-year level data suggests using a competing risks model, in which observations exit from the data through more than one type of event (promotion and non-promotional transfer in our case). However, below we treat non-promotional transfers as right-censoring by assuming that non-promotional transfers occurs randomly. Furthermore, we use a linear probability model of promotion, instead of a Cox proportional hazard model, by assuming that the duration of each leadership spell does not depend on connection status and growth. Although these assumptions are restrictive, the linear probability model allows us to control for unobservable heterogeneity across provinces and years that can differ between secretaries and governors, which we believe is important in our context to minimize the bias in the estimation.

Therefore, to investigate how the promotion of provincial leaders is correlated with

5% level. This evidence suggests that real GDP growth, at least to some extent, reflects improvements in some dimensions of living standards captured by nighttime lights.

their connections and performance, we estimate the following linear probability model:

$$P_{iopt} = \alpha C_{it} + \beta(G_{iopt} - \bar{G}) + \gamma C_{it} * (G_{iopt} - \bar{G}) + \boldsymbol{x}'_{iopt}\boldsymbol{\delta} + (G_{iopt} - \bar{G}) * \boldsymbol{x}'_{iopt}\boldsymbol{\xi} + \sum_{\tau=2}^{12} \kappa_\tau T^{\tau}_{iopt} + \mu_{op} + \eta_{ot} + \varepsilon_{iopt} \quad (4.7)$$

The dependent variable, P_{iopt}, is the indicator that takes the value of one if provincial leader i in office o (secretary or governor) in province p is promoted in the period from July of year t to June of year $t+1$.[23] We have three regressors of interest. The first is C_{it}, the indicator of leader i being connected to the members of the PSC in office in June of year t. The second is G_{iopt}, the average annual growth rate of province p since leader i assumed office o until year t, measured as the deviation from the sample mean, $\bar{G}$ (11.4%). The last is the interaction term of these two variables. To facilitate the interpretation of the coefficient on the connection indicator, β, the growth variable is demeaned so that β measures the difference in the probability of promotion between connected and unconnected officials displaying average growth performance. For robustness checks, we control for $\boldsymbol{x}_{iopt}$, a vector of characteristics of provincial leader i and province p in year t, and its interaction with $(G_{iopt} - \bar{G})$, to investigate whether the endogeneity of C_{it} is driving our main results.

Province and year fixed effects are allowed to differ between secretaries and governors (μ_{op} and η_{ot}). Controlling for province fixed effects ensures that the coefficients of interest (α, β, γ) do not pick up the possibilities that certain provinces which grow more quickly also have their leaders more likely promoted and connected. Controlling for year fixed effects incorporates the possibility of relative performance evaluation, often discussed in the literature on Chinese political selection (e.g. Maskin et al., 2000, Chen et al., 2005). It also allows for higher promotion rates in Party Congress years, in which many personnel decisions are made.

Since the promotion probability changes non-linearly with the number of years in office, as suggested by Figure 4.1, we also control for a set of dummies for the number of years in office from two to twelve (T^{τ}_{iopt}).[24] Standard errors are clustered at the province level given that both growth rates, G_{iopt} and the error term are likely to be serially correlated within each province, and C_{it} tends to take the same value for the same province for a certain number of years. As the low number of Chinese provinces (31) may cause an underestimation of the standard errors, even with clustering, we also report p-values on the significance of the estimated γ by using the wild cluster bootstrap-t procedure (Cameron et al., 2008).

The coefficient on the interaction term of connections and growth, γ, is negative if the

[23] See footnote 15 for why we measure promotion during the period from July.

[24] These dummies may be endogenous. However, the results do not differ significantly if we do not control for these dummies.

two determinants of promotion are substitutes and is positive if they are complements.

4.5 Empirical Results

We present our empirical results in three steps. First, we show our baseline estimates. Second, we discuss the endogeneity concerns of connections and growth. Third, we present various robustness checks and additional results.

4.5.1 Baseline Estimates

Table 4.2 shows our main results from estimating equation (4.7) by adding regressors of interest one by one without controlling for any covariates ($\boldsymbol{x}_{iopt}$). The first three columns estimate the correlations of promotion with connections and growth without introducing their interaction term, thus making them directly comparable to previous studies. Column 1 shows that connected provincial leaders are more likely to be promoted. The difference is estimated at 4.9 percentage points – nearly 60% of the average promotion rate – and statistically significant at the 5% level. This finding confirms anecdotal pieces of evidence mentioned in the introduction. It is also in line with recent evidence by political scientists (Shih et al. 2012). In column 2, we find that the promotion probability increases with the provincial economic growth during the tenure, although it is not significantly different from zero. The point estimate suggests that one standard deviation increase in growth (by 2.4 percentage points) pushes up the probability of promotion by 1.7 percentage points. This magnitude is comparable to the one estimated by economists (Li and Zhou, 2005).[25] Column 3 includes both the connection indicator and provincial GDP growth as regressors. The results are similar to those in the previous columns, suggesting that the connection status is largely orthogonal to provincial GDP growth once province and year fixed effects are controlled for.

Column 4 presents our key finding by including the interaction term of connections and growth as a regressor. The coefficient on the interaction term is positive and statistically significant at the 5% level (or at the 10% level if we use the p-value by Cameron et al., 2008), suggesting that the two determinants of promotion are complements rather than substitutes. The point estimate indicates that a one standard deviation increase in growth raises the promotion probability by 5.3 percentage points more for connected officials than for unconnected ones. The coefficient on the connection indicator suggests that the promotion rate for provincial leaders with the sample average growth is 3.2 percentage points higher for those connected than for those unconnected, although this difference

[25] The result of Li and Zhou [2005] suggests that a one standard deviation increase in growth raises the promotion probability by 1.8 percentage points and is statistically significant from zero at the 1% level. The main differences to our estimates are that: their sample spans the period from 1979 to 2002, and they do not cluster standard errors at the province level.

is not statistically significant. The growth effect for unconnected officials is insignificant although the point estimate suggests that a one standard deviation increase in growth increases the probability of promotion by 0.8 percentage points, which is more than 10% of the sample mean promotion rate.

Figure 4.2 shows this main result graphically. We first regress both the promotion dummy and provincial growth since assuming office on dummies of the numbers of years in office, province-office fixed effects, and year-office fixed effects, and obtain the residuals from these regressions. Then, we divide the observations into tertiles according to the residual growth, irrespective of connection status. Finally, for each tertile, we plot the average residual promotion rate by connection status. We also use the bar graph in the background to show the distribution of connected observations across growth tertiles.

The figure shows that the complementary result is entirely driven by a large difference in the promotion rates between connected and unconnected provincial leaders among the top-third of growth performers. For unconnected provincial leaders, the worst third performers are slightly less likely to be promoted than the rest. In terms of the distribution of growth performances, connected officials are most likely to be in the middle tertile, but do not disproportionately perform better or worse than their unconnected peers. We discuss possible interpretations of this graphical result in Section 4.6.1 below.

Our key finding implies that the previous literature fails to recognize the importance of the interplay between connections and growth in determining the promotion of provincial leaders, and perhaps of Chinese Communist Party officials in general. Connected officials do have a higher likelihood of promotion on average, but this is solely due to their promotion probability being more responsive to performance. Put differently, officials with a better performance appear to be promoted more often, but this relationship mainly applies to those connected to top political leaders of China.

In terms of the theoretical framework introduced in Section 4.3, our results suggest that connections have more to do with fostering loyalty of provincial leaders to top leaders than with information about their ability.

4.5.2 Endogeneity of Connections and Growth

Endogeneity of Connections The connection status of provincial leaders may certainly be endogenous to their promotion probability. Table 4.3 compares the means of observable characteristics of provincial leadership spells between those connected and those unconnected, where the connection status is measured at the first year of the spell (as changes in the connection status during the tenure are due to the membership shuffling of the PSC and thus more likely to be exogenous). Connected leadership spells end with promotion 9 percentage points more often than unconnected ones. The average annual growth at the end of the spell is significantly higher for connected leaders. The length of tenure is slightly shorter for the connected.

In terms of individual characteristics, connected officials are significantly younger and more likely to have served in the central government. In terms of provincial characteristics, the provinces ruled by connected officials are more likely to have higher economic growth in the five-year period before the officials assume office, less likely to be the home province for provincial leaders, and more likely to be the one in which the current members of the PSC used to work.

However, many of these differences can be explained by differences in provinces and time periods when connected and unconnected leaders are in office. Column (4) reports the estimated coefficient on the connection indicator from regressing each of these variables on the connection indicator, office-by-province fixed effects, and office-by-year (where the year refers to the one when each leader assumes office, ranging from 1983 to 2009) fixed effects. Except for the age when assuming office, the indicator of having served in the central government, and whether the current PSC members used to work in the province, the connection coefficient is not significantly different from zero. In Section 4.5.3 below, we control for the aforementioned variables to check whether the connection status picks up the effect of observable differences between connected and unconnected provincial leaders.

However, these observable characteristics may not reflect how much support provincial leaders obtain from the central government due to their connection to the PSC members. PSC members may help connected provincial leaders achieve high growth so that they can promote them as if the decision were based on meritocracy. We do not observe all dimensions of the support by the central government to each province, but we do have the data on fiscal transfers from the center to each provincial government annually since 1994 (China Financial & Economic Publishing House, various years). The last row in Table 2 reports the difference in average annual fiscal transfers during the term between connected and unconnected provincial leaders. An unconditional comparison suggests that connected provincial leaders do obtain significantly more fiscal transfers from the central government. But, once we take into account province and year fixed effects, this difference is no longer significantly different from zero.

Endogeneity of Growth The provincial GDP growth data may not reflect the performance on basis of which the promotion decision is taken. For example, provincial growth may be higher for those who are promised promotion, because the central government offers support to them to boost the economic growth of their province so that their promotion will look merit-based. If so, we should see a significant drop in economic growth in the province after its leader has been promoted. Columns 1 and 2 of Table 4.4 investigate this issue. With balanced panel data of provinces for 1993-2009, we regress annual real GDP growth on indicators for one, two, and three years after promotion as well as on province and year fixed effects. These indicators are defined based on the promotion of provincial secretaries in column (1) and governors in column (2). The estimated coefficients on these indicators are, however, insignificant and positive in most cases.

Provincial growth may also reflect the strength of connections. Our measure of connections does not necessarily reflect the actual connections. PSC members may provide support to boost the economy only to those provincial leaders who are actually connected to them. Although we cannot entirely dismiss this possibility, we can use observable characteristics of connections that may be correlated with the strength of connections, to check if real GDP growth is higher for provinces whose leader's connection to the PSC is stronger. Columns (3) to (6) of Table 4.4 run provincial panel regressions of the following form:

$$g_{pt} = \phi C_{pt} + \xi C_{pt} * Z_{pt} + \psi_p + \omega_t + \varepsilon_{pt},$$

where g_{pt} is the annual real GDP growth of province p in year t, C_{pt} the indicator that the leader in province p in year t is connected to the PSC member(s), Z_{pt} the strength of connections of the leader in province p in year t, ψ_p the province fixed effect, and ω_t the year fixed effects. We use two variables to measure Z_{pt}. First, we measure the number of years that the provincial leader has worked together with his connected PSC member(s), assuming that a longer time together strengthens connections.[26] Second, we take the difference in ages between the connected pair by subtracting the age of the PSC member from the provincial leader, assuming that a larger age difference strengthens connections. Different generations of party officials do not compete with each other for power due to the seniority system of promotion.[27] We run this regression separately for provincial secretaries and provincial governors, the former reported in columns (3)-(4) and the latter in (5)-(6).

Columns (3) and (5) measure the strength of connections by the number of years working together. While this interaction term is insignificant for provincial secretaries, growth is significantly higher for provincial governors whose connection is stronger by this measure. Plotting the data, however, reveals that this result is driven by one observation (Fujian in 1993) where the governor has worked with a PSC member for 14 years (the maximum in the sample) and the provincial economy grew by 22.6%, almost twice the sample mean. If we drop this observation, however, there is no substantial change in our main results.[28]

---

[26] If the provincial leader has more than one connection, we take the average. The results do not substantially change if we take the maximum or the minimum.

[27] It might also be the case that the connections are stronger if the connected pair of individuals is more similar in age. To reflect this possibility, we also measure the age difference in absolute value. The results are similar.

[28] The coefficient on the interaction of connection and growth for column (4) of Table 4.2 is reduced

Columns (4) and (6) use the age difference variable (which is positive if the provincial leader is older than his connected PSC members) as a measure of the strength of connections. Its interaction term with the connection indicator is insignificant both for provincial secretaries and governors.[29]

These results suggest that provincial growth is not higher for those provincial leaders whose connections may be stronger in terms of observable dimensions, encouraging the interpretation of growth as a performance measure, rather than a measure of the strength of connections.

4.5.3 Robustness Checks

Definitions of Promotion The first set of robustness checks on our key finding concerns the definition of promotion. Tao et al. [2010] suggest that the definition of promotion of provincial leaders should include three additional appointments. Thus, we broaden the definition of promotion step by step in the final three columns in Table 4.2. Column 5 changes the definition of promotion so that becoming a minister is also regarded as promotion for provincial governors. Ministers are officially ranked equal to provincial secretaries. As governors are ranked below secretaries in each province, becoming ministers can be seen as promotion for governors. Seven additional leadership spells end with promotion in this definition.

Column 6 further changes the definition of promotion, including appointments of both secretaries and governors to become vice-chairmen of the National People's Congress (the lower house of the Chinese parliament). Four additional leadership spells are then coded as promotions. In column 7, six more cases where provincial leaders become vice-chairmen of the CPPCC (the upper house of the Chinese parliament) are also coded as promotion. Officially, these positions are ranked higher than provincial leadership positions. Due to the nature of parliaments in the Chinese political regime, they can also be regarded as ceremonial.

Our finding of the complementarity of connections and growth is robust to these different definitions of promotion, with estimated coefficients fairly stable across definitions.

