EMPIRICAL ESSAYS *on* HEALTH, WEALTH *and* POLITICS

Till Stowasser

Ludwig Maximilians University Munich

December 16th, 2011

DOCTORAL THESIS

Empirical essays on health, wealth and politics

Inaugural-Dissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der wirtschaftswissenschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

2011

vorgelegt von
Till Stowasser

Referent:	Prof. Dr. Joachim Winter
Korreferent:	Prof. Dr. Florian Englmaier
Mündliche Prüfung:	30. April 2012
Promotionsabschlussberatung:	16. Mai 2012

für Horst

Acknowledgements

First and foremost, I would like to thank my supervisor, Joachim Winter, for the support and guidance he provided throughout my dissertation. I am especially indebted to his notable decision of taking me under his wing, despite me being a lateral entrant to the field of econometrics. He gave me the important opportunity to steer my dissertation in a new direction. The same can be said about my second reviewer, Florian Englmaier, without whom my thesis would not have been the same. I am grateful for his constant encouragement and for not letting me give up on our joint project (chapter 3), even when all odds were against us. I also thank Alexander Danzer, who kindly agreed to serve as third reviewer on my dissertation committee.

In addition, I am grateful to my other two co-authors, Florian Heiss and Daniel McFadden, for their collaboration that is a source of inspiration and motivation. I very much appreciate the valuable suggestions and comments from colleagues and participants at seminars and conferences, of which Georg Gebhardt, Susanne Hoffmann, Michael Hurd, Francis Kramarz, Pierre-Carl Michaud, Nicolas Sauter, and Robert Willis deserve special mention. My gratitude also extends to the Labor and Population research division at the *RAND Corporation* (Santa Monica, CA), particularly to Arie Kapteyn, for hosting me as a visiting scholar. My stay at RAND was extremely helpful in shaping the research ideas that turned into chapters 1 and 2 of this thesis. Since sound empirical research is impossible without excellent data, I also would like to thank student assistants Johannes Kümmel, Kirill Lindt, Thomas Hattenbach, and Nikolaos Karygiannis for their contribution to gathering the data for chapter 3. And since looks matter, too, I thank Holger Gerhardt for helping me to squeeze every last drop of typographical compliance out of LaTeX.

It was a pleasure to share doctoral studies with Silvia Appelt, Susanne Hoffmann, Anno Stolper, Lisa Stadler, Joachim Klein, Johannes Maier, Felix Reinshagen, Jan Schikora, Matthias Fahn, Amelie Wuppermann, the entire team at the *Seminar for Empirical Economic Research*, and many others. I am

grateful to everyone at the *Munich Graduate School of Economics* and the Economics Department of *Ludwig Maximilians University*. A special thanks goes to Silke Englmaier. Financial support from the *German Research Foundation (DFG)* and the *German Academic Exchange Service (DAAD)* is gratefully acknowledged.

Finally, I would like to thank my family for all their love and encouragement. I am grateful to my parents who made me the person I am, and who supported me in all my pursuits. I am especially indebted to my loving husband, Kenji: Without your patience, support, and understanding, none of this would have been possible. Thank you.

Till Stowasser

Contents

List of tables

List of figures

Preface

This dissertation consists of three empirical essays on questions related to health, wealth, and politics. Each study constitutes a self-contained chapter. While chapters 1 and 2 contribute to the literature of health economics and are closely connected to one another, the final chapter 3 investigates an entirely unrelated topic from the area of political economics.

The first two essays deal with one of the most fundamental questions in health economics: Why are the economically successful not only wealthier but also healthier than the less affluent? There is little doubt about the existence of this so-called socio-economic gradient in health, but there remains a vivid debate about its source (see Cutler *et al.*, 2011). The traditional view, most prominently advocated within epidemiological research, argues that economic resources determine health. Social scientists were among the first to suggest that causality may also work in the opposite direction, with health influencing future economic success via the development of human capital. Yet another plausible explanation is that some unobserved third factors have a common impact on wealth and health, without there being any causal relationship between the two.

Certainly, these three views need not be mutually exclusive. In fact, there is reason to believe that reality is best described by a lifelong interplay between economic conditions and well-being, with different pathways being relevant at different stages of the life cycle. Similarly, the links between health and wealth are arguably dynamic in nature: Socio-economic status (SES) at one point in time may impact future health, which in turn has implications for economic outcomes that lie even further ahead. This complicates the already demanding task of disentangling the many plausible channels underlying the health-wealth nexus. Yet, the discrimination among these rivalling hypotheses is paramount since policy recommendations will critically depend on the nature and the sources of the gradient. For instance, if causal links between wealth and health were confirmed, society would likely benefit from more universal access to health care and redistributive economic policy. Yet, if the notion of causality were rebutted,

resources would be better spent on influencing health knowledge, preferences, and ultimately the behavior of individuals.

Chapter 1, which is based on joint work with Florian Heiss, Daniel McFadden and Joachim Winter, starts by reviewing the methodological challenges – simultaneity and omitted-variable bias – involved in testing the causal relationships between SES and health. This is followed by a discussion of the conventional solution to both of these problems, which consists of instrumental-variables (IV) approaches that exploit exogenous variation from natural experiments. Since IV strategies are not always persuasive in practice, we continue by describing an alternative approach of testing for the absence of causal channels developed by Adams *et al.* (2003). Their methodology seeks identification without the need to isolate exogenous variation in economic variables by resting on the purely statistical concept of Granger causality. Their finding that socio-economic status is unlikely to be causal for mortality and a wide range of health incidences in the US American population over the age of 70, has sparked much controversy in the subsequent literature. Given that – besides methodological issues – most of the criticism centers around the external validity of their results, we repeat their analysis using the full range of data that has since become available in the longitudinal *Health and Retirement Study* (HRS). Compared to the original study – which uses a subsample of HRS – this represents a much more comprehensive set of data, both in terms of observations years and age ranges covered. In stark contrast to the original study, we find that it is much harder to statistically reject the activity of causal links in more encompassing samples. Importantly, this result is not solely driven by the inclusion of younger individuals, as the mere growth in sample size already leads to higher rejection rates for causality tests, hinting at Adams *et al.*'s results being partly driven by insufficient test power.

In light of these findings, we discuss three important methodological drawbacks of Adams *et al.*'s original framework that merit attention in future research. First, the underlying notion of health dynamics, with health being modelled as a first-order Markov process, is arguably an inadequate description of latent health capital as envisioned by Grossman (1972). Second, the failure to account for individual heterogeneity complicates the discrimination between true causal links and unobserved common effects. Third, even if third factors were convincingly controlled for, their approach would fail to uncover the exact mechanisms through which SES influences health. Yet, while knowledge of the general presence of causal links is important in its own right, the identification of specific pathways is equally crucial from a policy perspective.

The second chapter aims at addressing all three of these methodological weaknesses by exploiting the recent availability of retrospective data within HRS: To begin with, the information on early-life circumstances enables the incorporation of longer health histories, which arguably provides a better description of health dynamics. Furthermore, ample information on family background and family SES should allow for reasonable proxy-control of individual heterogeneity that drives a wedge between Granger causality and causation in a more structural sense. Moreover, conditioning on both historic and contemporary information, will shed light on the question *when* the SES gradient in health is established, which, once again, has important implications for optimal policy. In addition to this, I decompose the broad effect of SES on morbidity to gauge *how* wealth, income and education leave their mark on an individual's general health.

Succinctly, the main conclusions of chapter 1 remain intact if longer health histories and proxies for individual effects are incorporated, which lends support to a causal interpretation of observed correlations. While there is not much evidence for causal links from SES to acute health innovations in an elderly population, causality appears likely for mortality, changes in mental health and general well-being. Results for chronic diseases and functional health are a bit mixed, as they suffer from low test power. In line with the literature on early life circumstance (see Almond and Currie, 2011), I find that childhood health has lasting effects for adult outcomes. This, however, does not render contemporary factors unimportant. Furthermore, it seems that economic status exerts its influence both through the accumulation of health conditions, as well as via more direct channels. Interestingly, there is no evidence that wealth or income can act as a shield from the health consequences of illnesses – at least not in a post-retirement US population that enjoys near-universal health care coverage through Medicare. However, the prevalence of chronic conditions appears to have stronger adverse effects on overall health among the lower educated. This is indicative of difficulties for this group to take measures of disease management that are necessary to alleviate further negative health consequences from illnesses such as lung disease or diabetes.

For the third and final chapter of this thesis, which is joint work with Florian Englmaier, we abandon the nexus of health and wealth and turn our attention to something completely different: the political economics of savings bank lending. We contribute to the literature of political business cycles (see Nordhaus, 1975; and MacRae, 1977), by testing the hypothesis that even *local* incumbent politicians try to boost economic conditions in pre-election periods. We exploit a peculiarity in the German public bank-

ing system where municipal politicians are by law involved in the management of local savings banks. We use the different timing of municipal elections across states and the existence of cooperative banks as a control group to estimate the causal effect of elections on lending policy of German savings banks. To this end, we use a novel, largely hand-collected dataset, merging municipal election outcomes for a subset of German states with accounting information of local banks. Econometrically, we conduct difference-in-difference as well as triple-difference estimation embedded in a fixed-effects panel data setup.

In line with our predictions, we find that savings banks systematically extend more credits in pre-election periods. This effect is not only statistically significant but also economically relevant: it amounts to a 2%–3% election-induced increase in the stock of lending, when our preferred empirical specification is used. Note that this increase is relative to the total stock in bank lending. If we were to model the extension of new credit contracts alone, relative effect sizes would certainly be substantially larger. To provide a better sense for the actual magnitude of the effect, consider that its absolute size amounts to an average of EUR 30.6 million extra stock in lending per bank.

The election effect is robust to various alternative specifications. Importantly, it is not present with cooperative banks that are very similar to savings banks but that lack their political connectedness. Our hypotheses are further strengthened by the fact that lending cycles only occur with municipal but not with state or federal elections. Furthermore, we find weak evidence for overly prudent lending policies after elections, consistent with a binding credit constraint that banks face so that they have to make up for excessive pre-election lending. Our evidence also suggests that the ability to induce political lending cycles depends on the degree of dominance of the incumbent party. Finally, we find significant differences in effect sizes across states, suggesting an interaction with other details of the institutional environment, which we intend to further investigate in future research.

Our results stand in contrast to the empirical finance literature, which has long suspected that the behavior of government-controlled banks is rather different from that of private financial institutions, but which has not yet found compelling evidence for politically induced lending in highly developed countries such as Germany (see Dinç, 2005). Given that savings banks are holding the largest market share in the German private-customer deposit market and that they are the most important lender to the *Mittelstand* (SMEs) – that is considered the backbone of the German economy – it is potentially worrisome to find their policies distorted.

Chapter 1

"Healthy, wealthy, and wise" revisited: An analysis of the causal pathways from socio-economic status to health

1.1 Introduction

In health economics, there is little dispute that the socio-economic status (SES) of individuals is positively correlated with their health status. The size of the body of literature documenting that wealthy and well-educated people generally enjoy better health and longer life is impressive.[1] The robustness of this association is underscored by the fact that the so-called socio-economic gradient in health has been detected in different times, countries, populations, age-structures, and for both men and women. Moreover, the results are largely insensitive to the choice of SES measures (such as wealth, income, education, occupation, or social class) and health outcomes.

While the existence of the gradient may be uncontroversial, the same cannot be said about its explanation. Medical researchers, economists and other social scientists have developed a large number of competing theories that can broadly be categorized as follows: there may be causal effects from SES to health, causal effects that work in the opposite direction, and unobserved common factors that influence both variables in the same direction without a causal link between the two. Distinguishing among these explanations is important since they have different implications for public policy aimed at improving overall well-being. For instance, if causal links between wealth and health were confirmed, society would likely benefit

[1] Smith (1999) and Goldsmith (2001) provide extensive surveys of the earlier literature. A brief summary of more recent contributions to this field can be found in Michaud and van Soest (2008).

from more universal access to health care and redistributive economic policy. Yet, if such causal links were rebutted, resources would be better spent on influencing health knowledge, preferences, and ultimately the behavior of individuals.

Besides its importance, the discrimination between these alternative hypotheses also poses a great methodological challenge since the variation found in observational data is typically endogenous. This is especially true for cross-sectional data, which only offers a snapshot of the association between health and wealth. Without further information on the history of both variables, the researcher faces a fundamental simultaneity problem, which makes the identification of causal paths a hopeless venture. A possible remedy consists of finding some sort of exogenous variation in SES or health to infer causality and the direction of its flow. This search, however, is typically quite difficult because convincing instrumental variables are very hard to come by. As a consequence, researchers often face the unattractive choice between the easy path of ignoring the endogeneity problem, which casts serious doubts on any drawn conclusions, and the more involved use of IV strategies that critically rely on the untestable quality of the instruments.

The nexus of health, wealth, and wisdom is also the subject of the study by Adams *et al.* (2003) (HWW henceforth). The authors propose an innovative approach that attempts to solve the above trade-off, on the premise that causal inference may be possible without having to isolate exogenous variation in SES. Their identification strategy consists of two main ingredients: First, they exploit the dynamic nature of panel data, focusing on health *innovations* rather than the prevalence of medical conditions. Second, they make use of the so-called Granger causality framework, which represents a purely statistical approach to the theory of causation. The great advantage of working with this alternative concept is that the detection of potential Granger causality is a rather easy task. While knowledge on the existence of Granger causality may not be useful in its own right, it allows for tests on the *absence* of "true" causality in a structural sense.

Applying this framework to the first three waves of the *Asset and Health Dynamics among the Oldest Old (AHEAD)* survey study, HWW find that in an elderly US population, causal channels that operate from wealth to health are an exception rather than the rule: while causality cannot be ruled out for some chronic and mental conditions for which health insurance coverage is not universal, SES is unlikely to be causal for mortality and most other illnesses. Considering these strong results, as well as the methodological novelty of HWW's approach, it is not surprising that their work has

subsequently been the subject of vivid debate within the literature.[2] So far, the focus has clearly been on the validity of HWW's identification strategy in general, with some calling into question the ability to truly infer causality with a concept that arguably is a rather sparse characterization of causal properties.

We certainly agree that HWW's model would benefit from certain methodological refinements and plan to implement these in future research. For the present project, however, we deliberately leave the econometrics unchanged, to study a different aspect that also merits attention: the stability of HWW's results when confronted with new data that allows for hypothesis tests of greater statistical power. Special interest lies in assessing whether the somewhat surprising absence of direct causal links from SES to most medical conditions is a robust finding or perhaps the artifact of a particular data sample. Since the publication of HWW's original article, the AHEAD survey has been incorporated into the more-encompassing *Health and Retirement Study (HRS)*. This permits deviations from HWW's data benchmark along the following dimensions: the same individuals can be tracked for a longer period of time, the analysis can be extended to new cohorts of respondents, and the working sample can be widened by including younger individuals aged 50 and older. The last point is of special interest as it offers variation in health insurance status that is not available in a Medicare-eligible population. To understand which of these data changes contribute to any deviating conclusions, we do not apply the whole bundle of modifications at once. Instead, we estimate the model multiple times, by applying it to several different data samples, which are gradually augmented along the dimensions just outlined.

We lay out the theoretical background of our analysis in section 1.2, where we review the potential explanations for the association between SES and health and specify the econometric challenges that arise when trying to discriminate among them. This is followed by a discussion of how to address these challenges. The conventional solution consists of employing IV methods whose identification strategy is assessed in section 1.3. Section 1.4 describes the approach proposed by HWW. A reanalysis of HWW with new data is presented in section 1.5. Section 1.6 concludes and outlines topics for future research.

[2] As an example, consider the comments to HWW by Adda *et al.* (2003), Florens (2003), Geweke (2003), Granger (2003), Hausman (2003), Heckman (2003), Hoover (2003), Poterba (2003), Robins (2003), and Rubin and Mealli (2003) published in the same issue as the original article.

1.2 The difficulty of causal inference

1.2.1 The issue: Potential channels between SES and health

Correlation does not necessarily imply causation. This insight is one of the main lessons every empiricist needs to internalize. At times, however, it can be tempting to neglect this admonition, especially when a causal interpretation of a joint motion of two variables is very intuitive. The relationship between SES and health is a prime example for such a situation. As an illustration, consider table 1.1, which lists household median wealth of HRS respondents arrayed against self-reported health status. Here, the socio-economic gradient is prominently on display as median wealth monotonically decreases with impairing health self-reports – an observation that is remarkably stable over time.

What could be more natural than to interpret this strong correlation as a causal influence of wealth on health? After all, it is the explanation best in line with conventional wisdom: money can buy (almost) anything – even better health. Yet, the most intuitive conclusion may not necessarily be the only valid one. In fact, there are two additional hypotheses for the association of SES and medical conditions: the causation could flow from the latter to the former, and the correlation may actually be spurious, with third factors affecting health and wealth in a similar way. This section describes these rivalling theories and gives an overview of the most commonly-cited potential pathways between SES and health (see Adler and Ostrove, 1999; Smith, 1999; Goldsmith, 2001; and Cutler *et al.*, 2011 for more extensive reviews).

Table 1.1. *The socio-economic gradient in health*
Median wealth by self-rated health status

Self-rated health	1992	1996	2000	2004	2008
Excellent	155.6	192.0	256.0	331.4	363.0
Very good	122.1	159.0	202.6	240.0	304.0
Good	82.5	106.2	130.6	160.0	194.0
Poor	46.7	62.2	69.0	75.1	86.1
Fair	19.5	35.0	36.5	39.7	48.1

Notes: Calculations by authors based on HRS data. Numbers reported in thousands of 1997 USD.

Hypothesis A: SES has a causal influence on health ouctomes

This is the hypothesis most energetically advocated within the epidemiological literature. While it is true that the main contribution from economists consists of formulating alternative interpretations of the socio-economic health gradient (see hypothesis B, below), it should be emphasized that they are not on record of categorically challenging hypothesis A, either. Below, we list the most prominent theories of channels through which SES may have a causal effect on health.

Channel A1: Affordability of health care. This potential channel is arguably one of the most intuitive explanations and may be active both before and after an individual is diagnosed with an illness. For one, varying SES may be responsible for differentials in the onset of health conditions as poorer people may be overly sensitive to the costs of preventive health care. In addition, wealth could play a crucial role in determining the quality or even the plain affordability of medical treatments, once they become necessary.

Channel A2: The psychological burden of being poor. Medical scientists increasingly emphasize the importance of psychological consequences of low SES. They argue that low-wage employment is typically associated with a high degree of work monotonicity and low job control, leading to psychosocial stress. Similarly, economically disadvantaged individuals are believed to be repeatedly exposed to episodes of high emotional discomfort, either due to long phases of unemployment or a general feeling of social injustice. When accumulated, these stressful experiences may well have strong adverse effects on physical health as well. Furthermore, adverse wealth shocks – such as the loss of life savings in a stock market crash – are likely to cause anxiety and depression, representing a more immediate avenue through which SES may impact health.

Channel A3: Environmental hazards. Another line of argument is that the exposure to perilous environments is considerably higher for the poor. This may concern job-related risks since it can be argued that workplace safety is lower and physical strain higher for poorly-paid occupations. The reasoning also extends to people's living environments as neighborhood safety, dwelling condition, air and water quality, etc. are usually much better in exclusive residential areas.

Channel A4: Health knowledge. Considering that education is an integral component of SES, it is conceivable that part of the correlation be-

tween SES and health is attributable to differences in health knowledge. According to this argument, information on medical risk factors or the importance of preventative care may be more widespread among the highly educated and wealthy, leading to healthier lifestyles and lower morbidity rates among this group.

Channel A5: Risk behaviors. An often-cited pathway through which SES may influence health is the asymmetric distribution of unhealthy lifestyles such as smoking, drinking and poor diet. To the extent that all of these vices are less common among the rich, health differentials may in fact be driven by SES variables. Note that the question of why smoking, excess alcohol consumption, and obesity are especially prevalent in lower social classes, is interesting in its own right, with channels A2 and A4 potentially accounting for part of this relationship.

Hypothesis B: Health has a causal influence on SES outcomes

Economists and other social scientists were among the first to challenge the conception that causal mechanisms would work their way exclusively from SES to health. Much of this research is inspired by Grossman's (1972) health production framework, which models the impact of health capital on savings, labor market participation, and retirement decisions. We believe the following three channels to be the most important in describing causal effects from health to SES outcomes.

Channel B1: Productivity and labor supply. Arguably, the most relevant reason why health may be causal for SES outcomes can be found on the labor market. The productivity of an individual in poor health is generally lower than that of someone whose physical robustness allows for longer working hours, less absenteeism, and better career options. As a consequence, frail people will tend to earn lower wages and accumulate less assets throughout their life course. Adverse health shocks may even be so severe that people are forced to leave the labor market altogether, depriving them from any realistic chance to improve their SES.

Channel B2: Life expectancy and time preferences. To the extent that severe illnesses increase mortality risks, there may be an impact of poor health on time preferences. Life-cycle models predict that the optimal response to a perceived reduction in life expectancy is to move consumption from an uncertain future towards the presence. Thus, a

history of dire medical events may induce individuals to dissave faster, establishing a causal link from health to SES.

Channel B3: Medical care expenditures. The most immediate form of impact health events can have on financial endowments are out-of-pocket costs of medical care. While it can be argued that the influence of this pathway should only be modest in size, this is certainly untrue for people without health insurance. In many cases, not even the insured are completely shielded from medical bills: the existence of deductibles and lifetime coverage limits poses great financial threats especially for the chronically ill.

Hypothesis C: SES and health are jointly caused by an unobserved third factor

This hypothesis makes the case that the association between health and wealth could have other reasons than causal mechanisms between the two: There may be hidden third factors with a common influence on both SES and health, rendering the correlation among the latter spurious. This distinction is vital since policies that aim at improving health outcomes by, say, redistributing wealth are bound to be ineffective, as long as the true common cause remains unaffected.

Channel C1: Unobserved genetic heterogeneity. A good candidate for an unobserved common cause is genetic disposition. For instance, genetic frailty may reduce the physical resistance as well as the intellectual and professional skills of an individual. In such cases, health will be poorer and SES will be lower despite the absence of causal links among the two.

Channel C2: Unobserved family background. Genetic endowment is not the only determinant of people's physical and personal traits. Similarly influential are matters of parentage and upbringing. Especially prenatal and early-childhood nutrition as well as stress are believed to have lasting negative effects on well-being and functional abilities, establishing an association between health and SES that is similar to that of Channel C1.

Channel C3: Unobserved preferences. Irrespective of whether they are inherited or learned, preferences that influence certain behavior and lifestyles are another often-cited source of common effects. The prime example are descendants of dysfunctional families, who adopt both

the unhealthy lifestyles (such as poor diet or smoking) and the unambitious attitudes towards education and work by which they are surrounded. Another example are time preferences: overly myopic people will underinvest in preventative medical care and in education since in both cases pay-offs will materialize in a distant future, to which only little importance is attached.

1.2.2 The challenges: Simultaneity and omitted variables

The fact that all of the aforementioned hypotheses are generally plausible, makes the inference on causation a methodologically challenging task. Suppose – as is the case for the remainder of this paper – we were interested in testing the validity of hypothesis A, that is whether SES has a causal effect on health outcomes. Ideally, we would want our analysis to rely on truly exogenous variation in SES variables, similar to that attained in controlled experiments. The reality for economists, however, is far from being ideal since the sources of variation we find in observational data is unknown to us. As a consequence, causal variables are potentially endogenous themselves.

The possible sources of endogeneity in the wealth-health case have been described in section 1.2.1. Ultimately, they generate two fundamental econometric challenges: we have to distinguish hypothesis A from hypothesis B, and hypothesis A from hypothesis C. As we discuss below, the first consists of dealing with a simultaneity problem, and the second of finding a solution to the problem of omitted variables.

Challenge 1: The simultaneity problem (hypothesis A vs. hypothesis B)

Imagine for a moment that hypothesis C could be dismissed, so that any association between SES and health had to be due to either hypothesis A or B. Even with this kind of simplification in place, the identification of SES causality for health is still difficult. Of course, we could regress our health variable of interest (H_i) on SES (S_i) and a vector of exogenous control variables ($\mathbf{X_i}$), estimating the following equation with OLS:

$$H_i = \theta_0 + \theta_s S_i + \mathbf{X_i'}\boldsymbol{\theta_x} + \eta_i, \tag{1.1}$$

where i denotes the unit of observation and η_i is the residual. Yet, the crucial question is if we could interpret the parameter estimate $\widehat{\theta_s}$ as the causal

effect of SES on health. The answer would be affirmative if the structural model were to look like

$$\begin{aligned} \mathrm{E}(H_i|S_i, \mathbf{X_i}) &= \alpha + \beta S_i + \mathbf{X_i'}\gamma, \\ \mathrm{E}(S_i|H_i, \mathbf{X_i}) &= \mathrm{E}(S_i|\mathbf{X_i}). \end{aligned}$$

This model describes a world in which causality only flows from SES to health, with β capturing the true causal effect. In this world, $\widehat{\theta_s}$ would indeed have a causal interpretation, with $\operatorname{plim}\widehat{\theta_s} = \beta$. However, the existence of hypothesis B indicates that the above model may not be a realistic description of reality. In fact, the true structural model is likelier to look like

$$\mathrm{E}(H_i|S_i, \mathbf{X_i}) = \alpha + \beta S_i + \mathbf{X_i'}\gamma, \tag{1.2}$$

$$\mathrm{E}(S_i|H_i, \mathbf{X_i}) = a + bH_i + \mathbf{X_i'}\boldsymbol{c}, \tag{1.3}$$

with β again measuring the true causal effect of SES on health, and b capturing any causation working its way in the opposite direction. Equations 1.2 and 1.3 describe a standard simultaneous-equation model (SEM) as both dependent variables are jointly determined with each being a function of the other. When trying to estimate this SEM by simply running regression equation 1.1, $\widehat{\theta_s}$ will be subject to simultaneous-equation bias, picking up the information conveyed in b as well. As a result, the parameter of interest, β, is not identified, making a test for causation of SES to health all but impossible.

Challenge 2: The omitted-variable problem (hypothesis A vs. hypothesis C)

Even in the absence of challenge 1, we would still face the problem of having to discriminate between hypotheses A and C. Presume we were able to plausibly exclude causal paths from health to SES. In this case, the identification problem no longer consists of confounding the causal effect of wealth on health with reverse causality. Instead, the question arises if an association between both variables is attributable to causality at all since it could also stem from a joint reaction to a third factor. As the review of hypothesis C has shown, all of these potential common causes (such as genetics or preferences) are inherently unobservable, rendering challenge 2 an omitted-variable problem.

Suppose the true structural model is best described by

$$\mathrm{E}(H_i|S_i, \mathbf{X_i}, C_i) = \alpha + \beta S_i + \mathbf{X_i'}\gamma + \delta C_i, \tag{1.4}$$

with C_i standing for an individual-specific variable that influences both SES and health. If this common cause were observable, we could simply include

it in our regression function and the causal effect, β, would be readily identified. However, given its omitted-variable nature, C_i will be swamped into the error term, as the comparison of the structural model in error form (equation 1.5) with the estimable model (equation 1.6) demonstrates:

$$H_i = \alpha + \beta S_i + \mathbf{X_i'}\gamma + \delta C_i + \epsilon_i, \tag{1.5}$$

$$H_i = \alpha + \beta S_i + \mathbf{X_i'}\gamma + u_i. \tag{1.6}$$

Here, the well-behaved structural error is denoted by ϵ_i, whereas the composite residual is $u_i = \delta C_i + \epsilon_i$. Given that C_i has an impact on our explanatory variable of interest, S_i, the latter will be endogenous since $\text{cov}(S_i, u_i) \neq 0$. As a consequence, the estimation of this model by means of regression equation 1.1 will yield a parameter estimate $\widehat{\theta_s}$ that suffers from omitted-variable bias, with $\text{plim}\,\widehat{\theta_s} \neq \beta$. Importantly, $\widehat{\theta_s}$ will absorb any causal impact that C_i may have on H_i. As a result, the presence of common effects could easily lead to erroneous conclusions of active causal links between wealth and health in cases where β actually equals zero.