Individual and Province Characteristics Table 4.5 conducts a series of robustness checks by controlling for individual or province characteristics and their interactions with (demeaned) growth. Column 1 controls for the age of each provincial leader when assuming office (this variable is thus time-invariant for each leadership spell). Since connected provincial leaders are significantly younger (see Table 4.3), the connection indicator may

by one-fifth in size, but it remains significant at the 5% level.

[29] Since the age difference is correlated with the provincial leader's own age, we also run regressions where we control for the age and its interaction with the connection indicator. The coefficients on the age difference interacted with the connection indicator remain insignificant.

pick up the effect of the leader's age. Columns 2 and 3 control for the indicator of having served in the central government before assuming provincial leadership and the indicator of ruling the native province, respectively. Connected provincial leaders may be on the elite career track that includes positions in the central government and various provinces while unconnected ones may tend to rule their native province only. Column 4 controls for provincial growth over the five-year period before assuming leadership. The growth performance of connected officials may reflect the skills to boost the stagnant provincial economy (if connected leaders are assigned to slow-growing provinces) or the ability to rule politically important provinces (if the growth performance of a province, either good or bad, indicates its importance for the central government). Column 5 controls for the indicator of provincial leaders being a princeling, the son or son-in-law of a prominent Communist Party official. Princelings are known as a powerful faction within the Party.[30] Connections to the PSC members may simply reflect the political advantage of being a princeling. Columns 6 and 7 check if connected leaders are simply assigned to provinces that PSC members have a great deal of knowledge about. If this is the case, growth is a stronger signal of the leader's ability, thus explaining the stronger responsiveness of the promotion rate to growth. To measure this feature, we use two variables: an indicator of provinces where current members of the PSC used to work in column 6, and an indicator of provinces where the current members of the PSC were born in column 7. Finally, column 8 controls for all these variables and their interaction with growth. In all columns, the estimated coefficients on the connection indicator and its interaction term with growth change little from the estimates in column 4 of Table 4, suggesting that the main result is not driven by these omitted variables.

Being the Political Elite Table 4.6 checks whether our measure of connections simply reflects being a member of China's political elite. A certain set of workplaces may be the home of every top leader in China. As a result, those destined for promotion have worked with the current top leaders in the PSC, and this has nothing to do with social connections between PSC members and provincial leaders.

We check this possibility in two ways. In column 1, we include a measure of connection between provincial leaders and past or future PSC members rather than current members. In column 2, we instead add an indicator for provincial leaders who used to work in the same place as current PSC members but in a different period. They should thus pick up the effect of the elite status, but not of connections. These dummies are set to zero if our main connection indicator is one. Therefore, the coefficients on these variables and their interaction with growth will be the same as those on the connection indicator if

[30]The data on princelings come from China Vitae [2012]. Xi Jinping, General Secretary since 2012, is a princeling. The media often reported factional struggles between princelings and the former members of the Communist Youth League headed by the outgoing General Secretary Hu Jintao in the lead-up to the 2012 Party Congress.

connections simply reflect the elite status.

Table 4.6 shows that the coefficients on these additional regressors are not significantly different from zero. The F-test rejects the null that the effect of being connected to the current members of the PSC is the same as that of being connected to the past or future members of the PSC (column 1) or of working in the same place as current PSC members in a different period (column 2) at the 5% and 10% levels, respectively. The result in column 1 also suggests that our main result cannot be explained by the effect of being trained on the job by a mentor who later joins the PSC.

In column 3, we look at the impact of being connected to current members of the Politburo, the second highest decision-making body in the Communist Party. Connections to the PSC members may simply reflect connections to top politicians in general. If so, connections to the Politburo should also be of importance. However, coefficients on the Politburo connection indicator (set to be zero if the PSC connection indicator is one) and its interaction with growth are not significantly different from zero, and the F-test rejects the equality of these coefficients to those for PSC connections. This result implies that connections to the people with decision-making power count the most.

4.5.4 Other Connection Sources

Table 4.7 investigates whether other sources of social connections are of importance for promotion. For this purpose, we replace C_{it} in equation (4.7) with alternative independent variables. In column 1, we use a dummy that equals one if a provincial leader graduated from the same college as a current PSC member within a period of three years before or after. Such provincial leaders may have met a PSC member while in college. In column 2, we use an indicator of provincial leaders having graduated from the same college as a current PSC member at any point in time. Graduating from the same college may reduce the cost of communication, even if they did not attend the college at the same time.[31] In column 3, we define connections as being born in the same province, which may also reduce the cost of communication. Table 4.7 shows that none of these sources of connections have any significant effect on the probability of promotion. One interpretation of these results is that sharing working experience is of greater importance than sharing the same birthplace or knowing each other at college. Another interpretation is that our measures of school and birth place connections are coarser than those based on shared working experience.

[31] Many of the top politicians in China graduated from Tsinghua University, one of the most prestigious colleges in China, and they are known as the Tsinghua clique.

4.5.5 Heterogeneous Impacts by Age Difference

In light of our theoretical framework in Section 4.3, the estimated complementarity between connections and performance reflects the loyalty-fostering role of connections. However, it is also consistent with connections increasing the PSC's marginal benefit from promoting more able provincial leaders in more general senses. To narrow down the number of possible interpretations of our empirical finding, we investigate whether the complementarity result is stronger for connected pairs in which the provincial leader is much younger than his connected PSC member. Since the 1990s, the Communist Party has undergone generational changes of leadership every ten years. This suggests that party officials of a similar age compete with each other for high office, but different generations of officials do not. Thus, it is plausible to assume that provincial leaders show more loyalty towards connected PSC members whose age is a lot higher than their own. If the complementarity between connections and performance in the promotion process is due to the loyalty-fostering role of connections, we expect a stronger complementarity for the connected pairs in which provincial leaders are substantially younger than the PSC members.

As in Section 4.5.2, we obtain the age difference between provincial leaders and their connected PSC members by subtracting the PSC member's age from that of the provincial leader.[32] In our sample, provincial leaders are on average 7.1 years younger than their connected PSC members, with a standard deviation of 5.8 years. We include the interactions of this variable with the connection indicator and with the connection-growth interaction term as additional regressors to equation (4.7). If the age difference makes the complementarity of connections and growth stronger, the coefficient on its interaction with the connection-growth interaction term will be negative.

Table 4.8 reports the results from this estimation. Column (1) shows that the complementarity between connections and performance are indeed stronger for pairs where provincial leaders are much younger than their connected PSC members. The estimates suggest that a one standard deviation larger age difference raises the coefficient on the interaction between connections and growth by 1.44. As this result may be driven by the provincial leader's own age, independent of the PSC member's age, column (2) controls for the provincial leader's age when he assumed office (the same variable as the one used in column (1) of Table 4.5) interacted with the connection indicator and with the connection-growth interaction term. The size of the coefficient on the age difference interacted with the connection-growth interaction term changes little and remains significant at the 10% level. These results imply that the complementarity between connections and performance is indeed due to the loyalty-fostering role of connections rather than to more general benefits of promoting connected officials for the PSC.

[32] If there are two connected PSC members, we take the minimum (i.e. the maximum in absolute terms).

4.6 Discussion

4.6.1 Interpretations of the Graphical Result

As shown in Figure 4.2, it is the top one-third of provincial leaders in terms of performance that entirely drive the estimated complementarity between connections and performance. Our theoretical framework in Section 4.3 can explain this observed pattern by allowing the PSC's survival probability to increase with ability if the promoted official is connected. Specifically, $p(1)$ increases with a_i while $p(0)$ does not. More talented officials may be better at helping top leaders survive as long as they are loyal to them due to their connection.

Figure 4.2 can also be interpreted as an outcome of bargaining among PSC members. Each member of the PSC may want to promote provincial leaders with whom he is connected. However, the PSC makes decision by consensus, as observed by political scientists (e.g. Shirk, 1993). For a provincial leader who is connected to a particular member of the PSC, other PSC members may agree to promote him only when he has achieved high economic growth in his province.

As discussed in Section 4.3, due to the lack of information on the actual bargaining process of the PSC, together with too few changes of the PSC membership during our sample period, we are unable to empirically disentangle these two interpretations. We leave this issue to future research.

4.6.2 Implications on Efficiency

What is the implication of our findings for the allocation of talent? Unfortunately, we do not have any good measure of the ability of Chinese politicians once they leave the provincial leadership positions. However, we can look at provincial leaders with more than one spell: (1) secretaries and governors who get transferred to another province without promotion, (2) governors who are promoted and become secretaries of the same or a different province, and (3) secretaries who are promoted by joining the Politburo without leaving the provincial secretary office (see footnote 17). We observe the performance of these leaders after their promotion or non-promotional transfer, measured by the real GDP growth of the new province in which they assume leadership. Although it is a selected sample of provincial leaders, analyzing this sample sheds some light on whether connected officials are more or less talented than unconnected ones, conditional on the initial performance.

We first regress annual provincial real GDP growth on province and year fixed effects with the full balanced panel data of 31 provinces from 1993 to 2009, and calculate the residuals from this regression. Then, we estimate the following equation

$$\hat{y}_i = \phi C_i + \xi(\hat{g}_i - \bar{g}) + \psi C_i * (\hat{g}_i - \bar{g}) + \omega_i, \tag{4.8}$$

where $\hat{y}_i$ is average conditional annual real GDP growth for official i during his second term, C_i the connection indicator in the last year of i's first term, $\hat{g}_i$ the average conditional annual real GDP growth for i during his first term, and $\bar{g}$ the sample average of $\hat{g}_i$. A few officials also serve a third term. Such a case is treated as one additional observation for i in our sample so that $\hat{y}_i$ refers to the third term and C_i and $\hat{g}_i$ refer to the second term.

If $\phi \leq 0$ and $\psi < 0$, connected officials with more than the average performance during the first term perform worse than unconnected ones with a similar first-term performance, indicating that the promotion pattern that we observe is inefficient. If $\phi \geq 0$ and $\psi \geq 0$, connected officials perform equally or better than unconnected ones if their first-term performance is more than the average, suggesting that promoting connected officials rather than unconnected ones among best performers may indeed be efficient.

In total, 58 officials serve more than one term of provincial leadership between 1993 and 2009, 13 of which serve three terms. In terms of leader-years, these officials account for 453 observations with a promotion probability of 12.4%. This number is higher than that in the main sample (7%) because part of the sample selection criteria requires the promotion from governors to secretaries and from secretaries to Politburo-member secretaries. In terms of performance and connections, this subsample is more or less comparable to the main sample, however. The average growth measured as the deviation from the whole sample average is 0.005 percentage points, and 18.1% of the leader-years are connected.

Column 1 of Table 4.9 replicates our main result by restricting the sample to leader-years served by these 58 officials. With this subsample, connected officials are significantly more likely to be promoted than those who are unconnected if their growth performance is average. For unconnected officials, higher growth reduces the promotion probability although this negative correlation is not significant. However, the complementarity between connection and growth does apply to this subset of provincial leaders.

Column 2 of Table 4.9 reports the result of estimating equation (4.8). Since first-term growth is demeaned, the coefficient on the connection indicator tells us the difference in the second-term conditional growth between those connected and unconnected whose first-term growth is average, and it is not significantly different from zero. The higher the first-term growth, the higher is the second-term growth, with the coefficient being significant at the 10% level. The coefficient on the connection-growth interaction term is positive but not significantly different from zero. Although not conclusive, these results suggest that promoting connected officials with high performance instead of unconnected ones with similarly high performance does not appear to be inefficient.

4.7 Conclusions

The past literature on the promotion of Chinese Communist Party officials looks at the impact of their performance and that of social connections to top politicians separately or assumes no interplay between the two. In this paper, we theoretically show that these two factors can interact, and empirically find that the positive correlation between promotion and performance is robustly stronger for connected officials than for unconnected ones.

Political selection in autocracy often may reflect a trade-off between competence and loyalty (Egorov and Sonin, 2011). Appointing competent officials to high office threatens the power of an autocrat. As a result, incompetent but loyal subordinates tend to surround the autocrat, which is one contributing factor to a poor quality of government in autocracy.

Our evidence might suggest that China avoids this trap. A system of job rotation and promotion within the Communist Party might help pairs of officials build trust by working together. Within a pool of officials with such connections, top officials may then be able to pick the most able without being threatened. In this view, what we may call patronage or nepotism does not necessarily result in an inefficient allocation of talent. We leave testing this hypothesis to future research.

Bibliography

D. Acemoglu, G. Egorov, and K. Sonin. Political selection and persistence of bad governments. *Quarterly Journal of Economics*, 125:1511–1575, 2010.

M. Bertrand and S. Mullainathan. Are ceos rewarded for luck? the ones without principals are. *Quarterly Journal of Economics*, 116:901–932, 2001.

T. Besley. Political selection. *Journal of Economic Perspectives*, 19:43–60, 2005.

T. Besley and M. Reynal-Querol. Do democracies select more educated leaders? *American Political Science Review*, 105:552–566, 2011.

T. Besley, O. Folke, T. Persson, and J. Rickne. Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden, 2012.

O. Blanchard and A. Shleifer. Federalism with and without political centralization: China versus russia. *IMF Staff Papers*, 47, 2001.

A. Cameron, J. B. Gelbach, and D. L. Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics & Statistics*, 90:414–427, 2008.

Y. Chen, H. Li, and L-A. Zhou. Relative performance evaluation and the turnover of provincial leaders in china. *Economics Letters*, 88:421–425, 2005.

China Financial & Economic Publishing House. Finance yearbook of china, various years.

China Vitae, 2012. URL `http://www.chinavitae.com`. Accessed February 17 2012.

L. Dittmer. Chinese informal politics. *China Journal*, 34:1–34, 1995.

M. Edin. State capacity and local agent control in china: Ccp cadre management from a township perspective. *China Quarterly*, 173:35–52, 2003.

G. Egorov and K. Sonin. Dictators and their viziers: Endogenizing the loyalty-competence trade-off. *Journal of the European Economic Association*, 9:903–930, 2011.