Causal inference in the face of both challenges

Naturally, there is no reason to believe that both econometric problems are mutually exclusive. As a rule, they will be present at the same time, aggravating causal inference even more. Ultimately, we have to estimate a structural model that takes the following form:

$$H_i = \alpha + \beta S_i + \mathbf{X_i'}\gamma + \underbrace{\delta C_i + \epsilon_i}_{=u_i}, \tag{1.7}$$

$$S_i = a + bH_i + \mathbf{X_i'}\boldsymbol{c} + \underbrace{dC_i + e_i}_{=v_i}, \tag{1.8}$$

with e_i denoting a structural error and v_i representing the composite unobservable. Given this multitude of potential confounders, we truly cannot expect the simple regression function 1.1 to uncover β, the structural parameter of interest. While this assessment is certainly sobering, it also sets a clearly defined bar for any alternative identification strategy: in order to be convincing, it has to live up to the challenges of simultaneity and omitted variables.

1.3 The conventional solution: Instrumental variables

A common way of dealing with the potential endogeneity of SES is the use of IV estimators. The virtue of this approach is that – at least in theory – it

solves both of the above challenges at once. All we need is an instrumental variable, Z_i, with the following three properties:

Exogeneity in the population. Z_i should be as good as randomly assigned and must not suffer from the same problems as the endogenous variable it seeks to "replace". In our case, it should neither be influenced by health status nor the unobserved common causes.

Relevance for S_i. Z_i must be partially correlated with the endogenous variable, once the effect of the other exogenous regressors is controlled. So, loosely speaking, the instrument must have a direct influence on the SES variable under consideration.

Redundancy for H_i. Z_i does not belong in the structural equation 1.7, which means that there is no direct effect of Z_i on the dependent variable, H_i. This exclusion restriction guarantees that any impact the instrument may have on health is exclusively through the indirect route via SES.

With a valid instrument at our disposal, we can consistently estimate β by substituting the endogenous S_i in equation 1.7 by its reduced-form prediction, $\widehat{S}_i(Z_i, \mathbf{X_i})$. Intuitively, instead of exploiting the whole observed variation in SES (which also contains endogenous parts), we concentrate on the fraction that is explained by the exogenous instrument.

Yet, as good as this solution sounds in theory, IV estimation typically causes great headaches in practice. The greatest caveat is that a convincing instrument is usually very hard to find. In their search, social scientists often turn to unexpected real-world events that could be interpreted as natural experiments. Whether these provide valid instrumental variables, however, is not always clear and more often than not a matter of faith. The fact that the exogeneity and redundancy assumptions above are inherently non-testable, makes matters even worse. Furthermore, even the availability of a valid instrument does not solve all of our problems. For one, the use of IV estimators is accompanied by considerable loss of precision since the exogenous variation in SES evoked by the instrument will not be devoid of noise. Moreover, IV estimators are biased in finite samples even if asymptotically consistent. As is well-known, both problems are especially pronounced if the correlation between instrument and endogenous variable is only weak.

Besides these textbook-established weaknesses, the IV approach can additionally be problematic if the instrument of choice is an exogenous wealth shock that occurs late in the life cycle. In fact, it is exactly this sort of natural experiment that is predominantly exploited in the literature: Meer *et al.*

(2003) as well as Michaud and van Soest (2008) use inheritances, arguing that – unless anticipated – they provide a good instrument for wealth changes. In a similar vein, Smith (2005) interprets the strong stock-market surge in the 1990s as a positive wealth shock, and it is probably just a matter of time until we will see the first papers that make use of the exogenous variation in wealth caused by the recent global financial crisis. While all of these instruments may be exogenous and certainly have an impact on wealth, it is not entirely clear if the SES variation they induce is really that relevant for *health*. According to Grossman's (1972) standard economic model of health, an individual's general health status can be viewed as a latent capital stock that reflects the entire history of medically relevant events and behaviors. As a result, the human body will certainly react to current influences but it will not forget how it was treated in the past either. This "memory effect" likely extends to any influence SES may have had during one's lifetime. In light of this, it is questionable whether sudden changes in wealth are really that informative when testing for causal links between SES and health.

As an illustration, consider the somewhat hyperbolic case of an individual without a high-school degree whose low-income occupation did not allow for significant asset accumulation or any of the amenities that usually come with money. Now, in a happy turn of events, this person wins the lottery at the age of 62 – a natural experiment that could arguably be used for an IV approach as well. Which of the two available SES measures will be more relevant for predicting the jackpot winner's health status: the long history of education, income, and wealth (which is not used for identification of β due to the potential endogeneity of these variables) or the one-time wealth shock close to retirement age? Sure enough, the IV estimator would – provided the instrument is truly exogenous and sample size sufficient – avoid the comparison of apples and oranges by effectively matching the health of the above person with that of a non-winner whose pre-shock SES is equally scant. But the central problem remains: while wealth may have had a causal, incremental impact on well-being throughout life, this may no longer be the case at a later stage because one's physical condition is predetermined after years of continuous investment (or the lack thereof) into the health capital stock. Given this inertia, it is unrealistic that our fictitious lottery winner – who is at risk of being in poor health, considering his subpar economic history – will suddenly be regenerated just because her bank balance improved.

Since an IV estimator makes use of this potentially ineffective variation in wealth to identify β, there is a great chance that causal links from SES to

health are statistically rejected, even though they are operating in general.[3] Admittedly, an IV estimator will still capture any *instantaneous* impact a wealth shock would have on health outcomes. As a renewed look at the potential causal pathways for hypothesis A suggests, immediate effects are most likely to arise through channel A.2 if wealth shocks are severe enough to have direct psychological consequences.

1.4 The approach of the HWW study

The previous section demonstrates that the identification of causal paths between health and wealth with IV approaches is not always feasible. Especially the isolation of truly exogenous and yet meaningful variation in SES poses considerable problems. On this account, HWW propose an alternative identification strategy that avoids this critical step altogether. In fact, they make use of the entire observed variation in SES variables, tacitly accepting that some of it may well be of endogenous nature. The authors argue that, in spite of this methodological simplification, their approach still allows for at least indirect inference of causal links from SES to health.[4]

Naturally, HWW need to find convincing answers to the two econometric challenges described in section 1.2.2. When testing hypothesis A, they face challenge 1 of excluding the possibility that any observed comovement of wealth and health is in reality due to reverse causality. In addition, they have to tackle challenge 2 of ruling out that the association is driven by unobserved common effects.

Challenge 1: Ruling out hypothesis B

Distinguishing hypotheses A and B without the aid of instrumental variables is a difficult task. We may observe that the poor are less healthy but we have no information on which happened first: were people already poor before they got sick, or were they already sick before they became poor? With cross-sectional data that only offers a snapshot of this association, there is no way of finding out. Panel data, on the other hand, provides valu-

[3] In this light, it is not too surprising that none of the aforementioned studies using wealth shocks as an instrument for SES, was able to find evidence supportive of hypothesis A.

[4] In their article, HWW also formulate tests on causality working in the opposite direction. However, the authors themselves are quite skeptical about this part of their analysis, admitting that it is likely subject to model misspecification. As they stop short of endorsing their own results, we follow their lead and concentrate on the more promising test of hypothesis A.

able information on transitions in health and wealth, making it possible to analyze the dynamics of their relationship and to identify the direction of the causality flow. Imagine we were to analyze the dependence of health *innovations* on past levels of SES. As long as one agrees that a cause must precede its effect, we can be sure that the (unanticipated) onset of an illness at time t cannot have caused the amount of wealth or education at time $t-1$. Given there is any causation at work, it must flow from the past to the present, or – as in this case – from SES to health innovations.

HWW take this insight to heart by applying their framework to the first three panel waves of the aforementioned AHEAD survey study, which spans the years between 1993 and 1998 and is representative of the US population aged 70 and older. They propose a dynamic model of health incidence that takes the following form:

$$f\left(HI_{it}^{j} | \mathbf{HI}_{it}^{k<j}, \mathbf{H}_{it-1}, \mathbf{S}_{it-1}, \mathbf{X}_{it-1}\right), \tag{1.9}$$

where i once again stands for the unit of observation (in this case: household) and the newly introduced t denotes time. The index j stands for the respective health condition as the authors apply their model to 20 different medical outcomes.[5] The dependent variable, HI_{it}^{j} measures a new incidence of a given health condition.[6] According to this model, a health innovation is potentially influenced by the following explanatory variables:

Past level of SES: The vector $\mathbf{S}_{it-1}$ includes five SES variables, namely wealth, income, years of education, dwelling condition, and neighborhood safety. These are the variables of main interest. Conceptually, if SES had any direct causal impact on health, we would expect to observe that rich individuals are less likely to develop a new medical condition compared with poor individuals. While this finding alone would not yet prove the existence of a causal link from SES to health, confounding with reverse causality could be ruled out since $\mathbf{S}_{it-1}$ precedes HI_{it}^{j}.

[5] These include acute illnesses (cancer, heart disease, stroke), mortality, chronic conditions (lung disease, diabetes, high blood pressure, arthritis), accident-related events (incontinence, severe fall, hip fracture), mental problems (cognitive impairment, psychiatric disease, depression), as well as information on interview status (self vs. by proxy), BMI, smoking behavior, ADL/IADL impairments, and self-rated health.

[6] Note that this measure of health innovation *cannot* be interpreted as a simple change in health status ($\Delta H_{it}^{j} = H_{it}^{j} - H_{it-1}^{j}$) since HI_{it}^{j} generally captures deteriorations in health only. For chronic illnesses, such as diabetes, it measures when the condition was first diagnosed. For acute health events, such as stroke, HI_{it}^{j} indicates every new occurrence.

Past health status. New medical events are likely influenced by the respondent's health history as well. This may take the form of state dependence (e.g., past cancer influences the onset of new cancer) and co-morbidities (e.g., past cancer influences the onset of depression). For this reason, HWW control for vector $\mathbf{H}_{it-1}$, containing the past levels of all 20 health conditions.

Current health incidences with immediate impact. In theory, health innovations could also be influenced by contemporaneous shocks in SES or other health conditions. This constitutes a problem for HWW's concept of dealing with simultaneity as it critically relies on the ability to observe the timing of innovations in both variables. HWW solve this problem by imposing further structure: On the one hand they make the assumption of no instantaneous causation of SES to health shocks, arguing that any causal action as described by channels A.1 to A.5 takes time.[7] On the other hand, they impose a chain structure on contemporaneous health innovations, grouping them in the order in which instantaneous causality is most likely to flow.[8] Thus, they include the vector $\mathbf{HI}_{it}^{k<j}$ containing the incidence variables for all health conditions $(1, ..., k)$ that are causally arranged upstream of condition j.

Demographic control variables. Finally, the authors control for a number of demographic factors that could have an impact on health events, too. The corresponding vector, $\mathbf{X}_{it-1}$, includes the respondent's age, marital status, as well as information on the parent's mortality and age at death.

Building on model 1.9, HWW design a test for non-causality of SES in the spirit of Granger (1969) and Sims (1972). This so-called Granger causality (or G-causality) approach is a purely statistical take on the con-

[7] The authors themselves make the point that this assumption loses its innocuousness if the time intervals between panel waves become too large since even the more inertial causal links will then have enough time to unfold. Given that the AHEAD study is conducted biennially, the time aggregation to observation intervals may indeed reintroduce some degree of simultaneity.

[8] HWW list cancer, heart disease, and stroke first because they can have an immediate impact on mortality. The other medical conditions are grouped such that degenerative illnesses can cause chronic diseases, which in turn may influence accidents and finally mental health. Importantly, instantaneous causality is not designed to flow in the opposite direction.

cept of causation, having its origin in the time-series literature. Formally, SES is not Granger causal for health condition j if

$$f(HI_{it}^{j}|\mathbf{HI}_{it}^{k<j},\mathbf{H}_{it-1},\mathbf{S}_{it-1},\mathbf{X}_{it-1}) = f(HI_{it}^{j}|\mathbf{HI}_{it}^{k<j},\mathbf{H}_{it-1},\mathbf{X}_{it-1}), \qquad (1.10)$$

i.e., HI_{it}^{j} is conditionally independent of $\mathbf{S}_{it-1}$, given $\mathbf{HI}_{it}^{k<j}$,$\mathbf{H}_{it-1}$, and $\mathbf{X}_{it-1}$. Intuitively, given health history, knowledge of SES history must not contribute to the predictability of health innovations. The test is implemented by estimating the model by maximum likelihood (ML) both unconstrained (with $\mathbf{S}_{it-1}$ as regressors) and constrained (without $\mathbf{S}_{it-1}$) and by subsequently comparing the log likelihoods of both versions. The motivation for this likelihood ratio test is that the two values should be the same if the null hypothesis of non-causality is true.

The detection of Granger causality, however, does not guarantee the presence of "true" causality in a structural sense, which is the concept we are ultimately interested in.[9] Admittedly, information on the presence of G-causality is helpful when predicting health innovations for an individual with *given* health and SES history. However, the reduced-form nature of G-causality renders it unsuitable to predict the effects of (economic) policy interventions. If SES is Granger causal for health innovations, we only know that, for instance, the onset of an illness is likelier for a person with low SES. Yet, we do not know if this statistical dependence is due to a real causal link from wealth to health (hypothesis A) or due to unobserved common effects (hypothesis C). Given the diverging policy conclusions both interpretations would trigger, HWW also need to address the second methodological challenge of dealing with the omitted-variable problem.

Challenge 2: Ruling out hypothesis C

Most of the omitted variables identified in section 1.2.1 to potentially have a common influence on health and SES are unobservable by definition. As a result, challenge 2 cannot simply be resolved by improvements in data quality and the addition of missing variables to the vector of covariates.

[9] There are three major "schools" of causal analysis: The structural approach (S-causality) described by Hoover (2001) and Hausman (2003) that is grounded in econometric simultaneous equations models, the potential-outcomes approach (P-causality) characterized by Rubin (1974) and Heckman (2000) that is based on the analysis of experimental treatments and the time-series prediction approach (G-causality) employed here. The conventional interpretation of "true" causality is arguably best described by S- and P-causality treatments. In fact, Pearl (2000) demonstrates a formal equivalence between the two concepts. Both of these schools are critical of G-causality, arguing that its purely positivistic approach does not realistically characterize causal properties.

HWW also refrain from making use of fixed-effects estimation, which represents another common strategy to heal omitted-variable bias in cases where panel data is available. In fact, the efforts made by the authors to distinguish between structural causality and common effects are limited to using a rich set of covariates in the hope that this will mitigate the importance of unobservables. They argue (p. 6) that,

> [f]or example, genetic frailty that is causal to both health problems and low wages, leading to low wealth, may be expressed through a health condition such as diabetes. Then, onset of new health conditions that are also linked to genetic frailty may be only weakly associated with low wealth, once diabetic condition has been entered as a covariate.

Despite this conciliating argument, HWW acknowledge that the failure to cleanly identify causal structures questions their approach's ability to gauge the effects of "out-of-sample" policy changes. To address this issue, they scrutinize the generality of their results by adding invariance tests to the analysis. Intuitively, a model is only suitable for the sort of predictions HWW have in mind if it remains valid under different scenarios than those covered by the data, or – as the authors put it – if it has the invariance property of being valid for each possible history. For instance, if the application of the model to different populations, time periods, and policy regimes had a negligible impact on estimation results, there would be reasonable hope that the Granger non-causality tests are indeed informative. The invariance tests as implemented by HWW mainly inspect the stability of findings across time. Model 1.9 is estimated by stacking the data for the two available panel wave transitions (i.e., W1→W2 and W2→W3) above another. The same model is also estimated for both wave transitions individually, and a test statistic is constructed that compares the log-likelihoods of these three estimations. The motivation for this likelihood ratio test is similar to that of a Chow test. If the null hypothesis of model invariance is true, estimated parameters of the stacked model should not differ from those of the two single-transition models.

All told, HWW apply the following system of non-causality and invariance tests to the estimations of all 20 health conditions: First, they test for Granger non-causality of SES for health innovations in the stacked version of the model under the maintained assumption of invariance (S|I). Then, they employ an unconditional invariance test, as described above (I), followed by an invariance test with non-causality imposed (I|noS). Finally they implement a joint test of invariance and non-causality (S&I). Conceptually, HWW condition the validity of their non-causality tests on the outcome

of the corresponding invariance test: only if invariance is confirmed, they will put faith in the model's results. The authors are optimistic that with these refinements in place, their model is well-placed to make meaningful predictions even if it fails to identify true causal links, stating (p. 10) that

> [i]t is unnecessary for this policy purpose to answer the question of whether the analysis has uncovered a causal structure in any deeper sense. Econometric analysis is better matched to the modest task of testing invariance and non-causality in limited domains than to the grander enterprise of discovering universal causal laws. However, our emphasis on invariance properties of the model, and on tests for Granger causality within invariant families, is consistent with the view of philosophers of science that causality is embedded in "laws" whose validity as a description of the true data generation process is characterized by their invariance properties.

They even go a step further and suggest that their approach – while not powerful enough to distinguish between causation and common effects – permits at least the one-sided test for the *absence* of true causal links. Essentially, they view Granger causality as a necessary but insufficient condition for a structural causal pathway from SES to health. Their decision criteria when interpreting results are as follows: If the invariance test fails, one should question the validity of the model for this particular health variable and refrain from drawing any conclusions. If invariance holds and Granger causality is present, one cannot distinguish between a direct causal link and a common factor. Yet, if invariance holds and Granger causality is ruled out, it should be safe to deduce that SES does not have a causal impact on the health condition under consideration.

Summary of HWW's findings

Contrary to conventional wisdom, the evidence from applying HWW's approach to the elderly US population is not *universally* supportive of hypothesis A. In fact, they find that SES is unlikely to be causal for mortality, most acute health conditions, accidents, and a large number of degenerative diseases. Medical conditions, for which direct causal links cannot be ruled out, include self-rated health status, most mental illnesses and some chronic conditions such as diabetes, lung disease and arthritis. This pattern loses some of its mysteriousness when viewed in the context of US health-policy characteristics. The population under examination is of advanced age and eligible for Medicare, which will likely weaken any causal impact wealth could have on well-being via the affordability of health care. Yet, even Medicare coverage is not fully comprehensive and tends to focus on

acute care procedures, while generally failing to limit out-of-pocket costs for treatments of chronic and psychological conditions.[10] This lends indirect evidence for the importance of channel A.1 since the socio-economic gradient emerges exactly for those health conditions, for which the ability to pay is most likely to be an issue.

Reflecting the substantial degree of ambiguity in these results, the policy conclusions formulated by HWW are rather contained in both phrasing and substance. On the one hand, they cannot overcome the methodological challenge of inferring true causality when G-causality is detected. This leaves open whether SES-linked preventive care induces onset of chronic and mental illnesses or whether persistent unobserved factors are to blame for the observed health-wealth association. On the other hand, even convincing evidence for the absence of direct causal links might not necessarily warrant the bluntest form of policy recommendation. Sure enough, SES-linked therapies for acute diseases do not appear to induce health and mortality differentials, which – to quote HWW (p. 10) – should *theoretically* permit the strong conclusion that

> policy interventions in the Medicare system to increase access or reduce out-of-pocket medical expenses will not alter the conditional probabilities of new health events[.]

However, the authors stop short of actually drawing this conclusion, which reflects their reluctance to base overly aggressive policy proposals on a concept whose ability to simulate the effect of system shocks is not indisputible.

Discussion of HWW's approach

All things considered, what should we make of HWW's approach of inferring causality and yet avoiding the cumbersome search for exogenous variation in SES? Does their reliance on Granger causality and their decision to focus on health innovations really do the trick of solving the endogeneity problem, or have they entered a methodological dead-end street? Overall, the response within the literature has been fairly critical, albeit not excoriating, pointing out a number of issues briefly discussed below.

Existence vs. activation of channels It is important to understand the limitations of an approach that focusses on *innovations* in health, rather than health status itself. HWW detect a strong and ubiquitous association of SES and prevalence of health conditions in the initial

[10] Note that the study was conducted well before the introduction of Medicare Part D in 2006 that especially benefited the chronically ill by improving the coverage of prescription drugs.

wave of their sample. This suggests that the elderly population under consideration has potentially been affected by some of the causal channels between health and wealth in the past. This history, however, remains a blind spot for HWW's model: by concentrating on future health events, they are unable to explain what factors lead to the pre-existing SES gradient. By contrast, they study the question whether SES has an impact on the onset of *additional* medical conditions, given an individual is already old, still alive, and has gone through a long and unexplained health-wealth history. While the analysis of an elderly population is not illegitimate and certainly interesting in its own right, one should entertain some doubts about its external validity. In theory, HWW's findings could – if extrapolated backwards – also provide a retrospective explanation for the early relation between SES and health. However, as pointed out by Adda *et al.* (2003), Heckman (2003), Poterba (2003), and HWW themselves, this extreme form of time invariance over the entire life cycle is unlikely to hold as certain causal channels are probably relevant at different stages in one's life.[11] In light of this, an accepted non-causality test should perhaps not be taken as evidence against the plain existence of a causal link but rather against its *activation* within the class of invariances under consideration.

Unobserved common effects. As argued above, one weakness of HWW's approach is that it cannot separate true causality from hidden common effects. Yet, according to the authors, this will only constitute a problem if Granger causality is detected. In the absence of G-causality, causation in a structural sense should be ruled out as well. This interpretation implies that the detection of conditional dependence is a prerequisite for an active causal link – an assumption that is questioned by Heckman (2003), who argues that persistent hidden factors may also work in the opposite direction of causal pathways and offset them. If this were the case, information on G-causality might actually not tell us anything about true causal mechanisms, rendering HWW's strategy ineffective. However, the likelihood of direct causal effects being *exactly* offset by unobserved common factors should be practically zero, making this argument irrelevant for identification. Then

[11] For retirees, pension income is not affected by (contemporary) ability to work, occupational hazards vanished on the day of retirement, and Medicare provides basic health insurance, rendering channels B.1, A.3, and A.1, respectively, of little importance when late in the life cycle. At younger ages, however, all of these pathways may well have played an important role.

again, there are obvious limits to this defence in finite samples, so that statistical inference of causation could indeed be seriously jeopardized by the failure to account for hidden common causes.

Invariance tests. Anticipating that their framework might fall short of inferring deep causal structures, HWW subject their model to the aforementioned invariance tests. On a conceptual level, model invariance would arguably justify predictions of policy effects but there are legitimate concerns whether the actual tests implemented in their paper are statistically powerful enough. Granger (2003), Hausman (2003), Heckman (2003) and, once more, HWW themselves point out that invariance under historical interventions is of little use when the panel is as short as AHEAD, offering hardly any in-sample variation in populations, age structures, and – most importantly – policy regimes. As a consequence, an accepted invariance test as implemented by HWW is unlikely to be a sufficient condition for the sort of model validity necessary to make out-of-sample predictions. On top of that, Poterba (2003) even questions whether one should view the acceptance of HWW's invariance tests as a *necessary* condition for meaningful analysis. Instead of discarding results when invariance tests are rejected, one could follow up on the reasons for time invariance failures as they may be informative of structural breaks in causal relationships. For instance, certain causal pathways may switch on or off in the course of policy changes or as the observed cohort grows older. In such cases, failed invariance tests would actually shed light on the circumstances under which causal links will be active or unexpressed, allowing for sharper, channel-specific causality tests.

Health dynamics. Another reason for concern is the fact that HWW model health dynamics as a first-order Markov process, which cannot be expected to properly capture the medium and long-run evolution of health. Intuitively, this is because the Markov model assumes that all relevant information about the whole past is captured in the observed variables one period ago. This is unrealistic since knowledge of longer histories would better capture the stock characteristics of health capital as envisioned by Grossman (1972). Taking functional limitations as an example, a respondent who reported difficulties with walking one year ago and no limitations previously has a different outlook than a respondent who consistently reported difficulties with walking for the last ten years.

Instantaneous causality. Finally, Florens (2003), Geweke (2003), and Heckman (2003) express their skepticism about HWW's handling of instantaneous causality. The hierarchy imposed on health conditions (with the assumption that incidence of each condition is conditioned on upstream incidences but not on downstream ones) may be acceptable as a reduced-form assumption and is etiologically fairly reasonable. Yet, it likely falls short of the structural stability explored by invariance tests and is a potential source of serious model misspecification, making it a prime target for methodological improvements in the course of future research.

1.5 Reanalysis of HWW with new data

The preceding discussion indicates that HWW's approach of disentangling the association between health and wealth while avoiding the often futile struggle of finding exogenous variation in SES comes at the price of limited methodological persuasiveness. However, since the generic alternative – instrumental variables – is not exempt from substantial criticism either, we certainly feel that this identification concept merits methodological refinement rather than being dismissed altogether. Some weaknesses, such as the treatment of common effects, health dynamics, or instantaneous causality, require significant modifications to the original model and we plan to implement these in future research.

Yet, one of the major downsides of HWW's study – the lack of invariance test power – can be addressed without the need for complex changes but instead by applying the largely unaltered model to a more apposite set of data. Recall that the root of this problem is that the invariance tests are based on rather limited variation in "histories" of states relative to the universe of potential histories. Increasing the N as well as the T dimension of the panel data will arguably raise the number of histories and enhance the power of these tests. Of course, we can also expect larger sample sizes to boost the statistical power of non-causality tests, effectively reducing the risk of committing type-II errors. But sample size is not everything. We believe that the analysis will also greatly benefit from larger sample "diversity", with data covering different kinds of populations that are subject to varying institutional setups. For instance, the inclusion of younger respondents could shed light on the question if the activation of causal links is stable throughout the life cycle or if reaching the retirement age induces some sort of structural break.

Given that the HRS survey study provides panel data that meets all of the above requirements, the present analysis keeps methodological changes to an absolute minimum and assesses the stability of HWW's results when applying their model to new and more encompassing data.[12] Of particular interest is the question whether HWW's somewhat surprising result of SES not having any direct causal impact on most health conditions is confirmed as test power increases.

1.5.1 The HRS panel data

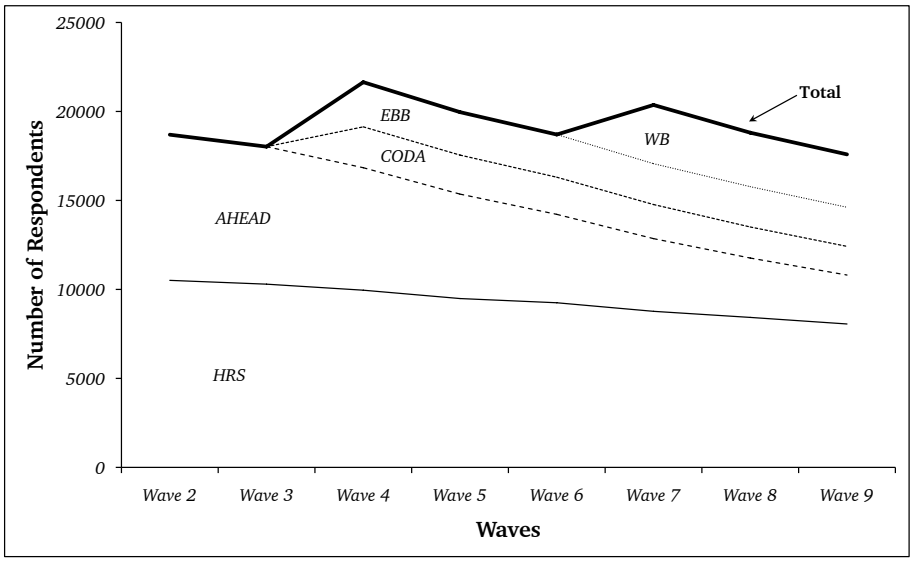
Sample characteristics

Our data – which is representative of the non-institutionalized US population over the age of 50 – comes from the *Health and Retirement Study (HRS)*, a large-scale longitudinal survey project that studies the labor force participation and health transitions of individuals toward the end of their work lives and in the years that follow. While the data is collected by the *University of Michigan Survey Research Center for the National Institute of Aging*, we use the public-release file from the *RAND Corporation* that merged records from the nine panel waves available to date. The wave 1 interviews were conducted in 1992 and then repeated every two years, so that HRS incorporates data from 1992–2008. Due to significant changes to the survey design between waves 1 and 2, the first cross-section cannot be directly compared to subsequent observations and is therefore not used in our analysis. To ensure that HRS stays representative of the population as time goes by, the panel is periodically refreshed with new cohorts of respondents. Up to now, the sample consists of five different entry cohorts: the original 1992 "HRS" cohort (born 1931–1941), the 1993 "AHEAD" cohort (born 1923 or earlier), the "CODA" (born 1924–1930) and "WB" (born 1942–1947) cohorts entering in 1998, and the "EBB" cohort (born 1948-1953) added in 2004.