C. Ferraz and F. Finan. Motivating politicians: The impacts of monetary incentives on quality and performance, 2011.

R. Fisman. Estimating the value of political connections. *American Economic Review*, 91:1095–1102, 2001.

P. Francois, I. Rainer, and F. Trebbi. How is power shared in africa?, 2012.

V. Galasso and T. Nannicini. Competing on good politicians. *American Political Science Review*, 105:79–99, 2011.

C. Garcia-Jimeno and J. A. Robinson. Prosperity, inequality, and elites: The determinants of political office-holding in nineteenth century antioquia. In A. Meisel and Maria Teresa Ramirez, editors, *Colombian Economic History in the Nineteenth Century*. Bogota: Fondo de Cultura Economica, 2010.

J. V. Henderson, A. Storeygard, and D. N. Weil. Measuring growth from outer space. *American Economic Review*, 102:994–1028, 2012.

B. Holmstrom. Managerial incentive problems: A dynamic perspective, 1982. Helsinki: Swedish School of Economics.

Y. Huang. The industrial organization of chinese government. *Harvard Business School Working Paper 99–076*, 1998.

J. Blanes i Vidal and C. Leaver. Are tenured judges insulated from political pressure? *Journal of Public Economics*, 95:570–586, 2011.

R. Jia. Pollution for promotion, 2013.

A. I. Khwaja and A. Mian. Do lenders favor politically connected firms? rent provision in an emerging financial market. *Quarterly Journal of Economics*, 120:1371–1411, 2005.

P. F. Landry. *Decentralized Authoritarianism in China*. Cambridge University Press, 2008.

H. Li and L-A. Zhou. Political turnover and economic performance: The incentive role of personnel control in china. *Journal of Public Economics*, 89:1743–1762, 2005.

Y. Li. China's political business cycle, 2012.

E. Maskin, Y. Qian, and C. Xu. Incentives, information and organizational form. *Review of Economic Studies*, 67:359–378, 2000.

A. Mattozzi and A. Merlo. Political careers or career politicians? *Journal of Public Economics*, 92:597–608, 2008.

A. Mattozzi and A. Merlo. Mediocracy, 2011.

R. McGregor. *The Party: The Secret World of China's Communist Rulers*. HarperCollins Publishers, 2010.

M. Messner and M. K. Polborn. Paying politicians. *Journal of Public Economics*, 88: 2423–2445, 2004.

A. L. Miller. Hu jintao and the party politburo. *China Leadership Monitor*, 9, 2004. URL http://www.hoover.org/publications/china-leadership-monitor/article/6954. accessed 1 June 2012.

National Bureau of Statistics of China. Comprehensive statistical data and materials for 60 years of new china, 2009.

National Bureau of Statistics of China. China statistical yearbook 2010, 2011.

National Geophysical Data Center. Dmsp-ols nighttime lights time series (version 4), 2010. URL http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html. accessed 20 January 2013.

Natural Earth. 1:10m cultural vectors admin 1 - states, provinces, 2012. URL http://www.naturalearthdata.com. accessed 15 January 2013.

P. Persson and E. V. Zhuravskaya. Elite capture in the absence of democracy: Evidence from backgrounds of chinese provincial leaders, 2012.

T. Persson and G. Tabellini. *Political Economics: Explaining Eonomic Policy*. Cambridge, Massachusetts: MIT Press, 2000.

V. Shih. Factions matter: Personal networks and the distribution of bank loans in china. *Journal of Contemporary China*, 13:3–19, 2004.

V. Shih. *Factions and Finance in China? Elite Conflict and Inflation*. Cambridge: Cambridge University Press, 2008.

V. Shih, C. Adolph, and M. Liu. Getting ahead in the communist party: Explaining the advancement of central committe members in china. *American Political Science Review*, 106:166–187, 2012.

S. Shirk. *The Political Logic of Economic Reform in China*. University of California Press, 1993.

R. Tao, F. Su, L. Xi, and Y. Zhu. Jingjizengzhang nenggoudailai shengqian ma?: dui jinshengjinbiaojingsaililun de luojitiaozhan yu shengjishizhengchonggu (does economic growth lead to promotion?: A challenge of the logic of tournaments and a reevaluation of the provincial-level data). *Guanlishijie (Management World)*, 12:13–26, 2010.

S. H. Whiting. *Power and Wealth in Rural China*. Cambridge University Press, 2000.

E. Wong. Family ties and hobnobbing trump merit at china helm, 2012. URL `http://www.nytimes.com/2012/11/18/world/asia/family-ties-and-hobnobbing-are-keys-to-power-in-china.html`. November 17.

C. Xu. The fundamental institutions of china's reforms and development. *Journal of Economic Literature*, 49:1076–1151, 2011.

W. Zhang. Meritocracy versus democracy. New York Times, 2012. URL `http://www.nytimes.com/2012/11/10/opinion/meritocracy-versus-democracy.html`. November 9.

Bo Zhiyue. Economic performance and political mobility: Chinese provincial leaders. *Journal of Contemporary China*, 5:135–154, 1996.

Bo Zhiyue. *Chinese Provincial Leaders: Economic Performance and Political Mobility Since 1949*. M.E. Sharpe, 2002.

A Appendix

A.1 A model of promotion as an incentive scheme

An alternative model of promotion is that the PSC cares about provincial economic growth per se, not the ability of those to be promoted. Promotion is used as an incentive scheme where growth is determined by provincial leaders' effort, not their ability.

Suppose that the PSC derives the utility from the share of tax revenues in province i that its leader i (with his connection status C_i) contributes to them. Assuming that tax revenues increase with growth, g_i, we can write the PSC's payoff as follows:

$$\alpha(C_i)g_i,$$

where $\alpha(C_i)$ is the extent to which provincial leader i shares his province's tax revenue with the PSC. We have $\alpha(1) > \alpha(0)$ if connections make provincial leaders more obliged to share their tax revenue with the PSC. If unconnected provincial leaders need to share more tax revenues with the PSC to compensate for the lack of loyalty through connections, we have $\alpha(1) < \alpha(0)$.

Economic growth in province i is determined by:

$$g_i = e_i + \varepsilon_i,$$

where e_i is the effort exerted by the leader of province i and ε_i the stochastic shock to growth, distributed by the cumulative distribution function G with mean 0.

Provincial leader i obtains the payoff of r if promoted and zero otherwise, and the disutility from making an effort for i is $\kappa(e_i)$ with $\kappa' > 0$ and $\kappa'' > 0$.

Assume that the PSC can commit to promoting provincial leader i if $\alpha(C_i)g_i \geq \bar{u}$, where $\bar{u}$ is the performance of an alternative candidate. For simplicity, we assume $\bar{u}$ to be uniformly distributed in the interval $[-1/2\nu, 1/2\nu]$.

We first analyze the provincial leader's behavior. The probability of promotion given e_i is:

$$\begin{aligned} Pr(\alpha(C_i)(e_i + \varepsilon_i) \geq \bar{u}) &= \int \big[\frac{1}{2} + \nu\alpha(C_i)(e_i + \varepsilon_i)\big] dG(\varepsilon_i) \\ &= \frac{1}{2} + \nu\alpha(C_i)e_i, \end{aligned}$$

with the last equality by $\int \varepsilon_i dG(\varepsilon_i) = 0$.

Provincial leader i chooses e_i to maximize

$$\big[\frac{1}{2} + \nu\alpha(C_i)e_i\big]r - \kappa(e_i).$$

The first-order condition is given by

$$\nu\alpha(C_i)r = \kappa'(e_i).$$

By $\kappa'' > 0$, there is the unique solution for e_i, $e_i^*(C_i)$, with $e_i^*(1) > e_i^*(0)$ if and only if $\alpha(1) > \alpha(0)$.

Given this optimal behavior, we now look at the marginal probability of promotion with respect to growth. Once ε_i is observed, the probability of promotion is given by

$$Pr(\alpha(C_i)g_i \geq \bar{u}) = \frac{1}{2} + \nu\alpha(C_i)g_i.$$

Differentiating this expression with respect to g_i yields:

$$\frac{\partial Pr(\alpha(C_i)g_i \geq \bar{u})}{\partial g_i} = \nu\alpha(C_i).$$

Consequently, if we have $\alpha(1) > \alpha(0)$, this expression is larger for $C_i = 1$ and thus connections and growth are complementary. If $\alpha(1) < \alpha(0)$, they are substitutes.

The above argument depends on the assumption that the PSC's commitment to this promotion scheme is credible. Once growth has been realized, the PSC is indifferent between promoting the high-performing provincial leader and reneging on the promise (and even prefers not promoting if promotion is costly). Credible commitment is plausible if the PSC expects to remain in power for a certain period of time and thus needs to build the reputation to reward good performance for future provincial leaders. Otherwise, we need an assumption that the PSC cares about the ability of those promoted so that the PSC has an incentive to promote those who have achieved high growth (see Fairburn and Malcomson 2001).

A.2 Model extensions

General distribution of $\bar{u}$ In the main text, we assume that $\bar{u}$, the payoff of not promoting provincial leader i, is uniformly distributed. If we instead impose no restriction on the cumulative distribution function of $\bar{u}$, F, equation (4.4) becomes

$$\frac{\partial F(W_i^{C_i})}{\partial g_i} = f(W_i^{C_i})\eta p(C_i)h(C_i),$$

where f is the probability density function of $\bar{u}$. As a result, connections and growth are complementary if

$$\frac{p(1)}{p(0)} > \frac{f(W_i^0)}{f(W_i^1)}\frac{h(0)}{h(1)},$$

and substitutes if the opposite inequality holds.

Since $W_i^{C_i}$ depends on g_i, whether connections and growth are complementary or substitutes may change with g_i. Denote the relative importance of the loyalty-fostering role of connections to their informational role by

$$\xi \equiv \frac{p(1)/p(0)}{h(0)/h(1)}$$

so that if $f(W_i^0)/f(W_i^1) < \xi$, connections and growth are complementary. Below we show that for sufficiently high ξ, complementarity holds for a wide range of g_i around the mean growth (i.e. $\bar{a}$). We also show that substitutability holds for a wide range of g_i around $\bar{a}$ for sufficiently low ξ.

First, suppose that $\xi > 1$. With the uniform distribution of $\bar{u}$, this condition implies complementarity of connections and growth. It also implies that the difference in payoffs from promoting connected and unconnected provincial leaders with the same growth performance, $W_i^1 - W_i^0$, monotonically increases with g_i. Let $\hat{g}$ be the growth rate that equates W_i^1 and W_i^0. Since $W_i^1 > W_i^0$ at $g_i = \bar{a}$, we know that $\hat{g} < \bar{a}$.

At $g_i = \hat{g}$, $f(W_i^0)/f(W_i^1) = 1$, suggesting that connections and growth are complements. When $\partial f(W_i^{C_i})/\partial W_i^{C_i} > 0$ (e.g. $\bar{u}$ being normally distributed with the equilibrium promotion probability less than a half), $f(W_i^0)/f(W_i^1)$ decreases with g_i because W_i^1 becomes larger than W_i^0. Therefore, there exists $\bar{g}^c < \hat{g}$ such that connections and growth are complementary for all $g_i > \bar{g}^c$. On the other hand, when $\partial f(W_i^{C_i})/\partial W_i^{C_i} < 0$ (e.g. $\bar{u}$ follows the Pareto distribution), $f(W_i^0)/f(W_i^1)$ increases with g_i. There exists $\tilde{g}^c > \hat{g}$ such that connections and growth are complementary for all $g_i < \tilde{g}^c$. As $\tilde{g}^c$ increases with ξ, we have $\tilde{g}^c > \bar{a}$ for a large enough ξ. Thus, for a range of g_i around $\bar{a}$, connections and growth are complementary under general distributions of $\bar{u}$ if $\xi > 1$.

Second, suppose instead that $\xi < 1$, under which connections and growth are substitutes if $\bar{u}$ follows the uniform distribution. In this case, $W_i^1 - W_i^0$ monotonically decreases with g_i. As $W_i^1 > W_i^0$ at $g_i = \bar{a}$, this implies that $\hat{g} > \bar{a}$.

At $g_i = \hat{g}$, $f(W_i^0)/f(W_i^1) = 1$, suggesting that connections and growth are substitutes. When $\partial f(W_i^{C_i})/\partial W_i^{C_i} > 0$, $f(W_i^0)/f(W_i^1)$ increases with g_i, suggesting that there exists $\bar{g}^s$ such that connections and growth are substitutes for all $g_i > \bar{g}^s$. With a small enough ξ, we have $\bar{g}^s < \bar{a}$. In the case of $\partial f(W_i^{C_i})/\partial W_i^{C_i} < 0$, $f(W_i^0)/f(W_i^1)$ decreases with g_i. There exists $\tilde{g}^s > \hat{g}$ such that connections and growth are substitutes for all $g_i < \tilde{g}^s$.

PSC's survival probability decreases with ability The probability for the PSC members of remaining in power, $p(C_i)$, may depend on a_i. If $p(C_i)$ increases with a_i, it would act in a similar way to a_i in the PSC's utility function. Consequently, our results are robust. A more interesting, and probably more plausible, case is when $p(0)$ decreases with a_i while $p(1)$ does not depend on a_i. In other words, more able officials threaten the power of the incumbent PSC members, and this effect is weaker if the officials are connected because connected officials are loyal to the PSC members.

Denote this probability by $p(C_i, a_i)$. Since a_i is unobservable, the PSC forms an expectation on the probability of survival from observed growth, which is given by

$$\int p(C_i, a_i)\phi(a_i|g_i)da_i,$$

where $\phi(a_i|g_i)$ is the posterior probability density function of a_i given g_i.

In this setting, equation (4.4) becomes

$$\begin{aligned}\frac{\partial F(W_i^{C_i})}{\partial g_i} =& f(W_i^{C_i})\Big[\eta h(C_i)\int p(C_i,a_i)\phi(a_i|g_i)da_i \\ &+ (R+\eta(h(C_i)g_i+(1-h(C_i))\bar{a}))\int p(C_i,a_i)\frac{\partial\phi(a_i|g_i)}{\partial g_i}da_i\Big].\end{aligned}$$

The second term in the square brackets is zero for $C_i = 1$. For $C_i = 0$, it is negative for $g_i > \bar{a}$ because we have $\partial p(0,a_i)/\partial a_i < 0$ and $\partial^2\phi(a_i|g_i)/\partial g_i a_i > 0$. It is positive if $g_i < \bar{a}$.