At baseline in wave 2 (covering interviews conducted between 1993 to 1994), the dataset contains 18,694 individuals with usable records. The panel is subject to considerable attrition, which reduces sample sizes from wave to wave – a trend that is only temporarily disrupted when a refreshment cohort is added to the sample (see figure 1.1). The two sources for at-

[12] In fact, this study exactly replicates HWW's model of health incidence with one notable exception. For simplicity, we skip their treatment of interview delay, which accounts for the fact that interview timing appears to depend on health status. While this potentially calls into question the comparability of responses from healthy and severely ill individuals, we find that results are virtually unaffected by this non-random distribution of time at risk.

Figure 1.1. *The five HRS cohorts*
Sample sizes per wave



Notes: Abbreviations are as follows: AHEAD=Asset and Health Dynamics among the Oldest Old; CODA=Children of Depression; EBB=Early Baby Boomers; WB=War Babies.

trition are mortality (especially for the elderly AHEAD cohort) and "sample fatigue". Death-related attritors are kept in the working sample since mortality is one of the key outcomes of interest. With respect to all others attritors, we apply two alternative sampling schemes. The first exactly mirrors HWW's benchmark in that it categorically excludes non-respondents from the working sample, irrespective of when their drop-off occurs or whether they rejoin the survey in later waves. As detailed below, the second sampling procedure assures that the information of these households is used for as long as they are part of the sample.

Much like in HWW's original study, we exclude all individuals with missing information on critical variables. This includes item nonresponse for key demographic variables as well as cases where information on health conditions is generally unavailable. If respondents merely fail to answer isolated health queries, these gaps are filled by means of simulation-based imputation. Certain health questions on cognitive ability, severe falls, and hip fractures are not asked to participants below the age of 65, which is why these variables are excluded from all estimations that include younger subpopulations. While HWW went to great lengths to impute a large number of wealth and income observations with first-order Markov cross-wave hot deck imputation methods, we are in the more convenient position to rely on the imputations that are now readily available within the RAND/HRS data.

We should note that, in spite of this data cleaning, self-reported wealth and income measures are still suspect of considerable measurement error. Summary statistics for all variables used in our analysis are given in appendix table 1.6.

Comparison with HWW's data benchmark

HWW's original data sample consists of the AHEAD cohort of US Americans aged 70 and older who are tracked through panel waves 2 to 4. Using the HRS data that is available to date, allows for deviations from this benchmark along several dimensions. Naturally, we can follow the same individuals for *more time periods* since the AHEAD cohort is now biennially observed between 1993/94 and 2008. Given the introduction of the four additional entry cohorts, the analysis can also be extended to *different individuals* with potentially diverging histories compared to those in the original study. In addition, it is now possible and certainly interesting to also widen the working sample by incorporating *younger individuals,* aged 50 and older. Finally, it should be noted that there is an additional, albeit minor, deviation from HWW's data benchmark even for the same observations as in the original study. One reason for this is that the early AHEAD data has subsequently been subject to data updates and revisions within the HRS project. Similarly, there may be differences between the SES imputations carried out by HWW and those conducted by RAND/HRS.

1.5.2 Results

Following the strategy described in section 1.4, we fit model 1.9 as binomial probits except for BMI and ADL/IADL impairments, which are estimated with OLS and ordered probit, respectively. Appendix tables 1.7 and 1.8 contain the empirical significance values for the system of non-causality and invariance tests specified above. For a more concise overview of results, refer to tables 1.2 and 1.3. In a nutshell, the reanalysis with fresher and more encompassing data suggests that direct causal links from SES to health can be ruled out for much fewer health conditions than in the original study. This casts some doubt on the stability of HWW's findings. In order to understand which of the data changes contribute to these deviating conclusions, we estimate the model multiple times, using several different datasets by augmenting them stepwise along the dimensions outlined above. In the first step, we rerun HWW's benchmark study for the same cohort and time

periods, yet reverting to the current version of HRS data instead.[13] This will detect any impact arising from data revisions and differences in imputations. The second step consists of extending the analysis to the other three cohorts in the sample, hence testing whether HWW's conclusions are also valid for different individuals. The third step addresses the question of how results are affected by increasing the number of time periods under consideration. Since in HWW's model there is no self-evident way to aggregate the information from multiple time intervals, we compare two different sampling approaches: one that refills the sample after each wave with applicable observations and one that does not. In the fourth and final step, we evaluate the impact on estimation results when younger individuals are included in the analysis as well. This stepwise decomposition of all data and sampling changes appears to be more informative than applying the whole bundle of modifications at once.

Step one: Re-estimating HWW with revised data

In order to gauge the result's sensitivity to data revisions and imputations, the model is re-estimated with fresh HRS data for the exact same cohort (AHEAD) and time periods (waves 2 to 4) as in the original study. The differences between the new results and HWW's benchmark are quite modest, and outcomes of causality test are mostly unchanged. The most notable exception is diabetes for which the non-causality test had to be previously rejected among male respondents. With revised data, however, a direct causal link from SES to diabetes seems unlikely to exist. For further details, compare columns (B) and (C) of table 1.7.

Step two: Adding new cohorts

While the relative stability of results in face of data revisions is certainly encouraging, a much stricter test is posed by extending the analysis to all available cohorts. To achieve this, we run three separate estimations on

[13] To ensure that none of the observed changes is confoundedly rooted in the way certain variables are constructed and program codes are implemented, we also reran HWW's study verbatim using their original data. While the goal to exactly reproduce HWW was ultimately achieved, it should be noted that results are identical to those published as log files within the appendix of the original 2003 paper, but not to those in the article itself. This difference is attributable to data revisions that HWW accounted for shortly after the paper was published, which means that the outcome from the appendix should be preferred as the ultimate benchmark. As is evident from comparing columns (A) and (B) of table 1.7, said differences are not always trivial in size. Most strikingly, invariance tests tend to fail less frequently when applied to HWW's post-publication dataset. Yet, the impact on causality tests is negligible, thus not challenging the author's main conclusions from the article.

the following samples: First, we revisit waves 2 to 4 but allow members of cohorts other than AHEAD to be part of the working sample. This barely changes the sample composition because the only other cohort that is part of the survey in this early stage is HRS, which hardly contains any individuals aged 70+. For the other two estimations, HWW's data benchmark is additionally changed inasmuch as later waves are used. The second estimation starts at wave 4, when the new cohorts CODA and EB are interviewed for the first time. Note that we do not restrict analysis to these two cohorts. Rather, all respondents who are at least 70 years old at wave 4 and who are not subject to subsequent sample attrition are followed until wave 7. This closely mirrors HWW's approach of analyzing a three-period panel, hence still keeping the deviations from the benchmark to a minimum. The third estimation repeats the second for waves 7 to 9, coinciding with the entry of the most recent cohort, namely WB.

Not surprisingly, the first estimation (table 1.7, column (D)) yields results that are almost identical to those of HWW's benchmark with revised data (table 1.7, column (C)). The hypothesis tests associated with the other two estimations, however, prove to be rather different. As far as non-causality tests are concerned, the differences seem to be unsystematic. For some medical conditions, such as depression and ADL impairments for females and incontinence for males, causality from SES can no longer be ruled out as non-causality tests are now consistently rejected. For other health conditions, namely diabetes and lung disease for males as well as psychiatric disease for females, the opposite holds true, as non-causality tests can no longer be rejected. In addition, there are a number of diseases for which the benchmark causality test results are not confirmed for only one of the sub-samples. For further details compare columns (D), (E), and (F), respectively with column (C) of table 1.7. Invariance tests, on the other hand, tend to be accepted more often than those under the benchmark scenario. At first glance, this may seem contradictory since causality tests have yielded fairly different results depending on which panel waves are under consideration. One should, however, not forget that the invariance tests merely check whether the model is time invariant within each of the three estimations but not among them. This is changed in the third step when the information from more than just three waves is incorporated.

Step three: Increasing the number of time periods

Step two has indicated that results depend on which panel waves are chosen to form the working sample. In order to reduce this arbitrary element

Table 1.2. *Results for steps 2 to 4*
Tests for Granger non-causality

Health condition	**Test results**											
	HWW (70+)		Step 2 (70+)						Step 3 (70+)		Step 4 (50+)	
	W234		W234		W456		W789		W2-9		W2-9	
	F	M	F	M	F	M	F	M	F	M	F	M
Cancer									•	•		
Heart							•		•		••	••
Stroke					•						••	
Mortality									•	••	•••	•••
Lung		•		•								••
Diabetes		•									•••	
High bp.					•						•	
Arthritis							••		•••	••	•	
Incontinence					•		•	•	•••	••	•••	•••
Fall									•	•	n.a.	n.a.
Hip fract.						••					n.a.	n.a.
Proxy		•		•••		••			••	•••	•••	•••
Cognition	••	•	••		•	•	••	•	•••	•••	n.a.	n.a.
Psychiatric	••		•					•••		•	•••	•
Depression	•			••	••	••	•••	•	•••	•••	•••	•••
BMI									•		••	••
Smoke now					•						•••	•••
ADL	•		•						•••		•••	•••
IADL		•••		••				•••		•••	•••	•••
S.-r. health	••	•	•••	•	••	••	•••	•••	•••	•••	•••	•••

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (•), rejected at 1% level (••), or rejected at 0.1% level (•••). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level. Blank cells indicate that Granger non-causality cannot be rejected. A "n.a." entry reflects the fact that information on certain health conditions is not available for the pre-Medicare population.

and to maximize the use of available information in the data, it makes sense to increase the number of panel waves. Since there is no unequivocal way to implement this in practice, we propose two different sampling approaches. The first approach is a simple extrapolation of HWW's sampling method. The working sample consists of all individuals who participated in the survey in wave 2 and who were not subject to sample-fatigue-related attrition in later waves. This cohort is then followed for as many waves as possible. This sampling scheme has two major disadvantages. First, by restricting the sample to individuals who were part of the survey from the very start, we exclude refreshment cohorts CODA, EB, and WB, basically discarding useful information. The second drawback is of a more practical nature: death-related attritors cause the sample to dramatically thin out over time, so that sample sizes eventually become too small to conduct

any meaningful analysis. Moreover, as time moves on, the sample arguably becomes less representative of the true population because the ongoing attrition will select against the most frail. Nevertheless, and for the sake of maximum comparability with HWW, we estimate two versions of this first approach. One that follows individuals from wave 2 until wave 6 (covering cohorts HRS and AHEAD) and another that follows individuals from wave 4 until wave 9 (covering HRS, AHEAD, CODA, and EB). The number of waves is chosen so that sample sizes in the last respective wave are still reasonably large.

The alternative sampling scheme directly addresses the downsides of the approach above. Instead of limiting the sample to respondents who are part of the survey from the very beginning, it is now refilled in each wave with all available respondents who meet the respective age criterion (i.e. 70+) and who answer all relevant questions for two consecutive waves. That way, all cohorts are used for analysis, sample sizes never diminish to levels too low for efficient estimation, and, consequently, all 8 waves can be used simultaneously. As a positive side effect, attrition bias is reduced as well, as the mortality-induced loss of observations is offset by filling up the sample with new respondents, once they become age-eligible. One might object that this approach reduces the power of panel analysis as it does not make much use of its potential time-series length. For the purpose of reproducing HWW, however, we deem it suitable since the original model does not use the theoretical length of the panel either, assuming that health and wealth trajectories are sufficiently described by single lags. Given that the models are estimated by simply stacking the data of all two-waves transitions above another, it is irrelevant how long an individual is part of the survey. The information conveyed in the responses of a person who only participates in, say, waves 4 and 5 is no less valuable than that of a respondent who participates from the very beginning to the end, and should therefore not be excluded.[14]

It is noteworthy that the invariance tests of both sampling schemes have slightly different interpretations. In both cases, they test whether parameter estimates stay constant over time, by comparing the log likelihood when all single wave transitions are pooled together with those when estimated separately. For a sample with refilling, an accepted invariance test indicates that the uncovered (non-)causal relationships hold for different populations at different times, underlining the generality of results. Invariance

[14]Of course, the validity of this argument relies on HWW's conceptualization of health trajectories as a first-order Markov process. The future development of more realistic models of health dynamics will require a more sophisticated sampling procedure as well.

tests for samples without refilling, however, cannot answer the question whether causal links hold for different populations since only one cohort is being followed. They rather check whether these links remain constant as a steadily diminishing cohort becomes older over time, ultimately comparing the frail (who exit the sample early) with the medically more robust.

The main change in results from using a larger number of sample waves is that causality from SES to health can no longer be ruled out for a large array of conditions, even for an elderly population, aged 70 and older. This observation holds, no matter which of the two above approaches is used, even though there are some differences. As columns (G), (H), and (J) of table 1.7 reveal, there are seven health conditions for which the samples without refilling yield a rejection of non-causality tests, even though this was not the case for shorter samples. This number even increases to ten conditions, if the sample with refilling is used instead. The most unambiguous evidence exists for six conditions (mortality and falls for males, proxy and BMI for females, and cancer irrespective of gender) for which both approaches suggest that, contrary to earlier evidence, causality may well play a role. The reversed case of causality becoming less likely to exist as panel length grows, is as good as non-existent. The influence of analyzing more time periods at once on invariance tests is about the same for both approaches and not very strong. If anything, invariance failures tend to be somewhat likelier – a result that makes sense because it is more demanding for a model to be valid for eight waves than a mere three.

Step four: Adding younger individuals

So far, the consequence of applying HWW's model to data that includes more individuals and time periods, is that the number of medical conditions for which SES causality may play a role has considerably increased. However, for a population aged 70 and older there remains a large number of diseases for which causal links are not detected, despite the fact that high SES is associated with a lower prevalence of these conditions. While this cross-sectional correlation cannot be interpreted causally, it indicates that causal channels may have been at work earlier in life, before the individual even entered the sample. In light of that, it is interesting to also include younger individuals, to test if the data will pick up additional causal links that are already mute in an elderly population.

First, the sample is opened up to people who are at least 65 years old, so that it represents (with some exceptions) the whole Medicare-eligible subpopulation. This yields a net-increase of 3 to 6 health conditions (de-

Table 1.3. *Pre- vs. post-retirement population*
Tests for Granger non-causality

Health condition	Test results			
	Step 4 (0-64)		Step 4 (65+)	
	W2-9		W2-9	
	F	M	F	M
Cancer				
Heart			●	
Stroke			●	
Mortality		●●	●	●●
Lung				●
Diabetes	●●●		●●	●
High bp.			●	
Arthritis			●●●	●●●
Incontinence	●●●		●●●	●●●
Fall	n.a.	n.a.	●	●
Hip fract.	n.a.	n.a.		
Proxy		●●●	●●●	●●●
Cognition	n.a.	n.a.	●●●	●●●
Psychiatric	●●●			●
Depression	●●●	●●●	●●●	●●●
BMI	●		●●	
Smoke now	●●●	●●●		
ADL	●●●	●●●	●●●	●●●
IADL	●●●	●●●	●	●●●
S.-r. health	●●●	●●●	●●●	●●●

Notes: The same abbreviations as in table 1.2 apply.

pending on whether samples with or without refilling are used) for which causality can no longer be rejected, affirming the speculation above. See columns (B), (F) and (K) of table 1.8 for details. A similar effect can be observed when the sample is opened up even further to include individuals aged 50+, exploiting the entire age range available within HRS. This time, the net-increase amounts to another 3 to 9 conditions, rendering cases for which causal links can be ruled out the exception rather than the rule. Among the latter are illnesses such as strokes and high blood pressure for males, lung disease for females, and cancer for both men and women. For all other health conditions the existence of causal links cannot be refuted. For further details, consider columns (C), (G) and (L) of table 1.8.

It is also worthwhile to split up the sample into older (65+) and younger (50–64) individuals to study how the activation of causal channels differs between a mostly retired, Medicare-eligible population and people who are typically still on the labor market and not quasi-universally health insured. As table 1.3 shows, there is quite a number of medical conditions for which SES may be a causal driving force irrespective of age. These include de-

pression fpr both genders, IADL impairments, incontinence, and diabetes for women as well as ADL impairments for men. For other conditions like arthritis, heart disease, strokes for females, or incontinence for males, SES is only a good predictor of new medical incidences at a higher age. On the other hand, smoking behavior as well as psychiatric problems for women are among the conditions for which a causal link may only be active at a pre-retirement age. Intriguingly, when young and old people are studied separately, results appear to be sensitive to whether samples with or without refilling are chosen (see table 1.8). For older individuals, the sample with refilling suggests more cases of Granger causality than its counterparts without refreshment do. The exact opposite, however, is true for younger individuals as for these, rejected causality is a less frequent outcome in samples without refilling. The latter observation is likely an artifactual side effect of the way the sampling methods are defined: Sampling with refilling effectively excludes people from the 50–64 sample once they become older than 65, whereas sampling without refilling follows all individuals until they die, even if they grow much older. As a consequence, unrefilled "young" samples may arguably pick up some of the causal effects that are exclusively active for the older subjects of the cohort.

Model invariance is not systematically influenced by adding younger individuals to the dataset. The fact that the seeming structural breaks in the relation of SES and health as people grow older are not detected by HWW's invariance tests, should, however, not be surprising. Recall that the test design does not pit the young against the old but the past against the future. The idea is to check parameter invariance as time progresses. Since the age structure within the sample varies only little from wave to wave (especially when it is regularly refreshed), the invariance test will not permit a direct comparison of, say, pre- and post-retirement populations. In light of this, the results in table 1.8 merely suggest that the *time* stability of the model is rather insensitive to changes in the age composition of the sample.

Changes in results of the underlying prediction models

Given the strong dependence of non-causality test results on both the size and age coverage of the estimation sample, it seems natural to investigate how these changes are related to the size and the precision of coefficients of the underlying prediction models.[15] As table 1.4 exemplifies, precision

[15] HWW did not report the coefficients of the prediction models, but these estimates are also available in their online appendix.

of SES coefficient estimates does generally increase with the size of the respective sample, even though this relation is not perfect. While standard errors remain fairly constant across estimations based on similarly sized 3-waves samples (step 2), they surprisingly spike upwards once all waves are pooled together in step 3. This observation may well be rooted in the aforementioned switch from a sampling procedure without (steps 1 and 2) to one with refilling (steps 3 and 4). Precision follows a more predictable pattern within step 4, as standard errors are invariably smaller, the larger the respective sample (note that $N_{50+} > N_{65+} > N_{70+} > N_{0-64}$). The same pattern emerges when comparing results by gender, as the number of women exceeds that of men in each of the subsamples.

Despite increased precision, table 1.5 suggests that the number of statistically significant SES coefficients does not seem to be systematically affected as samples become more encompassing. In HWW's benchmark sample, there is a total of 29 SES regression coefficients that are significant at the 5% level. This number stays rather constant across the performed steps. Yet, note that in the benchmark case there is a fair amount of cases for which coefficients have an unintuitive sign, suggesting that respondents with high SES are likelier to develop a new health condition. The share of these cases stays rather high for all 3-waves samples and drops significantly once all waves are used at once. This means that, while the overall number of significant SES coefficients does not increase, the direction of effects is now more in line with theory. This is an additional insight since non-causality tests as implemented here do merely check whether health innovations are conditionally independent of SES variables, whereas the quality of this dependence is not under consideration.

Table 1.4. *Prediction model*
Average standard errors of SES coefficients

Sex	**Standard errors of SES coefficients (average of all health cond. and SES regressors)**							
	HWW	Step 2 (70+)			Step 3 (70+)	Step 4 (W2-9)		
	W234	W234	W456	W789	W2-9	65+	50+	0-64
Female	0.081	0.083	0.087	0.091	0.126	0.107	0.085	0.172
Male	0.108	0.105	0.111	0.110	0.167	0.135	0.098	0.185

Notes: Reported are average standard errors of SES coefficients obtained from estimating model 1.9. Each entry is an average of 160 single standard errors (20 health variables and 8 SES regressors). Individual standard errors follow the depicted pattern quite uniformly.

Table 1.5. *Prediction model*
Significant SES coefficients

Health condition	SES regression coefficients that are significant at 5% level											
	HWW (70+)		Step 2 (70+)						Step 3 (70+)		Step 4 (50+)	
	W234		W234		W456		W789		W2-9		W2-9	
	F	M	F	M	F	M	F	M	F	M	F	M
Cancer		*E*		*E*			*H*		*W*			
Heart	**I**		**I**	*I*			**E**			**NH**		**N**
Stroke	*I*		*E*		*E*		**W**	*H*		**I**	**W**	
Mortality	*E*		*E*							**W**		**W**
Lung	**W**	**W**	**W**	**W**	**H**		*E*		**WH**		**W**	
Diabetes		**E**		**E**	**W**			**W**	**E**		**WE**	
High bp.	**E**			**E**	**W**			**E**				
Arthritis	**E**		**E**				**I**		**W**		*E*	
Incontinence	**N**				*E*	**N**	*I*	*E*	*I*		*W*	*H*
Fall							*E*	*E*			n.a.	n.a.
Hip fract.	**W**	**I**	**W**	**I**		**W**					n.a.	n.a.
Proxy	*I*	**E**	*I*	**E**		**E**	**W**		*I*		**E**	
Cognition	*W***IN**	**I**	*W***IN***I*		**E**		**E**		**E**		n.a.	n.a.
Psychiatric	**I**	**W**	**I**	**W**								
Depression		**N**		**N**	**E**	**IH**	**EH**		**E**	**E**	**WE**	
BMI					**H**		*I***H**		**E**	**E**		**H**
Smoke now		**W**		**W**	**W**	*N*	*W*		**W**	**E**	**W**	**I**
ADL					*E*	*E*	**I***E*				**WI**	**W**
IADL		**WI**		**WIH**	**I***E*	*N*		**E**	**H**	**W**	*N*	**W**
S.-r. health	**EH**	**WE**	**EH**	**WE**	**N**	**EH**	**INH**	**IEH**	**IEH**	*N***H**	**WIENH**	**IEH**

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Significant coefficients at 5% level for wealth (W), income (I), education (E), neighborhood (N), or housing quality (H). Boldfaced (e.g. **W**) symbols indicate that signs are in line with theory. Italic (e.g. *W*) symbols denote the opposite case. Empty cells indicate that none of the SES variables have a statistically impact on the respective health outcome. A "n.a." entry reflects the fact that information on certain health conditions is not available for the pre-Medicare population.

1.6 Conclusion and future research

All in all, re-estimating HWW's model of health incidence with new HRS data, alters conclusions about SES causation quite significantly. While the impact of data revisions within HRS is encouragingly small, the addition of new cohorts shows that causal inference critically depends on which time periods are used for estimation. Using the information of many – ideally all – waves at once has the greatest effect on results, with many health conditions moving to the column of illnesses for which SES causality may well play a role. Adding younger individuals to the sample has a very similar effect, reducing the number of medical conditions for which the existence of causal links can be statistically rejected even further. As a consequence,

the only health conditions for which SES causation can be ruled out when estimation is based on the most encompassing dataset with refilling, are cancer (irrespective of gender), lung disease for females and high blood pressure for males. For all other health incidences, SES is either G-causal or the failure of invariance tests does not permit reliable conclusions. This represents a stark contrast to HWW's original findings, where the rejection of structural causality was the most frequent outcome.

Given that the greatest changes are triggered by the addition of panel waves (step 3), the main driving force behind this reversal in results is most likely an increase in test power as sample sizes soar. After all, in HWW's stacking model, a longer panel is equivalent to a larger sample (with respect to N) since all waves are pooled together and treated as if they formed one cross section. This interpretation is corroborated by the fact that test results from long panels do not always reflect the average outcome of the respective three-wave panels they consist of. As the example of cancer in table 1.7 illustrates, non-causality tests are often rejected in the long samples, even though they are consistently accepted in each of the short panels. Similar observations can be made when comparing test results by age group. In some cases, a non-causality test is only rejected for the largest, most-encompassing sample of all individuals aged 50 and older. However, in all smaller sub-samples (50–64, 65+, and 70+) the same null hypothesis cannot be rejected. As an example for the latter case, see heart disease for females in table 1.8. All of this evidence permits the emergence of a clear picture: the larger the sample under consideration, the likelier the rejection of non-causality.

We also find that causal inference depends on the age structure of the underlying population, with certain conditions being Granger caused by SES at younger or older ages only. This yields at least indirect evidence that the activation of causal links may indeed change over the life cycle. However, we recommend to take these results with a grain of salt since their lack of robustness is far from comforting, as evidenced by the sensitivity to the choice of sampling schemes. In addition, we should note that the dataset for the 65+ population is about three times as large as that of the pre-retirement group. As a consequence, we face the risk of confounding the true effect of age structures with the impact of varying sample sizes identified above. This may well provide an alternative explanation for the failure to detect many cases of G-causality among the 50 to 64 year olds if estimation is based on a refilling sample – a result that is not confirmed if samples without refilling are used instead.

From a methodological point of view, the results of this study pose bad news for a model whose identification strategy relies on Granger causality. Recall that the reduced-form nature of G-causality cannot discriminate between structural causation and ecological association due to common unobserved effects if G-causality is detected. Ultimately, HWW's framework allows only the one-sided test for the absence of direct causal links, which is confirmed if G-causality is rejected as well. While HWW's original dataset provided us with a large number of such cases, the more-encompassing data samples analyzed here, do not do us this favor. As a result, we find ourselves in the unfortunate situation that little can be learned about the true links between SES and health, making it impossible to draw meaningful policy conclusions.[16]

In light of this, the need to improve the empirical model within future research so as to account for the confounding influence of hidden common factors, becomes even more pressing. In our view, there are two alternative ways to achieve this. The first mirrors the identification strategy of IV approaches: instead of using endogeneity-stricken SES histories as regressors, one could concentrate on the impact of clearly exogenous *changes* in these variables. If these SES innovations meet the critical IV assumptions described in section 1.3, we would be able to formulate two-sided tests that permit the clean identification of causality in a structural sense. Among the natural experiments one could exploit, are the major negative shock to housing and financial wealth that many people experienced during the ongoing financial market crisis of 2008/09, the positive shock Medicare households received as a result of the introduction of the heavily subsidized Medicare Part D program in 2006, and the shocks some employed individuals received from changes in employer-provided health insurance.[17] Particular attention could be given to the differential exposure to wealth shocks in the presence of health care delivery systems that vary in the financial impact of

[16] While invariance tests have arguably gained power by the inclusion of different time spans, cohorts, and age structures, we are still doubtful that their acceptance would attest the model the kind of stability necessary to make out-of-sample predictions of policy effects. The reason for this is that – with the notable exception of the introduction of Medicare Part D in 2006 – the observed variation in relevant policies remains rather low.