This implies that $F(W_i^0)$ is a concave function. If $\xi > 1$, the complementarity may not hold for a low value of g_i. If $\xi < 1$, substitutability may not hold for a high value of g_i. In both cases, however, a sufficiently high (low) ξ preserves the complementarity (substitutability) of connections and growth for a wide range of g_i.

Strategic provincial leader Given the promotion scheme, it is natural for provincial leaders to exert an effort to boost the economic growth of the province. Now, we modify the model to incorporate such strategic behavior of the provincial leader. Provincial leader i obtains the payoff of r if promoted and zero otherwise. Provincial growth, g_i, is now determined by the following equation:

$$g_i = a_i + e_i + \varepsilon_i, \tag{4.9}$$

where e_i is the effort made by i, and the function g is increasing in all arguments.[33] The disutility from making an effort for i is $\kappa(e_i)$ with $\kappa' > 0$ and $\kappa'' > 0$. The timing of events is as follows. First, nature picks the value of a_i, unobservable to both the PSC and the provincial leader.[34] Second, provincial leader i chooses e_i. Third, nature picks the value of ε_i, and thus g_i is observed by all players. Finally, the PSC decides whether to promote i.

The expected ability conditional on the observed growth is now given by

$$E(a_i|g_i) = h(C_i)(g_i - \tilde{e}_i) + [1 - h(C_i)]\bar{a}, \tag{4.10}$$

[33]We might consider a situation where ability and effort are complements: $g_i = a_i e_i + \varepsilon_i$. This case is intractable to analyze although it can be shown that the interdependence between connections and growth now depends on the equilibrium effort level as well as on $p(C_i)$ and $h(C_i)$, which may or may not offset the connection effects.

[34]The assumption that the provincial leader does not know his own ability follows the standard career-concern model (Holmstrom, 1982, Persson and Tabellini, 2000). It implies that a provincial leader does not know ex ante to what extent he is capable of running a provincial economy and of running the central government if promoted. This assumption certainly affects the optimal effort choice by provincial leaders. However, as we will see, the interdependent role of connections and growth in promotion does not hinge on the optimal effort level. Thus, this assumption is innocuous for our purpose.

where $\tilde{e_i}$ denotes the optimal choice of effort by i. When choosing e_i, provincial leader i knows that the PSC will promote i if

$$p(C_i)\Big[R+\eta\big[h(C_i)(a_i+e_i+\varepsilon_i-\tilde{e_i})+[1-h(C_i)]\bar{a}\big]\Big]\geq\bar{u}. \tag{4.11}$$

Provincial leader i maximizes the probability that this condition holds. Since i does not know his own ability, this condition suggests that the optimal effort level only differs by C_i. Denote this optimal effort by $e^*(C_i)$. Once g_i has been observed, the probability of i's promotion is

$$F\Big[p(C_i)\big[R+\eta\{h(C_i)[g_i-e_i^*(C_i)]+[1-h(C_i)]\bar{a}\}\big]\Big],$$

where we exploit $\tilde{e}=e_i^*(C_i)$ by rational expectation. Differentiating this expression with respect to g_i yields equation (4.4).

Connections affect the average ability The average ability, $\bar{a}$, may depend on C_i. Since connected provincial leaders are known to the PSC members when they assume provincial office, they are likely to be a selected sample of officials with higher ability: $\bar{a}(C_i=1)>\bar{a}(C_i=0)$. Alternatively, connected provincial leaders are those relying on connections to climb the promotion ladder in the Communist Party while unconnected ones are those relying on their competence. In this case, we would have $\bar{a}(C_i=1)<\bar{a}(C_i=0)$.

With the uniform distribution of $\bar{u}$, allowing $\bar{a}$ to depend on C_i does not affect the analysis as $\partial F(W_i^{C_i})/\partial g_i$ does not depend on a_i. However, the above analysis with more general distribution functions of $\bar{u}$ changes as follows. ξ needs to be much larger for complementarity to hold if $\xi>1$ and $\partial f(W_i^{C_i})/\partial W_i^{C_i}<0$ while it needs to be much lower for substitutability to hold if $\xi<1$ and $\partial f(W_i^{C_i})/\partial W_i^{C_i}>0$. This is because $\bar{a}(C_i=1)>\bar{a}(C_i=0)$ makes $W_i^1-W_i^0$ larger, thus moving $\hat{g}$ further away from $\bar{a}$.

Connections reduce the variance of growth shock The variance of the growth stochastic shock, σ_ε, may depend on C_i. If σ_ε is smaller for $C_i=1$, $h(C_i)$ in equation (4.3) can be larger for $C_i=1$. As a result, condition (4.6) is less likely to hold, and thus the substitutability between connections and performance is less likely to be observed even if we have $\sigma_a(1)<\sigma_a(0)$.

A.3 Definition of variable names used in tables

Measures of promotion

Promoted The indicator of getting promoted where promotion is defined as becoming a member of the Politburo, a Vice Premier, and a State Councilor for secretaries, a member of the PSC for Politburo-member secretaries, and a provincial secretary for governors.

Minister The indicator of getting promoted according to the definition of promotion that includes becoming a minister for governors.

NPC The indicator of getting promoted according to the definition of promotion that further includes becoming the vice-chairman of the NPC for both secretaries and governors.

CPPCC The indicator of getting promoted according to the definition of promotion that further includes becoming the vice-chairman of the CPPCC for both secretaries and governors.

Measures of connection

Connection The indicator of being connected to at least one of the current members of the PSC where connection is defined as working in the same workplace in the same period in the past.

Class The indicator of being connected to at least one of the current members of the PSC where connection is defined as graduating from the same college within a range of three years.

Alumni The indicator of being connected to at least one of the current members of the PSC where connection is defined as graduating from the same college irrespective of the graduation year.

Birth province The indicator of being connected to at least one of the current members of the PSC where connection is defined as being born in the same province.

Leadership spell level variables

Term length The number of years in office.

Other time-variant variables

Growth The average annual real provincial GDP growth since assuming office minus the sample mean of the 966 leader-year observations.

PSC work province The indicator of ruling the province where at least one current member of the PSC used to work.

PSC home province The indicator of ruling the province where at least one current member of the PSC was born.

Future/past connection The indicator of being unconnected to any current members of the PSC but being connected to at least one past or future member of the PSC, where connection is defined as working in the same workplace in the same period in the past.

Transfer from center The fiscal transfer from the central government to the province that the provincial leader is ruling.

Workplace The indicator of being unconnected to any current members of the PSC but having worked in the workplace in which at least one current member of the PSC used to work in a different period.

Politburo connection The indicator of being unconnected to any current members of the PSC but being connected to at least one member of the Politburo, where connection is defined as working in the same workplace in the same period in the past.

Other time-invariant variables

Age The age in the year of assuming office (thus time-invariant for each leadership spell).

Previous growth The average annual real provincial GDP growth for the five-year period before assuming office.

College graduate The indicator of having graduated from college.

Served in center The indicator of having assumed positions in the central government.

Home The indicator of ruling the province where the provincial leader was born.

Princeling The indicator of being a princeling (i.e. the son or the son-in-law of a prominent Communist Party official)

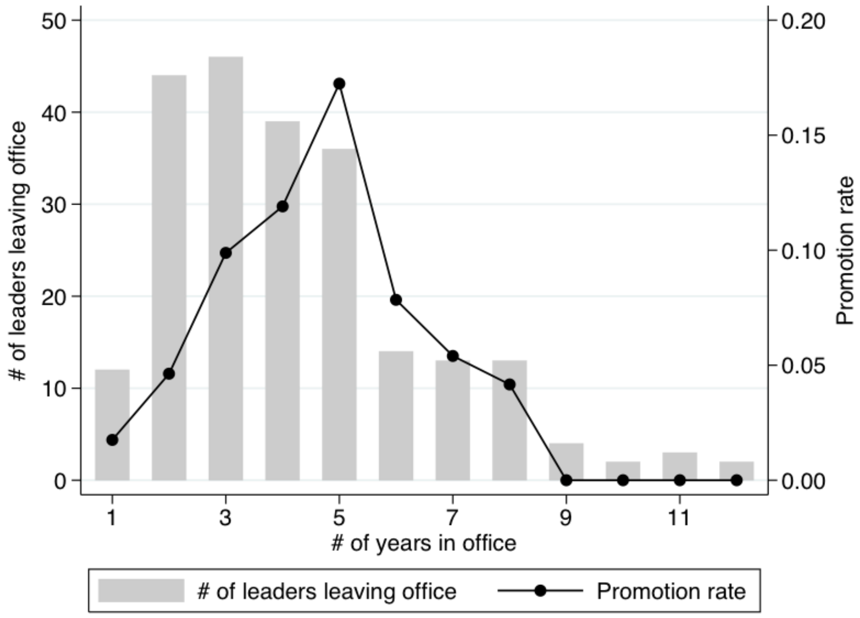


Figure 4.1: Distribution of term lengths and promotion rates by number of years in office

Notes: See the text for how this graph is constructed.

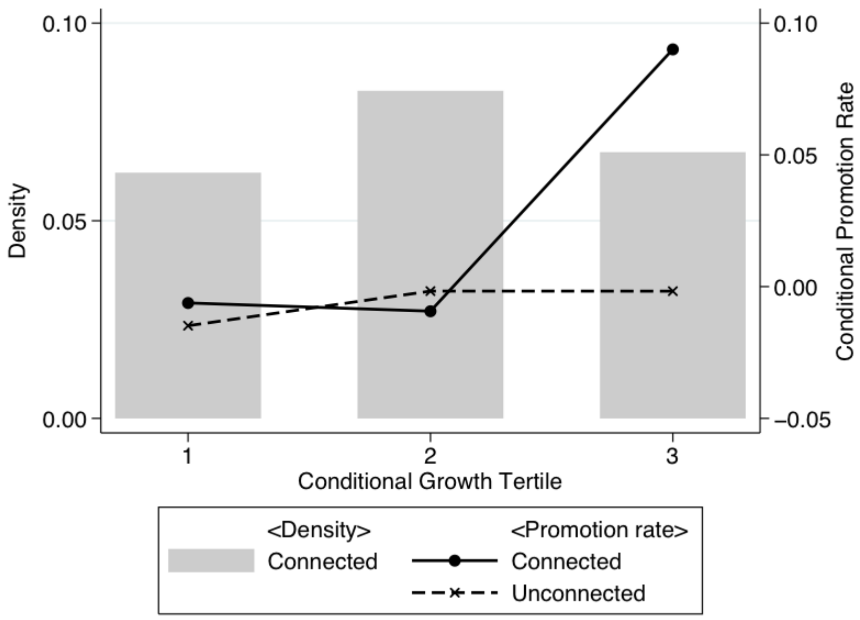


Figure 4.2: Promotion-growth relationship by connection

Notes: See the text for how this graph is constructed.

Table 4.1: Descriptive statistics

	(1) All spells	(2) Secretary spells	(3) Governor spells	(4) Leader years
Measures of promotion				
Promoted	0.26	0.18	0.34	0.07
Minister	0.29	0.18	0.40	0.08
NPC	0.31	0.19	0.42	0.08
CPPCC	0.33	0.22	0.45	0.09
Measures of connections				
Connection	0.25	0.27	0.23	0.21
Class	0.07	0.07	0.07	0.05
Alumni	0.12	0.13	0.11	0.10
Birth province	0.41	0.44	0.39	0.36
Growth	0.11	0.11	0.11	0.11
	(0.02)	(0.02)	(0.02)	(0.02)
Term length	4.29	4.39	4.18	3.23
	(2.27)	(2.42)	(2.12)	(2.14)
Time-invariant covariates				
Age	56.82	57.82	55.80	56.60
	(4.10)	(4.09)	(3.86)	(3.80)
Previous growth	0.11	0.11	0.11	0.11
	(0.02)	(0.02)	(0.03)	(0.02)
College graduate	0.82	0.82	0.82	0.82
Served in center	0.40	0.43	0.36	0.36
Home	0.23	0.15	0.31	0.26
Princeling	0.05	0.05	0.04	0.04
Time-variant covariates				
PSC work province	0.24	0.24	0.23	0.24
PSC home province	0.26	0.25	0.28	0.22
Future/past connection	0.21	0.22	0.21	0.18
Workplace	0.31	0.35	0.27	0.24
Politburo connection	0.27	0.27	0.27	0.21
Observations	258	130	128	966

Notes: Reported in each cell is the sample mean (and standard deviation in parentheses for continuous variables). The sample includes all leadership spells in column (1), provincial secretary spells in column (2), provincial governor spells in column (3) and leader-years in column (4). See Appendix Section A.3 for variable definitions. In columns (1)-(3), measures of promotion and *Growth* refer to the last year of the spell; measures of connections and time-variant covariates are the maximum value during the spell.

Table 4.2: Complementarity between connections and growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Promoted	Promoted	Promoted	Promoted	Minister	NPC	CPPCC
Connection	0.049**		0.048**	0.032	0.038	0.033	0.023
	(0.023)		(0.023)	(0.023)	(0.026)	(0.026)	(0.026)
Growth		0.702	0.690	0.356	0.441	0.328	0.273
		(0.465)	(0.447)	(0.427)	(0.452)	(0.460)	(0.423)
Connection * Growth				2.195**	2.309**	2.201**	2.374**
				(0.889)	(0.869)	(0.915)	(0.919)
				[0.056]*	[0.028]**	[0.050]*	[0.032]**
Fixed Effects	Y	Y	Y	Y	Y	Y	Y
# clusters	31	31	31	31	31	31	31
# observations	966	966	966	966	966	966	966

Notes: Standard errors clustered at the province level are reported in parenthesis. Reported in brackets are the p-values for the significance of the coefficient on *Connection * Growth* by the wild cluster bootstrap-t (Cameron et al., 2008). See Appendix Section A.3 for variable definitions. However, the variable *Growth* is normalized by subtracting the sample mean. All columns control for dummies of the number of years in office (two to twelve), office-by-province fixed effects, and office-by-year fixed effects.
* Significant at 10%, ** 5%, *** 1%.