[17] When it comes to the recent financial crisis, we acknowledge that the equity shock might not be large enough to provide strong identification. Using HRS data, Gustman *et al.* (2010) report that equity accounted only for about 15% of assets prior to the 2008/09 crisis. Whether this is sufficient exogenous variation would have to be scrutinized as part of future research. Alternatively, one could explore negative shocks to housing wealth which represent another aspect of the financial market crunch. Exogenous variation in these shocks is provided by regional differences in house prices and the severity of declines in real estate value during the crisis.

copayments, premiums, and coverage, particularly for chronic conditions and preventative and palliative therapies. Provided that the causal link in questions even exists, wealth shocks will take some time to affect health outcomes. Therefore, we expect any effects of the 2008/2009 financial crisis or Medicare Part D to leave their marks only in future waves of the HRS dataset.

However, the use of such natural experiments is not immune to objection, which leads to a fundamental trade-off. On the one hand, we can try to infer causality by relying on wealth shocks like the ones just described, which has the advantage of not having to worry about endogeneity issues. Yet, as argued in section 1.3, there is a risk that these shocks may not be all that *relevant* for health, especially when occurring late in life. On the other hand, the information contained in past levels of SES – the regressor used in HWW's G-causality framework – is certainly of great relevance, as it reflects the entire history between SES and health status. The disadvantage is that this pool of information may also include confounding elements, such as the impact of hidden common causes, calling into question the exogeneity of such explanatory variables.

The other alternative we deem feasible of discriminating among hypotheses A and C, seeks to solve this trade-off by exploiting the relevant information contained in SES histories, while eliminating the misleading influence of common effects. As extensively argued in the fixed- and random-effects literature, this may be achieved by interpreting the problem of common effects as an issue of unobserved individual heterogeneity, whose effect is controlled by fully exploiting the panel structure of HRS. This being said, the choice of a suitable estimator is not trivial because it needs to combine three important features that often tend to be mutually exclusive. First and foremost, the estimation strategy must allow heterogeneity to be correlated with SES, which makes FE estimators a logical candidate. However, FE estimation is generally ridden by matters of inconsistency, once confronted with the other two features, namely a dependent variable that is both binary (requiring a non-linear specification) and state-dependent (reflecting the dynamic nature of the model). A feasible way of tackling these three issues at once, promises to be a dynamic correlated RE Probit approach as implemented by Contoyannis *et al.* (2004). It solves the usual trade-offs between FE and RE setups, by allowing for correlated heterogeneity and the estimation of time-invariant regressors even when confronted with non-linear data structures and lagged dependent variables.[18] We ac-

[18] Michaud and van Soest (2008) adopt a similar strategy, by eliminating the effect of individual heterogeneity with GMM estimators in the spirit of Arellano and Bond (1991).

knowledge that this alternative strategy of coping with common effects is not devoid of criticism either, which is why we consider it reasonable to independently explore both routes in what lies ahead. This is especially true inasmuch as both approaches are expected to uncover different causal channels: while the latter strategy of modeling individual heterogeneity may allow the detection of average causal effects as manifest in SES histories, the exploitation of natural experiments will predominantly shed light on the most immediate (mental) health consequences of wealth shocks.

A second opportunity for future research lies in improving the limited microfoundation of causal pathways, which is inherent in the reduced-form nature of Granger causality. Even if we were able to univocally confirm the presence of causal effects from wealth to health, we still would not know the channels through which they operate. Yet, the latter information is absolutely critical from a policy perspective: interventions to increase the affordability of health insurance would be warranted if channel A1 were to be active, but would prove ineffective if the causal link were to work through, say, channel A3 instead. To address this issue, we intend to specify and test more differentiated hypotheses that may facilitate the discrimination among these channels. For instance, if channel A1 is truly relevant, we should observe a certain sensitivity of results to the availability and generosity of health care systems. Possible comparisons include the time before and after Medicare Part D or cross-country differences in health care regimes.[19] Another way of gauging the importance of health care affordability is to compare individuals with and without health insurance. Of particular interest will be the pre-retirement population not yet eligible for Medicare, as their insurance status will be endogenous unless they are covered by employer-provided health care. Even if health insurance proves to be of little importance for the *onset* of a health condition, it may well be decisive in determining whether and how it is treated, given that the individual has already gotten sick. On this account we intend to follow the health trajectories as well as medical care use of respondents that share the characteristic of having developed a certain medical condition.

In analyzing the HRS population aged 51–61, they find that causal effects of wealth on health can be ruled out if unobserved heterogeneity and a more realistic lag structure are accounted for. However, given that their approach is incompatible with non-linear models, it is not directly applicable to our research question.

[19] In fact, Hurd and Kapteyn (2003) find that causal effects from SES to health status are less pronounced in the Netherlands than in the USA. Given that the Dutch health care system is basically universal, they see this result as an indication of the general importance of differential access to health care: SES gradients in health are strongest in institutional environments in which affordability should a priori matter most.

Finally, the model would certainly benefit from addressing another of the methodological shortcomings identified in section 1.4: the treatment of health dynamics. In our view, there are several ways to accommodate the long memory effects that prove to be so critical for a realistic description of health trajectories. A straightforward fix consists of adding higher-order lags of health condition prevalences to the list of explanatory variables. A more demanding alternative is a hidden Markov structure in which health is controlled by a latent random process that drives the onset of health conditions, self-rated health and mortality. According to Heiss (2011), such models are parsimonious and capture the observed dynamics better than commonly applied random-effects or conditional Markov chain models.

1.A Appendix: Additional tables

The following three tables contain summary statistics for the dataset used in our analysis as well as detailed results of Granger non-causality tests. Due to their large dimensions, tables start on the next page.

Table 1.6. *Variables used for analysis*
Summary statistics

Variable	Wave 2 (N=18,694)		Wave 3 (N=18,022)		Wave 4 (N=21,645)		Wave 5 (N=19,961)		Wave 6 (N=18,703)		Wave 7 (N=20,365)		Wave 8 (N=18,806)		Wave 9 (N=17,588)	
	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.
Health prevalence																
Cancer (ever)	0.097	0.296	0.127	0.333	0.127	0.333	0.140	0.347	0.155	0.362	0.152	0.359	0.168	0.373	0.179	0.384
Heart disease (ever)	0.225	0.418	0.265	0.442	0.256	0.436	0.270	0.444	0.287	0.452	0.275	0.447	0.296	0.456	0.306	0.461
Stroke (ever)	0.066	0.248	0.090	0.286	0.090	0.287	0.098	0.297	0.105	0.306	0.099	0.298	0.107	0.309	0.113	0.316
Lung disease (ever)	0.107	0.309	0.115	0.319	0.102	0.302	0.103	0.304	0.106	0.308	0.105	0.307	0.112	0.316	0.120	0.325
Diabetes (ever)	0.133	0.340	0.150	0.357	0.140	0.357	0.161	0.368	0.183	0.387	0.187	0.390	0.211	0.408	0.229	0.420
High blood pressure (ever)	0.464	0.499	0.500	0.500	0.488	0.500	0.517	0.500	0.556	0.497	0.553	0.497	0.597	0.491	0.629	0.483
Arthritis (last 2 years)	0.190	0.392	0.227	0.419	0.220	0.414	0.223	0.416	0.237	0.425	0.228	0.420	0.231	0.421	0.240	0.427
Incontinence (ever)	0.146	0.353	0.214	0.410	0.237	0.425	0.290	0.454	0.331	0.471	0.324	0.468	0.364	0.481	0.398	0.490
Fall (ever)	0.075	0.263	0.148	0.355	0.169	0.374	0.203	0.402	0.225	0.418	0.247	0.431	0.262	0.440	0.288	0.453
Hip fracture (ever)	0.024	0.152	0.025	0.157	0.027	0.162	0.029	0.167	0.031	0.173	0.035	0.183	0.038	0.192	0.040	0.195
Proxy interview (now)	0.077	0.267	0.138	0.345	0.145	0.352	0.164	0.370	0.182	0.386	0.143	0.350	0.131	0.337	0.134	0.341
Cognitive impairment (ever)	0.221	0.415	0.151	0.358	0.208	0.406	0.257	0.437	0.254	0.435	0.198	0.398	0.234	0.424	0.230	0.421
Psychiatric disease (ever)	0.128	0.334	0.158	0.365	0.151	0.358	0.159	0.365	0.170	0.375	0.177	0.381	0.191	0.393	0.199	0.400
Depression (last 2 years)	0.130	0.337	0.096	0.295	0.110	0.313	0.111	0.314	0.112	0.315	0.111	0.314	0.114	0.318	0.104	0.305
BMI (now)	26.5	5.0	26.5	5.1	26.8	5.2	27.0	5.4	27.2	5.4	27.4	5.7	27.8	5.8	28.0	5.9
Smoker (now)	0.182	0.386	0.169	0.375	0.164	0.371	0.149	0.356	0.137	0.343	0.148	0.355	0.136	0.343	0.133	0.339
# of ADL impairments (now)	0.239	0.751	0.378	0.990	0.376	0.998	0.384	1.003	0.393	1.020	0.364	0.984	0.399	1.022	0.403	1.035
# of IADL impairments (now)	0.288	0.847	0.344	0.967	0.336	0.966	0.345	0.990	0.370	1.027	0.340	0.964	0.369	1.014	0.377	1.027
Poor/fair self-rated h. (now)	0.291	0.454	0.282	0.450	0.319	0.466	0.286	0.452	0.294	0.456	0.300	0.458	0.301	0.459	0.307	0.461
Health incidence																
Cancer (first/new)			0.040	0.197	0.046	0.209	0.039	0.194	0.044	0.206	0.041	0.198	0.042	0.201	0.046	0.209
Heart disease (first/new)			0.092	0.289	0.125	0.331	0.099	0.299	0.110	0.313	0.117	0.322	0.107	0.309	0.110	0.313
Stroke (first/new)			0.038	0.191	0.047	0.213	0.038	0.190	0.042	0.201	0.039	0.194	0.036	0.186	0.039	0.193
Died since last wave			0.054	0.225	0.055	0.229	0.065	0.246	0.076	0.264	0.058	0.234	0.066	0.249	0.071	0.257
Lung disease (first)			0.017	0.127	0.020	0.139	0.017	0.130	0.020	0.141	0.023	0.149	0.018	0.133	0.021	0.142
Diabetes (first)			0.023	0.151	0.023	0.150	0.023	0.150	0.032	0.176	0.028	0.166	0.032	0.176	0.030	0.171
High blood pressure (first)			0.043	0.204	0.045	0.207	0.050	0.218	0.059	0.235	0.054	0.227	0.055	0.227	0.052	0.221
Arthritis (first/new)			0.105	0.307	0.099	0.298	0.098	0.298	0.110	0.313	0.109	0.312	0.099	0.298	0.106	0.308
Incontinence (first/new)			0.159	0.365	0.166	0.372	0.183	0.387	0.201	0.401	0.201	0.401	0.220	0.414	0.252	0.440

Fall (first/new)	0.110	0.313		0.109	0.312		0.101	0.302		0.104	0.306		0.109	0.312		0.112	0.316		0.126	0.332	
Hip fracture (first/new)	0.008	0.089		0.013	0.112		0.009	0.094		0.013	0.111		0.014	0.119		0.015	0.121		0.016	0.124	
Proxy interview (now)	0.138	0.345		0.145	0.352		0.164	0.370		0.182	0.386		0.143	0.350		0.131	0.337		0.134	0.341	
Cognitive impairment (first)	0.096	0.295		0.074	0.261		0.058	0.233		0.062	0.240		0.047	0.212		0.047	0.212		0.482	0.214	
Psychiatric disease (first)	0.032	0.175		0.027	0.162		0.024	0.152		0.027	0.162		0.025	0.157		0.028	0.165		0.022	0.147	
Depression (first/new)	0.050	0.218		0.072	0.258		0.061	0.240		0.062	0.242		0.059	0.236		0.060	0.238		0.053	0.223	
BMI (now)	26.5	5.1		26.8	5.2		27.0	5.4		27.2	5.4		27.4	5.7		27.8	5.8		28.0	5.9	
Smoker (now)	0.169	0.375		0.164	0.371		0.149	0.356		0.137	0.343		0.148	0.355		0.136	0.343		0.133	0.339	
# of ADL impairments (now)	0.378	0.990		0.376	0.998		0.384	1.003		0.393	1.020		0.364	0.984		0.399	1.022		0.403	1.035	
# of IADL impairments (now)	0.344	0.967		0.336	0.966		0.345	0.990		0.370	1.027		0.340	0.964		0.369	1.014		0.377	1.027	
Poor/fair self-rated h. (now)	0.282	0.450		0.319	0.466		0.286	0.452		0.294	0.456		0.300	0.458		0.301	0.459		0.307	0.461	
SES variables																					
Wealth in 1997 USD (000)	225.9	469.6		273.8	704.3		288.2	1,057		315.7	853.1		310.0	789.0		348.6	1,226		418.8	1,877		360.6	1,034	
1st quartile wealth indicator	0.264	0.441		0.246	0.431		0.253	0.435		0.239	0.427		0.229	0.421		0.246	0.430		0.239	0.426		0.240	0.427	
4th quartile wealth indicator	0.190	0.393		0.215	0.411		0.236	0.425		0.261	0.439		0.275	0.447		0.283	0.451		0.311	0.463		0.304	0.450	
Income in 1997 USD (000)	44.1	79.5		45.2	65.9		47.2	90.2		47.7	93.1		45.0	80.8		50.1	85.3		50.6	255.4		49.0	359.1	
1st quartile income indicator	0.289	0.453		0.268	0.443		0.251	0.434		0.253	0.435		0.250	0.433		0.240	0.427		0.242	0.428		0.251	0.434	
4th quartile income indicator	0.213	0.410		0.230	0.421		0.247	0.432		0.249	0.432		0.236	0.425		0.263	0.441		0.249	0.432		0.253	0.435	
Poor/fair housing condition	0.108	0.310		0.119	0.324		0.113	0.317		0.105	0.307		0.095	0.294		0.124	0.330		0.114	0.318		0.109	0.311	
Poor/fair neighborhood safety	0.146	0.353		0.122	0.327		0.101	0.301		0.087	0.282		0.078	0.268		0.097	0.296		0.101	0.301		0.096	0.295	
Education (in years)	11.5	3.5		11.6	3.5		11.8	3.4		11.9	3.4		12.0	3.4		12.3	3.3		12.3	3.3		12.4	3.3	
High school (educ. > 10 y.)	0.701	0.458		0.706	0.456		0.737	0.440		0.746	0.435		0.758	0.429		0.787	0.410		0.795	0.404		0.801	0.400	
College (educ. > 14 y.)	0.181	0.385		0.184	0.387		0.204	0.403		0.211	0.408		0.218	0.413		0.242	0.428		0.244	0.429		0.251	0.434	
Demographic variables																					
Widowed	0.200	0.400		0.209	0.407		0.203	0.402		0.216	0.412		0.230	0.421		0.201	0.401		0.213	0.410		0.221	0.415	
Divorced/separated	0.087	0.282		0.087	0.281		0.095	0.293		0.095	0.294		0.097	0.296		0.109	0.311		0.109	0.312		0.114	0.318	
Never married	0.031	0.174		0.029	0.168		0.033	0.177		0.031	0.174		0.030	0.170		0.037	0.188		0.034	0.182		0.034	0.180	
Current age	65.9	11.0		67.2	10.7		66.5	10.9		67.7	10.6		69.0	10.3		67.0	11.5		68.4	11.1		69.6	10.8	
Mother's current/death age	73.7	15.3		74.4	15.3		75.0	15.0		75.4	15.1		75.8	15.2		75.6	14.7		76.1	14.8		76.5	14.8	
Father's current/death age	70.8	14.4		71.2	14.6		71.4	14.4		71.7	14.5		71.9	14.5		71.9	14.2		71.1	14.3		71.4	14.4	
Ever smoked	0.588	0.492		0.587	0.492		0.590	0.492		0.586	0.493		0.583	0.493		0.579	0.494		0.579	0.494		0.577	0.494	

Notes: Summary statistics are for full sample (50+). Abbreviations are as follows: ADL=Activities of daily living; IADL=Instrumental activities of daily living.