Table 4.3: Do connected leadership spells differ from unconnected ones?

	(1) Connected	(2) Unconnected	(3) t-value	(4) Conditional difference
Promoted	0.35	0.24	1.73*	0.21***
				[0.08]
Tenure length	3.76	4.36	-1.90*	0.18
	(1.66)	(2.38)		[0.29]
Growth	0.12	0.11	3.72***	0.00
	(0.02)	(0.02)		[0.00]
Age	55.26	57.20	-3.05***	-3.35***
	(4.21)	(3.99)		[0.88]
Previous growth	0.12	0.10	2.92***	0.00
	(0.02)	(0.03)		[0.00]
College graduate	0.88	0.82	1.27	0.05
				[0.07]
Served in center	0.53	0.37	2.35**	0.20*
				[0.12]
Home	0.10	0.25	-2.43**	-0.06
				[0.07]
Princeling	0.06	0.04	0.50	0.00
				[0.07]
PSC work province	0.51	0.14	6.12***	0.16*
				[0.08]
PSC home province	0.29	0.22	0.88	0.07
				[0.07]
Observations	50	208		258
Transfer from center	4649.74	3133.91	2.79***	26.53
(in million RMB)	(3855.64)	(3280.62)		[507.86]
Observations	49	198		247

Notes: The unit of observations is the leadership spell. Columns (1) and (2) report the mean (and standard deviation for continuous variables) for those spells where the provincial leader is connected and unconnected, respectively, to the PSC in the first year of the spell. Column (3) reports t-statistics for the null that the means in columns (1) and (2) are the same. Column (4) reports the estimated coefficient on the connection status in the first year of the spell (and robust standard errors in brackets) from a regression of each variable on the connection indicator, office-by-province dummies, and office-by-year dummies. See Appendix Section A.3 for variable definitions. To aggregate leader-year level data to the spell level, we take the last year observation for *Promoted* and *Growth*, the first year observation for *PSC work province* and *PSC home province*, and the annual average for *Transfer from center*.
* Significant at 10%, ** 5%, *** 1%.

Table 4.4: Does provincial economic growth reflect the support from the central government?

(Dependent variable: Annual real provincial GDP growth)

	(1) Secretaries	(2) Governors	(3) Secretaries	(4) Secretaries	(5) Governors	(6) Governors
1 year after promotion	0.006	0.001				
	(0.004)	(0.004)				
2 years after promotion	0.003	0.002				
	(0.005)	(0.004)				
3 years after promotion	0.003	-0.001				
	(0.003)	(0.003)				
Connection			-0.005	-0.004	-0.009	0.002
			(0.005)	(0.006)	(0.005)	(0.006)
Years of working together * Connection			0.001		0.003**	
			(0.001)		(0.001)	
Age difference * Connection				-0.000		-0.000
				(0.001)		(0.001)
Fixed Effects	Y	Y	Y	Y	Y	Y
# clusters	31	31	31	31	31	31
# observations	488	478	488	488	478	478

Notes: Standard errors clustered at the province level are reported in parenthesis. The unit of observations is the province-by-year. The variable *x year after promotion* is a dummy for x years after the previous leader in the same province has been promoted. *Connection* is the indicator of the provincial leader having worked together in the past with any of the current members of the Politburo Standing Committee. *Years of working together* is the (average, if more than one connection) number of years that the provincial leader has worked together with the current member(s) of the Politburo Standing Committee. *Age difference* is the (minimum, if more than one connection) difference in age between the provincial leader and his connected member(s) of the Politburo Standing Committee. In columns (1), (3) and (4), the provincial leaders to define these variables are provincial secretaries; in columns (2), (5), and (6), provincial governors. All columns control for province fixed effects and year fixed effects.

* Significant at 10%, ** 5%, *** 1%.

Table 4.5: Robustness to controlling for potential confounding variables with the connection indicator

(Dependent variable: Promoted)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Connection	0.012	0.029	0.029	0.029	0.026	0.043	0.029	0.014
	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.027)	(0.023)	(0.029)
Growth	1.530	0.210	0.347	2.815*	0.171	0.317	0.363	1.378
	(5.289)	(0.471)	(0.397)	(1.613)	(0.435)	(0.430)	(0.402)	(5.882)
Connection * Growth	2.096***	1.983**	2.173**	2.429**	2.150**	2.222**	2.222**	2.309**
	(0.732)	(0.769)	(0.817)	(0.985)	(0.901)	(1.004)	(0.882)	(0.845)
	[0.032]**	[0.058]*	[0.032]**	[0.064]*	[0.080]*	[0.098]*	[0.066]*	[0.042]**
Age	-0.009***							-0.008***
	(0.003)							(0.003)
Age * Growth	-0.023							0.018
	(0.094)							(0.105)
Served in center		0.026						0.033
		(0.025)						(0.024)
Served in center * Growth		0.607						0.707
		(0.670)						(0.840)
Home			-0.037					-0.040
			(0.029)					(0.030)
Home * Growth			0.074					0.316
			(0.840)					(0.939)
Previous Growth				0.155				-0.152
				(0.462)				(0.419)
Previous Growth * Growth				-22.428				-24.499*
				(13.519)				(13.914)
Princeling					-0.042			-0.077**
					(0.036)			(0.029)
Princeling * Growth					2.936			2.639
					(1.902)			(1.902)
PSC Work Province						-0.039		-0.062**
						(0.034)		(0.027)
PSC Work Province * Growth						-0.400		-0.704
						(1.237)		(1.133)
PSC Home Province							0.039*	0.046**
							(0.022)	(0.019)
PSC Home Province * Growth							-0.259	-0.346
							(0.842)	(0.712)
Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
# clusters	31	31	31	31	31	31	31	31
# observations	966	966	966	966	966	966	966	966

Notes: Standard errors clustered at the province level are reported in parenthesis. Reported in brackets are the p-values for the significance of the coefficient on *Connection * Growth* by the wild cluster bootstrap-t (Cameron et al., 2008). See Appendix Section A.3 for variable definitions. However, the variable *Growth* is normalized by subtracting the sample mean. All columns control for dummies of the number of years in office (two to twelve), office-by-province fixed effects, and office-by-year fixed effects.
* Significant at 10%, ** 5%, *** 1%.

Table 4.6: Connections or being part of the political elite?

(Dependent variable: Promoted)

	(1)	(2)	(3)
Connection	0.033	0.043*	0.037
	(0.028)	(0.022)	(0.027)
Growth	0.414	0.234	0.395
	(0.483)	(0.507)	(0.479)
Connection * Growth	2.125**	2.351**	2.157**
	(0.938)	(0.902)	(0.903)
	[0.090]*	[0.044]**	[0.058]*
Future/Past Connection	0.005		
	(0.027)		
Future/Past Connection * Growth	-0.265		
	(1.059)		
Workplace		0.026	
		(0.028)	
Workplace * Growth		0.285	
		(0.742)	
Politburo Connection			0.013
			(0.019)
Politburo Connection * Growth			-0.234
			(0.653)
F-test	3.27	2.61	3.96
	[0.052]	[0.090]	[0.030]
Fixed Effects	Y	Y	Y
# clusters	31	31	31
# observations	966	966	966

Notes: Standard errors clustered at the province level are reported in parenthesis. Reported in brackets are the p-values for the significance of the coefficient on *Connection * Growth* by wild cluster bootstrap-t (Cameron et al., 2008). See Appendix Section A.3 for variable definitions. However, the variable *Growth* is normalized by subtracting the sample mean. All columns control for dummies of the number of years in office (two to twelve), office-by-province fixed effects, and office-by-year fixed effects. *F-test* reports F-statistics and their associated p-values for the null that the coefficients on *Connection* and on *Future/Past Connection* in column (1), *Workplace* in column (2), or *Politburo Connection* in column (3), are the same and that the coefficients on their respective interaction terms with *Growth* are also the same.
* Significant at 10%, ** 5%, *** 1%.

Table 4.7: Other sources of connections

(Dependent variable: Promoted)

	(1)	(2)	(3)
Definition of connection:	Class	Almuni	Birth province
Other connection	0.117	0.061	0.000
	(0.071)	(0.046)	(0.020)
Growth	0.662	0.750	0.564
	(0.442)	(0.531)	(0.538)
Other connection * Growth	-0.177	-0.589	0.458
	(1.851)	(1.232)	(0.787)
	[0.940]	[0.634]	[0.590]
Fixed Effects	Y	Y	Y
# clusters	31	31	31
# observations	966	966	966

Notes: Standard errors clustered at the province level are reported in parenthesis. Reported in brackets are the *p*-values for the significance of the coefficient on *Other connection * Growth* by wild cluster bootstrap-t (Cameron et al., 2008). The variable *Other connection* refers to the variable mentioned at the top of each column. See Appendix Section A.3 for variable definitions. However, the variable *Growth* is normalized by subtracting the sample mean. All columns control for dummies of the number of years in office (two to twelve), office-by-province fixed effects, and office-by-year fixed effects.
* Significant at 10%, ** 5%, *** 1%.

Table 4.8: Does a larger age difference make complementarity stronger?

	(1)	(2)
Connection	-0.004	0.471
	(0.026)	(0.374)
Connection	-0.005*	-0.002
* Age difference	(0.003)	(0.003)
Connection		-0.008
* Age		(0.006)
Growth	0.295	0.372
	(0.446)	(0.440)
Connection	0.310	-5.230
* Growth	(0.992)	(14.305)
Connection		
* Growth	-0.247**	-0.271*
* Age difference	(0.102)	(0.135)
Connection		
* Growth		0.091
* Age		(0.237)
Fixed Effects	Y	Y
# clusters	31	31
# observations	966	966

Notes: Standard errors clustered at the province level are reported in parenthesis. *Age difference* is the (minimum, if more than one connections) difference in age between the provincial leader and his connected member(s) of the Politburo Standing Committee. See Appendix Section A.3 for other variable definitions. However, the variable *Growth* is normalized by subtracting the sample mean. All columns control for dummies of the number of years in office (two to twelve), office-by-province fixed effects, and office-by-year fixed effects.
* Significant at 10%, ** 5%, *** 1%.

Table 4.9: Is the complementarity of connections and performance inefficient?

	(1)	(2)
Dependent variable:	Promoted	Second-term Growth
Connection	0.081**	0.004
	(0.035)	(0.004)
Growth	-0.309	0.162*
	(1.007)	(0.094)
Connection * Growth	2.568*	0.167
	(1.352)	(0.121)
	[0.140]	
Observations	453	71

Notes: In column (1), the unit of observations is the leader-year. The sample is restricted to those who assume provincial leadership positions more than once. The definition of variables is the same as in column 4 of Table 4.2. Standard errors clustered at the province level are reported in parenthesis. Reported in brackets is the p-value for the significance of the coefficient on *Connection * Growth* by wild cluster bootstrap-t (Cameron et al., 2008). In column (2), the leadership spell is the unit of observation. The sample is restricted to those who serve the second or third term of provincial leadership. The dependent variable is average annual growth during the tenure conditional on province and year fixed effects in a balanced province panel regression. *Connection* is an indicator of being connected in the last year of the previous term. *Growth* is the deviation of average conditional annual growth during the previous term from the sample average. No other variables are included as regressors. Robust standard errors are reported.
* Significant at 10%, ** 5%, *** 1%.

Chapter 5

Does the Demand for Redistribution Rise or Fall with Cognitive Ability?*

5.1 Introduction

The extent to which economic resources are redistributed between citizens varies widely across countries. We know from previous studies that factors such as income (Romer, 1975, Meltzer and Richards, 1981, Alesina and Giuliano, 2011), race (Alesina and Giuliano, 2011, Alesina and Ferrara, 2005), gender (Alesina and Giuliano, 2011, Alesina and Ferrara, 2005) and cultural factors (Fong and Luttmer, 2011, Luttmer and Singhal, 2011) influence individuals' demand for redistribution, but much remains unknown about how preferences for redistribution are formed. In this paper, we ask what role an individual's cognitive ability plays in determining preferences for redistribution. Because the more cognitively able take a larger part in democratic decision making, both as voters (Deary et al., 2008) and as leaders (Connelly et al., 2000, Judge et al., 2004), this question is particularly important.

From what is known about how cognitive ability affects other types of preferences, it is not clear how a higher cognitive ability should be reflected in an individual's opinion on redistribution. For example, some previous research has shown that people with a higher cognitive ability are more prone to take on risk (Benjamin et al., 2013) which, all else equal, should make them want less redistribution. On the other hand, other research indicates that higher cognitive ability is associated with stronger altruism (Rustichini et al., 2012), which points in the other direction. Other factors, such as how the costs and benefits of redistribution are perceived, may also differ between people with different

*We are grateful to the Ragnar Söderberg Foundation, The Royal Academy of Sciences and the Lab for Economics Applications and Policy at Harvard for financial support. We thank Alberto Alesina, Dan Benjamin, Raj Chetty, David Cutler, Olle Folke, Ed Glaeser, Larry Katz, Anna Larsson Seim, Erzo Luttmer, Michael Norton, Torsten Persson, Aldo Rustichini, David Strömberg, Christoph Wollersheim and seminar participants at Harvard University and the IFN for very valuable comments. We also thank Christina Lönnblad for editorial assistance.

cognitive ability.

This study uses Swedish data to investigate the relation between cognitive ability and the demand for redistribution. The data set has two parts, the first being a tailor-made survey, which elicits preferences regarding income redistribution. The second part comprises administrative records that include a measure of cognitive ability collected for men upon military enlistment.

By linking the survey- and the administrative data at the individual level, we establish that the demand for redistribution – measured in percent, with 100 corresponding to full redistribution, i.e. everyone earns the same amount after taxes and redistribution – decreases linearly with cognitive ability. The effect is large: a one-standard deviation increase in cognitive ability test scores taken at the age of 18 decreases the demand for redistribution surveyed at ages 33–61 by about 6.5 percent. This difference is larger than that between singles and married individuals or between men and women.

Our extensive survey- and administrative data also allow us to explore potential channels through which cognitive ability influences preferences. A strand of literature in labor economics has shown that cognitive ability is strongly related to earnings (see e.g. Heckman et al., 2006 and Lindqvist and Vestman, 2011). In line with Romer [1975], Meltzer and Richards [1981] and Alesina and Giuliano [2011], it is thus natural to think of labor earnings as being one potential channel through which cognitive ability influences preferences. Therefore, we use historical administrative data on labor earnings as well as individual beliefs about future income to distinguish between the full effect of cognitive ability on the demand for redistribution, and the partial effect, controlling for both current and estimated long-run income. We find that income measures are negatively correlated with the demand for redistribution, as expected, and that income can only account for a fraction of the relation between cognitive ability and the demand for redistribution.

We proceed with our analysis by investigating whether the link between cognitive skills and redistributional demands arises because of differences in risk preferences and altruism. Using methods to uncover such preferences in surveys developed by Falk and Becker [2010], we find that although the partial correlations of these measures have the expected signs, including these variables in a regression where demand for redistribution is the dependent variable does not affect the coefficient on cognitive ability.

The next section briefly discusses previous literature and places cognitive ability into its context. Section 5.2 describes the survey and administrative data. Section 5.3 discusses our results and section 5.4 concludes.

5.2 Data

5.2.1 Survey Data

The survey was designed by us and implemented by Statistics Sweden (see the Appendix for the survey). It was sent by mail to a representative sample of Swedes aged above 18 in May 2011 with a response rate of 36 percent. In addition to background questions, the survey elicits some economic preferences (including altruistic preferences and attitudes towards risk) and asks about opinions on income taxation and the demand for redistribution. In the survey, we define redistribution as meaning ”that the public sector, through taxes and subsidies, makes income in society more equal between the citizens than what would have been the case without these taxes and subsidies”. Note that this definition confines the focus to redistribution by the public sector as opposed to redistribution through, for example, charitable giving. Subjects are asked to indicate their preferred level of income redistribution on a scale from 1 (no redistribution, defined in the survey instructions as meaning that ”the income after taxes and subsidies is the same as before”) to 10 (full redistribution, i.e. that ”everyone's income after taxes and subsidies is the same”). This question is similar to what has been used in other research on the demand for redistribution (Alesina and Giuliano, 2011, Alesina and Ferrara, 2005 and Alesina and Glaeser, 2004). However, our definition of the concept of redistribution, being exclusively restricted to being conducted by the public sector, is novel. In the analysis below, we rescale the variable so that it ranges from 0 to 100 in order to be able to interpret it as the percent of redistribution desired.

5.2.2 Administrative Data

To obtain a more complete picture of the respondents, we link the survey data to administrative data at the individual level. The main source of administrative data is the longitudinal integration database for health insurance and labor market studies (LISA by Swedish acronym). LISA contains information on taxable income and wealth, social benefits, occupation, wages and education. Our data on taxable income and wealth cover the years 1999-2010.

The survey data are additionally linked to military enlistment data. For the sample that we analyze (which encompasses enlistments between 1969 and 1994), military service was mandatory for all Swedish men (but not for women, so our sample is not fully representative of the Swedish population). Military enlistment normally takes place at the age of 18 and includes tests of physical and cognitive ability. The cognitive ability test consists of four sub-tests (logical ability, verbal ability, technological comprehension and metal folding) with 40 questions each and is an accepted measure of intelligence (see e.g. Carlstedt, 2000). Moreover, according to Lindqvist and Vestman [2011], it was not possible to avoid the military service by scoring low on the cognitive ability test. The military weights the scores from the subtests to create an index-variable, scaled from one to nine. We normalize this score within enlistment year so that it has zero mean and unit variance.[1]

5.2.3 Sample

A total of 1,549 people responded to the survey (corresponding to a response rate of 36 percent). Since the cognitive ability data only cover men who enlisted in the military between 1969 and 1994, this subsample comprises 271 men. Table 5.1 presents summary statistics for key variables for different samples. In column (1), we present key statistics for the population of Swedes aged above 16. This sample is confined to all men in column (2). Column (3) presents information on the respondents to our survey. Columns (4) and (5) restrict the sample to all men and all men with military enlistment data, respectively. Although the subjects who originally received the survey consitute a representative sample of the Swedish population, selected by Statistics Sweden, this is not true for the people who responded to the survey. Comparing columns (1) and (3) of Table 5.1, people who responded to the survey have higher annual earnings and more education.

Comparing men for whom we have data on cognitive ability to the other men in our sample, we note that they are very similar with the exception that the former have higher incomes. This is, to a large extent, explained by men with data on cognitive ability being aged between 33 and 61 when taking the survey. Although the mean age among the sample of men who responded to our survey and have cognitive skills data is virtually the same as that of the other groups, the standard deviation is considerably smaller. They are thus more likely to be part of the labor force.

5.3 Results

To analyze the data, we use the following specification:

$$D_i = \alpha + \beta c_i + \mathbb{X}_i \gamma + \varepsilon_i, \tag{5.1}$$

where D_i is individual i's demand for redistribution, measured discretely on a scale $0-100$ at ten equally spaced points. c_i refers to normalized cognitive ability and $\mathbb{X}_i$ represent basic controls. Although the cognitive ability test measures actual cognitive ability with noise, the cognitive ability tests have been subject to only minor changes. Thus, it is

[1] Given our sample, one may question the potential to extrapolate our analysis of cognitive ability and redistributional demands to women. Although we do not claim that our results are representative of the relation for women, previous research shows that fundamental determinants of the demand for redistribution, such as income or race, do not vary by gender. Therefore, it does not strike us as implausible that the results here may also generalize to women.

likely that any measurement error in the cognitive ability variable will bias the coefficient downwards.

Before estimating equation (5.1), we plot the relation between cognitive ability and the demand for redistribution without imposing any functional form assumptions. Figure 5.1 shows the residuals from regressing demand for redistribution on age, education dummies and mean income since 1999, against cognitive ability. The y-axis denotes the sum of the mean of demand for redistribution and the residuals. Each bin corresponds to 8.33 percent of the men with enlistment data. Notably, the relation between the variables is not only negative, but linear, suggesting that the above specification is correct.

Table 5.2 presents the results from estimating equation (5.1) using Ordinary Least Squares (OLS) with various sets of control variables.[2] The negative relation presented in Figure 5.1 is confirmed to be statistically significant. The coefficient in column (1) suggests that a one-standard deviation increase in cognitive ability reduces the demand for redistribution by 6.7 %. Including basic controls in column (2) only marginally changes the estimated coefficient.

The estimated coefficient is economically important. As an illustration, we compare the coefficient on cognitive ability with the correlations with gender (a factors which is well-known to be associated with the demand for redistribution). We find that a one-standard deviation rise in cognitive ability has a larger absolute effect on the demand for redistribution than gender. Table 5.3 illustrates that men prefer five percent less redistribution than their female counterparts (the coefficient ranges from -3.1 to -5.3 depending on the specification).

Our results are robust to the inclusion of various income measures. Column (3) controls for the sum of annual labor and capital income in 2010 (the latest year before the survey was taken), while columns (4)-(7) control for mean annual administrative income since 1999 or since the person appeared in the dataset. Additionally, column (7) controls for being single, spouse's income and perceived relative income during the subject's youth as well as expected relative income during the next ten years.[3] Even after extensively controlling for current and more permanent income measures, the coefficient on cognitive ability remains statistically significant and the magnitude is unchanged.

However, the possibility remains that our results are due to a mis-measurement of long-run income. To address this issue, we look at whether the absolute size of the partial correlation (i.e. controlling for permanent income) between cognitive ability and demand for redistribution decreases with age. The idea is that as historical earnings become more important relative to future expected earnings, the mis-measurement in

[2] As a robustness check, we also run all regressions using Ordered Probit (OP). As this does not change any of the signs of the estimates nor the significance levels, the results from these regressions are omitted here.

[3] Relative income when young corresponds to question $20a$ in the survey whereas the future income expectations are gauged by the average of questions $15d$ and $15e$.

permanent income becomes smaller. This would, in turn, imply a smaller and less precise partial correlation for older men. Dividing the sample into tertiles according to the age distribution and running the regressions separately in these groups in Table 5.4 reveals that this is not the case. If anything, the relation between cognitive ability and demand for redistribution is strongest and most precisely estimated within the oldest group.

Risk Aversion and Altruism Having established a negative relation between cognitive ability and demand for redistribution that is robust to the inclusion of extensive controls for long-run income, we now investigate other potential mechanisms behind this relation. Our survey elicits risk preferences and altruism through measures developed in a laboratory environment but adapted for survey use by Falk and Becker [2010].[4] We now analyze whether these aspects of preferences help explain the negative relation.

Our survey utilizes a measure of risk aversion where the respondent makes hypothetical choices between receiving a fixed amount of money and participating in a lottery. Whereas the lottery remains the same in all eight questions, the fixed amount varies from 1/3 to 5/3 of the expected value of the lottery. The number of times a participant chooses the fixed amount is a well-established measure of risk aversion (see Holt and Laury, 2002). In unreported correlations, we find that risk aversion is indeed positively related to the demand for redistribution, which is in line with the theoretical predictions in Benabou and Ok [2001]. However, when estimating equation (5.1), controlling for risk aversion, the coefficient on cognitive ability does not change, as is shown in columns (5) and (8) of Table 5.2. Therefore, we conclude that the negative relation between demand for redistribution and cognitive ability in our sample of Swedish men does not stem from any variation in risk aversion.

Following Falk and Becker [2010], we also elicit a standard measure of altruistic preferences in the survey, namely the willingness to give in a hypothetical dictator game where the respondent is a charitable organization. In line with Rustichini et al. [2012], we find a positive correlation between altruistic preferences and cognitive ability. Since the partial correlation between altruistic preferences and demand for redistribution is positive, the negative relation between cognitive ability and demand for redistribution arises despite the fact that the more cognitively able express more altruism. In columns (6) and (9), we see that the coefficient on cognitive ability remains more or less unchanged compared to previous specifications.

[4] Falk and Becker [2010] also show that, although not incentivized, these survey based measures of economic preferences, correlate strongly with incentivized measures.

5.4 Conclusion

To conclude, our study provides novel evidence of a negative relation between cognitive ability and the demand for redistribution. This result is robust to the inclusion of a set of controls from both the survey and from administrative data. Importantly, the negative relation holds also when we add measures of long-run income to the regressions. By including preference parameters on altruism and risk aversion, we explore whether these variables can explain our finding, but find that they cannot.

Why do Swedish men in our sample with higher cognitive ability want less redistribution? In addition to long-run income, risk aversion and altruism, that we investigate, there must be other links between cognitive ability and the demand for redistribution. A particularly interesting avenue for future research is to explore whether the perceptions of and the ability to understand relative costs and benefits of redistribution differ across skill groups. Another plausible explanation for our finding could be a potential link between cognitive ability and the individual's social context. It would be interesting to explore whether the social context associated with different types of education plays a role. Since income inequality in Sweden is low, we are hopeful that future studies will also focus on countries with higher income inequality.

Bibliography

A. Alesina and E. La Ferrara. Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89:897–931, 2005.

A. Alesina and P. Giuliano. Handbook of social economics, 2011. A. Bisin, J. Benhabib and M. Jackson.

A. Alesina and E. Glaeser. *Fighting Poverty in the US and Europe: A World of Difference*. Oxford University Press, 2004.

R. Benabou and E. Ok. Social mobility and the demand for redistribution: the poum hypothesis. *Quarterly Journal of Economics*, 116:447–487, 2001.

D. Benjamin, S. Brown, and J. Shapiro. Who is 'behavioral'? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 2013. URL http://economics.cornell.edu/dbenjamin/iq.pdf.

B. Carlstedt. Cognitive abilities: Aspects of structure, process and measurement, 2000.

M. S. Connelly, J. A. Gilbert, S. J. Zaccaro, K. V. Threlfall, M. A. Marks, and M. D. Mumford. Exploring the reationship of leadership skills and knowledge to leader performance. *Leadership Quarterly*, 11:65–86, 2000.

I. Deary, D. Batty, and C. Gale. Childhood intelligence predicts voter turnout, voting preferences, and political involvement in adulthood: The 1970 british cohort study. *Intelligence*, 36:548–555, 2008.

A. Falk and A. Becker. Economic preference module, 2010. University of Bonn.

C. Fong and E. Luttmer. Do fairness and race matter in generosity? evidence from a nationally representative charity experiment. *Journal of Public Economics*, 95:372–394, 2011.

J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24:411–482, 2006.

C. Holt and S. Laury. Riskaversion and incentive effects. *American Economic Review*, 92:1644–1655, 2002.

T. Judge, A. Colbert, and R. Ilies. Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89:545–552, 2004.

E. Lindqvist and R. Vestman. The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal*, 3:101–128, 2011.

E. Luttmer and M. Singhal. Culture, context and the taste for redistribution. *American Economic Journal: Economic Policy*, 3:157–179, 2011.

A. Meltzer and S. Richards. A rational theory of the size of the government. *Journal of Political Economy*, 89:914–927, 1981.

T. Romer. Individual welfare, majority voting and the properties of a linear income tax. *Journal of Public Economics*, 7:163–188, 1975.

A. Rustichini, C. DeYoung, J. Anderson, and S. Burks. Toward the integration of personality theory and decision theory in the explanation of economic and health behavior, 2012.

A Appendix

For Online Publication: Appendix 2: Survey

Part 1: Background Information

1. When were you born?

Year: 1 9 _ _

2. Are you a man or a woman?

1 ☐ Man
2 ☐ Woman

3. Which is your main occupation right now?
NB! State only one alternative.

1 ☐ Public employed
2 ☐ Private employed
3 ☐ Own business
4 ☐ Unemployed
5 ☐ Student
6 ☐ Retired
7 ☐ Other

4. Which is/was your main profession?

If you do not work right now, please state the profession you have had for the longest time. Please answer as detailed as possible.

For example, instead of assistant write sales assistant. Please use capital letters!

Example: Instead of driver write

B U S S C H A U F F Ö R

Your profession:

1 ☐ No profession.

5. Are you an active member of any of the following organizations?

		Yes 1	No 2
a.	Church	☐	☐
b.	Sports	☐	☐
c.	Art or Music	☐	☐
d.	Union	☐	☐
e.	Political party	☐	☐
f.	Environmental	☐	☐
g.	Charity	☐	☐
h.	Consumer	☐	☐
i.	Other	☐	☐

6. Which is your main source of news and how often do you use it?

		Every day 1	Every week 2	Every month 3	Seldom/never 4
a.	Newspapers	☐	☐	☐	☐
b.	News on radio/TV	☐	☐	☐	☐
c.	Printed magazines	☐	☐	☐	☐

d.	Other radio/TV programs	☐	☐	☐	☐
e.	Internet	☐	☐	☐	☐
f.	Friends and colleuges	☐	☐	☐	☐

7. **To what extent do you agree with the following statements?**

a. **I am interested in politics**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

b. **I often discuss politics with family and friends**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

8. **Which party would you vote for if there were to be an election today?**
The parties are listed alphabetically

01 ☐ Centerpartiet
02 ☐ Feministiskt initiativ
03 ☐ Folkpartiet
04 ☐ Kristdemokraterna
05 ☐ Miljöpartiet
06 ☐ Moderaterna
07 ☐ Socialdemokraterna
08 ☐ Sverigedemokraterna
09 ☐ Vänsterpartiet
10 ☐ Other party
11 ☐ Would leave a blank vote
12 ☐ Would not vote
13 ☐ Don't know / Don't want to answer

Part 2: Economic Redistribution

Now follow some statements and questions about economic redistribution. Convey your opinion on each question by marking the alternative that is most right for you!

Economic redistribution means that the public sector, through taxes and subsidies, makes income in society more equal between citizens that what would have been the case without these taxes and subsidies. **The Public Sector** means the activities that all cities, regions and the federal state represent.

9. **a. How much economic redistribution do you want in society?**
No redistribution means that the public sector doesn't influence the income distribution at all. Full redistribution means that everyone gets the same amount after taxes and subsidies.

1 2 3 4 5 6 7 8 9 10

No redistribution. ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Full redistribution

b. If we were to ask question 9a to all Swedes who are 18 years or older, what do you think that the average answer would be?

1 2 3 4 5 6 7 8 9 10
No redistribution. ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Full redistribution.

10. **To what extent do you agree with the following statements?**

a. **I prefer the system of economic redistribution that means that I get the highest possible income after taxes and subsidies.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

b. **The public sector is responsible for making sure that the welfare of all citizens is above a certain minimum level.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

c. **I have a positive attitude to economic redistribution because I care about other people's standard of living.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

d. **I have a positive attitude to economic redistribution because I believe that it is good for me economically.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

e. **I have a positive attitude to economic redistribution because I feel that it gives safety if something unexpected were to happen.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

f. **I have a positive attitude to economic redistribution because I believe that it creates a more fair society.**

1 2 3 4 5 6 7 8 9 10
Disagree completely ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Agree completely

g. **I have a positive attitude to economic redistribution because I believe that it is good for me in some non-economic way.**
Example: if those with the lowest income get a higher income through redistribution it may reduce crime which is positive for me.

	1	2	3	4	5	6	7	8	9	10	
Disagree completely	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Agree completely

11. Would you like to change the income taxes that we have in Sweden today, and if so in what way?

1 ☐ Lower taxes
2 ☐ No change
3 ☐ Higher taxes

12. If the income taxes were to increase in Sweden, would you change your amount of work?

1 ☐ Yes, I would work more
2 ☐ No, I would work as much as I do now
3 ☐ Yes, I would work less

13. How would you like to change the economic redistribution in Sweden?

1 ☐ Less redistribution
2 ☐ No change
3 ☐ More redistribution

14. Do you think that you are a ”winner” or a ”loser” from the economic redistribution in Sweden?
A ”winner” is someone who uses services and gets subsidies from the public sector with a higher economic value than what he/she pays though taxes.

		"Loser" 1	Neither 2	"Winner" 3
a.	This year	☐	☐	☐
b.	During your whole life	☐	☐	☐

Part 3: Income currently and historically

15. Imagine that we divide all yearly incomes of Swedes on a scale between 1 and 10 so that 1 is the lowest income and 10 is the highest.

The income we refer to is the total yearly income which contains income from labor and capital before tax. Pensions before tax are also in this category. Subsidies like public unemployment payment are not part of the total yearly income.

a. **Where do you think that your income during last year (i.e. 2010) was on this scale?**

	1	2	3	4	5	6	7	8	9	10	
Individual 1 (lowest income)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Individual 10 (highest income)

b. **Where do you think that your income TEN YEARS AGO would be on the scale?**
Imagine that we use a scale that represents the incomes of Swedes 10 years ago. Leave the question blank if you were less than 18 years old 10 years ago.

	1	2	3	4	5	6	7	8	9	10	
Individual 1 (lowest income)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Individual 10 (highest income)

c. **Where do you think that your income FIVE YEARS AGO would be on the scale?**
Imagine that we use a scale that represents the incomes of Swedes 10 years ago. Leave the question blank if you were less than 18 years old 5 years ago.

	1	2	3	4	5	6	7	8	9	10	
Individual 1 (lowest income)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Individual 10 (highest income)

d. **Where do you think that your income IN FIVE YEARS would be on the scale?**

Imagine that we use a scale that represents the incomes of Swedes 5 years from now.

	1	2	3	4	5	6	7	8	9	10	
Individual 1 (lowest income)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Individual 10 (highest income)

e. **Where do you think that your income IN TEN YEARS would be on the scale?**

Imagine that we use a scale that represents the incomes of Swedes 5 years from now.

	1	2	3	4	5	6	7	8	9	10	
Individual 1 (lowest income)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Individual 10 (highest income)

16. How many percent of the Swedish population (18 years or older) do you think have a total yearly income which is lower than yours?
The income we refer to is the total yearly income which contains income from labor and capital before tax. Pensions before tax are also in this category. Subsidies like public unemployment payment are not part of the total yearly income.

I believe that [ ][ ][ ] percent have a lower income than I do.

17. Approximately how large was your income during 2010?
Yearly income is defined as in question 16.If your monthly salary is 18000 and you didn't have any other income your yearly income was 216000.

My yearly income was [ ][ ][ ][ ][ ][ ][ ] SEK in 2010.

18. a. Are you married or living with a partner?

1 ☐ Yes
2 ☐ No → *Go to question 19*

b. If you are married or living with a partner, approximately how big was your partner's total income before tax during 2010?
Yearly income is defined as in question 16. If you don't know the answer, please try to estimate it.

My partner's yearly income was [ ][ ][ ][ ][ ][ ][ ] SEK 2010.

19. What do you think was the average yearly income for Swedes aged 18 years or older during 2010?
Yearly income is defined as in question 16.

I believe that the average yearly income was [ ][ ][ ][ ][ ][ ][ ] SEK 2010.

20. How would you classify yourself in terms of class?
Please do not mark more than one alternative per question.

		"Working class" 1	"Lower middle class" 2	"Middle class" 3	"Upper middle class" 4	"Upper class" 5
a.	When you grew up	☐	☐	☐	☐	☐
b.	10 years ago	☐	☐	☐	☐	☐
c.	5 years ago	☐	☐	☐	☐	☐
d.	Currently	☐	☐	☐	☐	☐
e.	In 5 years	☐	☐	☐	☐	☐
f.	In 10 years	☐	☐	☐	☐	☐

21. To what extent do you agree with the following statements?

		Disagree completely 1	2	3	4	5	6	7	8	9	Agree completely 10
a.	If you are currently in a certain income group you will not belong to a different group in the future.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b.	Business and industries should be owned by the public sector.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c.	Competition between individuals is good, e.g. in school or in working life.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d.	Competition between businesses is good.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
e.	When born, all individuals have the same possibility to become economically successful.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
f.	One can only become rich at the expense of others.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

22. a. Is it mostly effort or luck that matters for how well an individual does economically in life?
Luck can for example mean having contacts.

	1	2	3	4	5	6	7	8	9	10	
Only luck	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Only effort

b. If we were to ask question 22a to all Swedes aged 18 and older, what do you think that the average answer would be?

	1	2	3	4	5	6	7	8	9	10	
Only luck	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Only effort

23. a. For your well-being, how dependent do you feel that you are on the public sector?

	1	2	3	4	5	6	7	8	9	10	
Not at all dependent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very dependent

b. If we were to ask question 23a to all Swedes aged 18 and older, what do you think that the average answer would be?

	1	2	3	4	5	6	7	8	9	10	
Not at all dependent	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very dependent

24. Do you believe that you are more or less dependent on the public sector than the average Swede?

1 ☐ Less dependent
2 ☐ As dependent as the average Swede
3 ☐ More dependent

25. How many percent of the public sector budget do you believe are used for health care, school and care for the elderly?

0-10%	11-20%	21-30%	31-40%	41-50%	51-60%	61-70%	71-80%	81-90%	91-100%
1	2	3	4	5	6	7	8	9	10
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

26. To what extent do you agree with the following statements?

		Disagree completely 1	2	3	4	5	6	7	8	9	Agree completely 10
a.	Changes in income taxes influence how much people choose to work	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b.	The public sector is efficient when redistributing money (no money is lost on the way)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c.	The public sector in Sweden spends money on the right things	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d.	People working in the public sector are generally doing a good job.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

27. What do you believe that one should teach one's children about the relative importance of luck and effort for economic success?
If you don't have children yourself, answer what you think that those with children should do.

	1	2	3	4	5	6	7	8	9	10	
Only luck matters	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Only effort matters

Part 4: Some final questions

28. Below we ask you a few questions where you can choose between getting a sum of money for sure or to take part in a lottery where you have a 50% chance of winning 3000 SEK and a 50% chance of not winning anything. We vary the alternative that you can get for sure but the lottery stays the same. Please note that all choices are hypothetical!

Mark the first square if you want the certain alternative and the second square if you want the lottery. Choose one alternative on each row.

	Certain alternative 1	Lottery 2
a.	500 SEK for sure ☐	50% chance of SEK 3000 ☐
b.	1000 SEK for sure ☐	50% chance of SEK 3000 ☐
c.	1200 SEK for sure ☐	50% chance of SEK 3000 ☐
d.	1400 SEK for sure ☐	50% chance of SEK 3000 ☐
e.	1600 SEK for sure ☐	50% chance of SEK 3000 ☐
f.	1800 SEK for sure ☐	50% chance of SEK 3000 ☐
g.	2000 SEK for sure ☐	50% chance of SEK 3000 ☐
h.	2500 SEK for sure ☐	50% chance of SEK 3000 ☐

29. In general, are you a person who is willing or unwilling to take risks?
Mark your answer below.

	1	2	3	4	5	6	7	8	9	10	
Not willing to take risks	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very willing to take risks

30. If you could choose, would you then prefer...

		Money today 1	Money in 12 months 2
a.	...1000 SEK today or 1000 SEK in 12 months?	☐	☐
b.	...1000 SEK today or 1170 SEK in 12 months?	☐	☐
c.	...1000 SEK today or 1340 SEK in 12 months?	☐	☐
d.	...1000 SEK today or 1510 SEK in 12 months?	☐	☐
e.	...1000 SEK today or 1680 SEK in 12 months?	☐	☐
f.	...1000 SEK today or 1850 SEK in 12 months?	☐	☐

g.	...1000 SEK today or 2170 SEK in 12 months?	☐	☐

31. **To what extent is the following statement true for you? I often postpone boring things, for example paying bills, and instead do something that is more fun.**

.

	1	2	3	4	5	6	7	8	9	10	
Not true at all	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Completely true

32. **Indicate your willingness to give money to charities.**

	1	2	3	4	5	6	7	8	9	10	
Not willing to give.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Very willing to give.

33. **a. If you were to win SEK 10 000, would you give anything to a charity?**

1 ☐ Yes

2 ☐ No → *Go to question 34*

b. How much of the SEK 10 000 would you give to charity?

I would give ☐☐☐☐☐ SEK.

34. To what extent are the following statements true for you?

a. **I always assume that other people have good intentions, if I don't get clear signals that this is not the case.**

	1	2	3	4	5	6	7	8	9	10	
Not true at all.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Completely true.

b. **If someone has helped me before, I go out of my way to help them.**

	1	2	3	4	5	6	7	8	9	10	
Not true at all.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	Completely true.

Thank you for answering the survey!

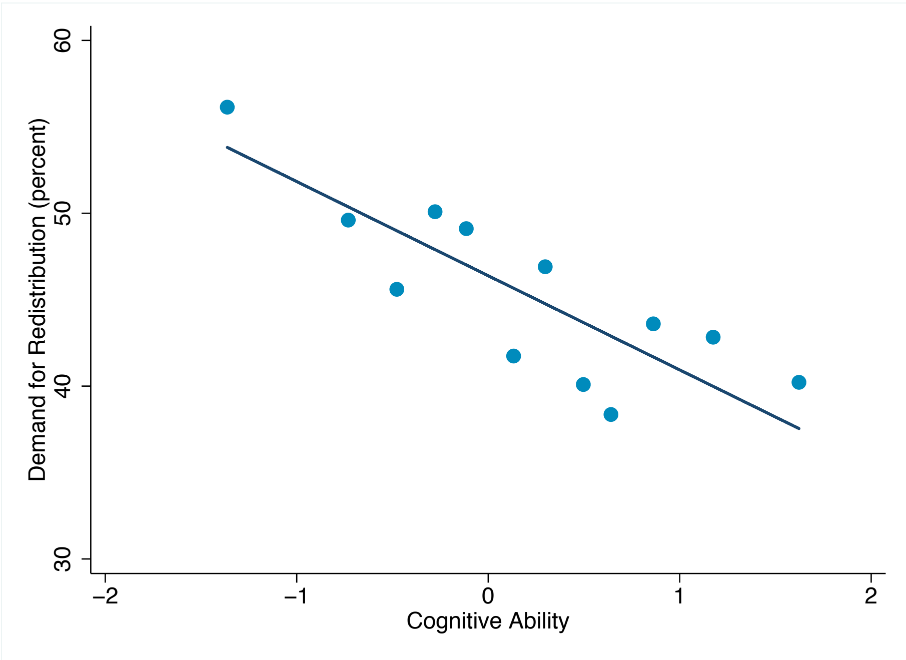


Figure 5.1: Cognitive Ability and Demand for Redistribution

Notes: We first regress demand for redistribution on age, education and mean income over twelve years. We add the mean to the residuals obtained from that regression and plot this variable against cognitive ability. Each bin corresponds to duo-deciles. $N = 271$.

Table 5.1: **Descriptive Statistics**

	(1) Population	(2) Male	(3) Survey Respondents	(4) Male Survey Respondents	(5) Male Resp. with Cogn. Skills
Labor Earnings	160,687	189,524	204,434	239,374	324,167
in 2009	(210,307)	(248,559)	(189,508)	(210,336)	(222,966)
Mean Income	120,858	161,738	160,676	188,966	261,311
1999 – 2009	(155,596)	(185,647)	(138,684)	(156,620)	(161,670)
Age	47.70	46.65	46.70	46.27	47.16
	(19.56)	(18.93)	(16.31)	(16.36)	(8.39)
Male	0.49	1	0.50	1	1
	(0.50)	(0)	(0.50)	(0)	(0)
Married	0.36	0.42	0.45	0.42	0.49
	(0.48)	(0.49)	(0.50)	(0.49)	(0.50)
Primary School	0.26	0.27	0.14	0.14	0.12
	(0.44)	(0.44)	(0.35)	(0.35)	(0.33)
Secondary School	0.44	0.46	0.44	0.50	0.47
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Higher Educ.	0.30	0.28	0.41	0.37	0.41
	(0.46)	(0.45)	(0.49)	(0.48)	(0.49)
Observations	8,832,171	3,792,911	1,549	779	271
Cogn. Skills	0.00	0.00	0.18	0.18	0.18
	(1.00)	(1.00)	(0.98)	(0.98)	(0.98)
Observations	1,335,130	1,335,130	271	271	271

Notes: Values in parenthesis represent standard deviations. All monetary variables expressed in SEK. Mean income refers to mean nominal annual income since 1999 or since the person entered the database. 1 USD $\approx$ 7 SEK.

Table 5.2: **Dependent Variable: Demand for Redistribution**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cognitive Skills	-6.665***	-6.488***	-5.375***	-5.448***	-5.035***	-5.432***	-5.638***	-5.930***	-6.526***
	(1.621)	(1.820)	(1.823)	(1.855)	(1.804)	(1.859)	(1.851)	(1.826)	
Annual Inc. 2010			-0.235**						
			(0.102)						
Annual Inc. 2010 Squared			0.001*						
			(0.000)						
Mean Inc.				-0.732***	-0.643***	-0.783***	-1.108***		
				(0.224)	(0.229)	(0.227)	(0.296)		
Mean Inc. Squared				0.005***	0.004**	0.005***	0.010***		
				(0.002)	(0.002)	(0.002)	(0.003)		
Spouse's Current Income							-0.999***		
							(0.458)		
Spouse's Current Income Squared							0.009		
							(0.006)		
Single							-15.228*		
							(9.052)		
Spouse's income zero							-2.244		
							(13.662)		
Future Relative Income							-0.050		
							(0.077)		
Relative Income Position in Youth							-1.005		
							(1.765)		
Mean Net Wealth							-0.021		
							(0.017)		
Risk Aversion					0.716			1.073	
					(0.843)			(0.848)	
Altruism						0.287			0.126
						(0.711)			(0.729)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Values in parenthesis represent standard errors estimated using the robust option in Stata. All monetary variables expressed in $10,000$ SEK. Mean income refers to mean nominal annual income since 1999 or since the person entered the database. 1 USD ≈ 7 SEK. *Spouse's Income Zero* is an indicator variable of spouse's income being zero interacted with a dummy of being married. *Future Relative Income* corresponds to the average of questions 15d and 15e in the survey and *Relative Income Position in Youth* represents the response to question 20a. *Mean Net Wealth* is the average of residential and financial wealth net of liabilities over the years for which there exists data between $1999-2006$. *Altruism* is self-reported willingness to give money to charities, taken from question 32. *Controls* denote age and education dummies.

Table 5.3: **Dependent Variable: Demand for Redistribution**

Male	-5.058***	-5.302***	-3.102**
	(1.334)	(1.341)	(1.389)
Mean Income			-0.312***
			(0.057)
Mean Income			0.000***
			(0.000)
Controls	No	Yes	Yes

Notes: Values in parenthesis represent standard errors estimated using the robust option in Stata. All monetary variables expressed in 10,000 SEK. Mean income refers to mean nominal annual income since 1999 or since the person entered the database. 1 USD ≈ 7 SEK. *Controls* denote age and education dummies. * significant at 10%, ** 5%, *** 1%.

Table 5.4: **Dependent Variable: Demand for Redistribution**

		Age	
	33 − 42	43 − 53	54 − 61
Cognitive Skills	-4.725	-3.673	-7.239***
	(3.188)	(2.793)	(3.269)
Mean Income	-1.961***	-0.909**	-0.560Â
	(0.510)	(0.351)	(0.489)
Mean Income Squared	0.024***	0.005**	0.005
	(0.007)	(0.002)	(0.004)
Controls	Yes	Yes	Yes
Observations	94	97	80

Notes: Values in parenthesis represent standard errors estimated using the robust option in Stata. All monetary variables expressed in 10,000 SEK. Mean income refers to mean nominal annual income since 1999 or since the person entered the database. 1 USD ≈ 7 SEK. *Controls* denote age and education dummies. * significant at 10%, ** 5%, *** 1%.

MONOGRAPH SERIES

1. Michaely, Michael *The Theory of Commercial Policy: Trade and Protection*, 1973
2. Söderström, Hans Tson *Studies in the Microdynamics of Production and Productivity Change*, 1974
3. Hamilton, Carl B. *Project Analysis in the Rural Sector with Special Reference to the Evaluation of Labour Cost*, 1974
4. Nyberg, Lars and Staffan Viotti *A Control Systems Approach to Macroeconomic Theory and Policy in the Open Economy*, 1975
5. Myhrman, Johan *Monetary Policy in Open Economies*, 1975
6. Krauss, Melvyn *International Trade and Economic Welfare*, 1975
7. Wihlborg, Clas *Capital Market Integration and Monetary Policy under Different Exchange Rate Regimes*, 1976
8. Svensson, Lars E.O. *On Competitive Markets and Intertemporal Resources Allocation*, 1976
9. Yeats, Alexander J. *Trade Barriers Facing Developing Countries*, 1978
10. Calmfors, Lars *Prices, Wages and Employment in the Open Economy*, 1978
11. Kornai, János *Economics of Shortage*, Vols I and II, 1979
12. Flam, Harry *Growth, Allocation and Trade in Sweden. An Empirical Application of the Heckscher-Ohlin Theory*, 1981
13. Persson, Torsten *Studies of Alternative Exchange Rate Systems. An Intertemporal General Equilibrium Approach*, 1982
14. Erzan, Refik *Turkey's Comparative Advantage, Production and Trade Patterns in Manufactures. An Application of the Factor Proportions Hypothesis with Some Qualifications*, 1983
15. Horn af Rantzien, Henrik *Imperfect Competition in Models of Wage Formation and International Trade*, 1983
16. Nandakumar, Parameswar *Macroeconomic Effects of Supply Side Policies and Disturbances in Open Economies*, 1985
17. Sellin, Peter *Asset Pricing and Portfolio Choice with International Investment Barriers*, 1990
18. Werner, Ingrid *International Capital Markets: Controls, Taxes and Resources Allocation*, 1990
19. Svedberg, Peter *Poverty and Undernutrition in Sub-Saharan Africa: Theory, Evidence, Policy*, 1991
20. Nordström, Håkan *Studies in Trade Policy and Economic Growth*, 1992

21. Hassler, John, Lundvik, Petter, Persson, Torsten and Söderlind, Paul *The Swedish Business Cycle: Stylized facts over 130 years*, 1992
22. Lundvik, Petter *Business Cycles and Growth*, 1992
23. Söderlind, Paul *Essays in Exchange Rates, Business Cycles and Growth*, 1993
24. Hassler, John A.A. *Effects of Variations in Risk on Demand and Measures of Business Cycle Comovements*, 1994
25. Daltung, Sonja *Risk, Efficiency, and Regulation of Banks*, 1994
26. Lindberg, Hans *Exchange Rates: Target Zones, Interventions and Regime Collapses*, 1994
27. Stennek, Johan *Essays on Information-Processing and Competition*, 1994
28. Jonsson, Gunnar *Institutions and Incentives in Monetary and Fiscal Policy*, 1995
29. Dahlquist, Magnus *Essays on the Term Structure of Interest Rates and Monetary Policy*, 1995
30. Svensson, Jakob *Political Economy and Macroeconomics: On Foreign Aid and Development*, 1996
31. Blix, Mårten *Rational Expectations and Regime Shifts in Macroeconometrics*, 1997
32. Lagerlöf, Nils-Petter *Intergenerational Transfers and Altruism*, 1997
33. Klein, Paul *Papers on the Macroeconomics of Fiscal Policy*, 1997
34. Jonsson, Magnus *Studies in Business Cycles*, 1997
35. Persson, Lars *Asset Ownership in Imperfectly Competitive Markets*, 1998
36. Persson, Joakim *Essays on Economic Growth*, 1998
37. Domeij, David *Essays on Optimal Taxation and Indeterminacy*, 1998
38. Flodén, Martin *Essays on Dynamic Macroeconomics*, 1999
39. Tangerås, Thomas *Essays in Economics and Politics: Regulation, Elections and International Conflict*, 2000
40. Lidbom, Per Pettersson *Elections, Party Politics and Economic Policy*, 2000
41. Vestin, David *Essays on Monetary Policy*, 2001
42. Olofsgård, Anders *Essays on Interregional and International Political Economics*, 2001
43. Johansson, Åsa *Essays on Macroeconomic Fluctuations and Nominal Wage Rigidity*, 2002
44. Groth, Charlotta *Topics on Monetary Policy*, 2002
45. Gancia, Gino A. *Essays on Growth, Trade and Inequality*, 2003
46. Harstad, Bård *Organizing Cooperation: Bargaining, Voting and Control*, 2003
47. Kohlscheen, Emanuel *Essays on Debts and Constitutions*, 2004
48. Olovsson, Conny *Essays on Dynamic Macroeconomics*, 2004

49. Stavlöt, Ulrika *Essays on Culture and Trade*, 2005
50. Herzing, Mathias *Essays on Uncertainty and Escape in Trade Agreements*, 2005
51. Bonfiglioli, Alessandra *Essays on Financial Markets and Macroeconomics*, 2005
52. Pienaar, Natalie *Economic Applications of Product Quality Regulation in WTO Trade Agreements*, 2005
53. Song, Zheng *Essays on Dynamic Political Economy*, 2005
54. Eisensee, Thomas *Essays on Public Finance: Retirement Behavior and Disaster Relief*, 2005
55. Favara, Giovanni *Credit and Finance in the Macroeconomy*, 2006
56. Björkman, Martina *Essays on Empirical Development Economics: Education, Health and Gender*, 2006
57. Larsson, Anna *Real Effects of Monetary Regimes*, 2007
58. Prado, Jr., Jose Mauricio *Essays on Public Macroeconomic Policy*, 2007
59. Tonin, Mirco *Essays on Labor Market Structures and Policies*, 2007
60. Queijo von Heideken, *Virgina Essays on Monetary Policy and Asset Markets*, 2007
61. Finocchiaro, Daria *Essays on Macroeconomics*, 2007
62. Waisman, Gisela *Essays on Discrimination and Corruption*, 2008
63. Holte, Martin Bech *Essays on Incentives and Leadership*, 2008
64. Damsgaard, Erika Färnstrand *Essays on Technological Choice and Spillovers*, 2008
65. Fredriksson, Anders *Bureaucracy, Informality and Taxation: Essays in Development Economics and Public Finance*, 2009
66. Folke, Olle *Parties, Power and Patronage: Papers in Political Economics*, 2010
67. Drott, David Yanagizawa *Information, Markets and Conflict: Essays on Development and Political Economics*, 2010
68. Meyersson, Erik *Religion, Politics and Development: Essays in Development and Political Economics*, 2010
69. Klingelhöfer, Jan *Models of Electoral Competition: Three Essays on Political Economics*, 2010
70. Perrotta, Maria Carmela *Aid, Education and Development*, 2010
71. Caldara, Dario *Essays on Empirical Macroeconomics*, 2011
72. Mueller, Andreas *Business Cycles, Unemployment and Job Search: Essays in Macroeconomics and Labor Economics*, 2011
73. Von Below, David *Essays in Climate and Labor Economics*, 2011
74. Gars, Johan *Essays on the Macroeconomics of Climate Change*, 2012
75. Spiro, Daniel *Some Aspects on Resource and Behavioral Economics*, 2012
76. Ge, Jinfeng *Essays on Macroeconomics and Political Economy*, 2012

77. Li, Yinan *Institutions, Political Cycles and Corruption: Essays on Dynamic Political Economy of Government*, 2013
78. Håkanson, Christina *Changes in Workplaces and Careers*, 2013
79. Qin, Bei *Essays on Empirical Development and Political Economics*, 2013
80. Jia, Ruixue *Essays on the Political Economy of China's Development*, 2013
81. Campa, Pamela *Media Influence on Pollution, and Gender Equality*, 2013