Table 1.7. *Tests for invariance and non-causality (Ages 70+)*
Empirical significance levels

Health condition	Test	Sex	Significance levels of tests								
			Benchmark: HWW (70+)			3-waves samples (70+)			Longer samples (70+)		
			Article	Appendix	New data	No refilling			No refilling		Refilling
			W234 (AHEAD)	W234 (AHEAD)	W234 (AHEAD)	W234 (All Coh.)	W456 (All Coh.)	W789 (All Coh.)	W2–6 (All Coh.)	W4–9 (All Coh.)	W2–9 (All Coh.)
			(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(J)
Cancer	I	F	0.056	0.528	0.933	0.930	0.870	0.605	0.683	0.777	0.591
	I|noS	F	0.023	0.309	0.895	0.896	0.845	0.543	0.628	0.788	0.489
	S|I	F	0.311	0.613	0.271	0.264	0.483	0.280	0.350	0.020	0.048
	I&S	F	0.057	0.600	0.855	0.846	0.854	0.515	0.653	0.546	0.441
	I	M	0.000	0.016	0.119	0.172	0.355	0.179	0.349	0.585	0.156
	I|noS	M	0.000	0.007	0.095	0.134	0.158	0.450	0.187	0.419	0.263
	S|I	M	0.225	0.198	0.203	0.178	0.806	0.302	0.047	0.294	0.030
	I&S	M	0.000	0.012	0.088	0.117	0.516	0.163	0.194	0.545	0.081
Heart disease	I	F	0.000	0.085	0.000	0.000	0.849	0.392	0.017	0.874	0.051
	I|noS	F	0.000	0.052	0.000	0.000	0.903	0.440	0.010	0.946	0.084
	S|I	F	0.812	0.398	0.068	0.059	0.701	0.010	0.301	0.406	0.020
	I&S	F	0.000	0.099	0.000	0.000	0.891	0.080	0.016	0.863	0.021
	I	M	0.017	0.690	0.505	0.603	0.248	0.635	0.484	0.262	0.185
	I|noS	M	0.020	0.619	0.458	0.566	0.197	0.856	0.384	0.702	0.343
	S|I	M	0.290	0.243	0.162	0.224	0.468	0.533	0.216	0.077	0.098
	I&S	M	0.018	0.571	0.349	0.478	0.279	0.662	0.413	0.168	0.130
Stroke	I	F	0.068	0.384	0.016	0.015	0.814	0.690	0.023	0.710	0.206
	I|noS	F	0.034	0.266	0.030	0.028	0.817	0.578	0.035	0.791	0.313
	S|I	F	0.056	0.657	0.193	0.147	0.044	0.197	0.062	0.311	0.065
	I&S	F	0.023	0.481	0.012	0.009	0.461	0.540	0.010	0.675	0.135
	I	M	0.086	0.204	0.030	0.017	0.403	0.446	0.036	0.745	0.015
	I|noS	M	0.042	0.135	0.006	0.003	0.276	0.334	0.004	0.437	0.014
	S|I	M	0.290	0.059	0.213	0.195	0.750	0.664	0.215	0.529	0.682
	I&S	M	0.080	0.080	0.024	0.013	0.535	0.543	0.029	0.757	0.020

Mortality	I	F	0.010	0.221	0.092	0.085	0.169	0.517	0.165	0.491	0.011
	I|noS	F	0.006	0.315	0.161	0.162	0.220	0.403	0.321	0.355	0.010
	S|I	F	0.812	0.652	0.331	0.309	0.453	0.108	0.248	0.092	0.029
	I&S	F	0.030	0.308	0.093	0.083	0.195	0.312	0.140	0.368	0.005
	I	M	0.021	0.378	0.306	0.262	0.846	0.370	0.491	0.156	0.077
	I|noS	M	0.032	0.492	0.236	0.227	0.594	0.395	0.318	0.064	0.072
	S|I	M	0.228	0.364	0.274	0.247	0.371	0.154	0.331	0.002	0.009
	I&S	M	0.018	0.360	0.259	0.212	0.796	0.244	0.462	0.033	0.027
Lung disease	I	F	0.493	0.552	0.386	0.403	0.154	0.526	0.421	0.253	0.496
	I|noS	F	0.479	0.483	0.486	0.487	0.154	0.324	0.435	0.273	0.732
	S|I	F	0.381	0.343	0.218	0.190	0.669	0.057	0.185	0.219	0.329
	I&S	F	0.470	0.504	0.295	0.291	0.235	0.250	0.343	0.213	0.474
	I	M	0.603	0.689	0.210	0.273	0.059	0.493	0.102	0.069	0.241
	I|noS	M	0.620	0.502	0.064	0.094	0.119	0.428	0.082	0.074	0.192
	S|I	M	0.013	0.010	0.023	0.026	0.084	0.617	0.161	0.584	0.255
	I&S	M	0.174	0.203	0.052	0.076	0.026	0.569	0.073	0.085	0.215
Diabetes	I	F	0.110	0.189	0.344	0.379	0.572	0.667	0.158	0.282	0.384
	I|noS	F	0.177	0.281	0.407	0.403	0.477	0.655	0.098	0.143	0.593
	S|I	F	0.246	0.110	0.657	0.601	0.086	0.672	0.483	0.134	0.071
	I&S	F	0.091	0.101	0.439	0.454	0.329	0.740	0.174	0.211	0.280
	I	M	0.243	0.234	0.377	0.269	0.603	0.393	0.029	0.103	0.063
	I|noS	M	0.199	0.128	0.284	0.179	0.585	0.403	0.021	0.061	0.068
	S|I	M	0.043	0.025	0.188	0.149	0.609	0.304	0.794	0.419	0.115
	I&S	M	0.085	0.062	0.271	0.171	0.663	0.345	0.049	0.107	0.044
High blood pressure	I	F	0.004	0.007	0.008	0.010	0.154	0.217	0.006	0.683	0.099
	I|noS	F	0.009	0.024	0.043	0.047	0.051	0.138	0.018	0.416	0.074
	S|I	F	0.777	0.534	0.488	0.516	0.018	0.465	0.835	0.084	0.228
	I&S	F	0.012	0.013	0.014	0.018	0.032	0.246	0.012	0.550	0.084
	I	M	0.220	0.393	0.074	0.083	0.053	0.759	0.049	0.111	0.186
	I|noS	M	0.121	0.172	0.014	0.017	0.029	0.579	0.029	0.041	0.052
	S|I	M	0.668	0.990	0.697	0.633	0.474	0.283	0.327	0.035	0.091
	I&S	M	0.310	0.666	0.132	0.133	0.072	0.667	0.047	0.052	0.130

Arthritis	I	F	0.041	0.046	0.093	0.087	0.270	0.928	0.363	0.399	0.053
	I|noS	F	0.017	0.032	0.104	0.109	0.156	0.663	0.234	0.352	0.015
	S|I	F	0.042	0.085	0.034	0.055	0.622	0.003	0.011	0.035	0.000
	I&S	F	0.012	0.020	0.025	0.031	0.349	0.342	0.145	0.240	0.005
	I	M	0.187	0.071	0.081	0.102	0.259	0.960	0.125	0.581	0.884
	I|noS	M	0.276	0.167	0.110	0.151	0.246	0.963	0.112	0.325	0.690
	S|I	M	0.145	0.395	0.382	0.313	0.472	0.409	0.404	0.276	0.009
	I&S	M	0.116	0.082	0.090	0.098	0.289	0.935	0.128	0.538	0.718
Incontinence	I	F	0.357	0.781	0.285	0.271	0.494	0.798	0.195	0.540	0.185
	I|noS	F	0.329	0.597	0.235	0.238	0.646	0.793	0.199	0.727	0.402
	S|I	F	0.080	0.163	0.082	0.077	0.010	0.015	0.004	0.000	0.000
	I&S	F	0.183	0.615	0.142	0.130	0.126	0.351	0.048	0.134	0.000
	I	M	0.161	0.351	0.681	0.614	0.257	0.236	0.553	0.022	0.393
	I|noS	M	0.486	0.682	0.924	0.903	0.396	0.337	0.730	0.312	0.616
	S|I	M	0.237	0.463	0.296	0.374	0.316	0.010	0.175	0.048	0.002
	I&S	M	0.129	0.373	0.602	0.578	0.234	0.043	0.464	0.010	0.169
Fall	I	F	0.864	0.904	0.616	0.626	0.100	0.955	0.089	0.444	0.359
	I|noS	F	0.763	0.790	0.372	0.384	0.193	0.997	0.037	0.525	0.478
	S|I	F	0.515	0.600	0.494	0.482	0.507	0.064	0.296	0.261	0.015
	I&S	F	0.861	0.913	0.630	0.634	0.131	0.764	0.081	0.402	0.205
	I	M	0.402	0.263	0.126	0.248	0.044	0.549	0.013	0.275	0.301
	I|noS	M	0.383	0.207	0.092	0.159	0.028	0.481	0.005	0.111	0.073
	S|I	M	0.507	0.592	0.347	0.474	0.338	0.103	0.009	0.079	0.015
	I&S	M	0.438	0.329	0.126	0.278	0.047	0.342	0.003	0.185	0.164
Hip fracture	I	F	0.604	0.492	0.260	0.256	0.711	0.790	0.083	0.569	0.143
	I|noS	F	0.470	0.280	0.283	0.284	0.646	0.907	0.086	0.302	0.043
	S|I	F	0.275	0.159	0.275	0.259	0.597	0.983	0.195	0.821	0.689
	I&S	F	0.520	0.348	0.222	0.213	0.750	0.925	0.064	0.643	0.170
	I	M	0.056	0.126	0.010	0.017	0.359	0.104	0.000	0.097	0.006
	I|noS	M	0.051	0.053	0.105	0.111	0.203	0.094	0.036	0.155	0.006
	S|I	M	0.305	0.430	0.069	0.078	0.003	0.686	0.011	0.048	0.203
	I&S	M	0.055	0.143	0.004	0.007	0.048	0.167	0.000	0.051	0.005

Proxy interview	I	F	0.345	0.301	0.597	0.582	0.283	0.764	0.953	0.482	0.444
	I|noS	F	0.442	0.465	0.922	0.922	0.506	0.663	0.998	0.782	0.652
	S|I	F	0.250	0.254	0.130	0.144	0.236	0.134	0.009	0.182	0.001
	I&S	F	0.283	0.248	0.413	0.412	0.226	0.577	0.784	0.414	0.185
	I	M	0.424	0.591	0.013	0.018	0.040	0.642	0.000	0.001	0.001
	I|noS	M	0.326	0.378	0.009	0.013	0.017	0.884	0.000	0.001	0.010
	S|I	M	0.019	0.032	0.000	0.000	0.002	0.925	0.000	0.004	0.000
	I&S	M	0.131	0.261	0.000	0.000	0.001	0.803	0.000	0.000	0.000
Cognitive impairment	I	F	0.020	0.005	0.278	0.287	0.417	0.057	0.445	0.053	0.090
	I|noS	F	0.007	0.008	0.436	0.435	0.239	0.108	0.430	0.013	0.127
	S|I	F	0.001	0.002	0.002	0.002	0.034	0.007	0.052	0.000	0.000
	I&S	F	0.000	0.000	0.027	0.026	0.160	0.006	0.285	0.004	0.001
	I	M	0.429	0.288	0.129	0.167	0.357	0.772	0.219	0.034	0.157
	I|noS	M	0.245	0.143	0.068	0.113	0.408	0.705	0.166	0.052	0.228
	S|I	M	0.022	0.026	0.049	0.060	0.031	0.010	0.036	0.001	0.000
	I&S	M	0.140	0.089	0.046	0.068	0.126	0.291	0.109	0.005	0.003
Psychiatric disease	I	F	0.075	0.127	0.164	0.154	0.711	0.075	0.074	0.945	0.175
	I|noS	F	0.031	0.052	0.147	0.150	0.491	0.296	0.063	0.950	0.269
	S|I	F	0.012	0.004	0.012	0.011	0.270	0.522	0.041	0.737	0.299
	I&S	F	0.012	0.012	0.031	0.028	0.622	0.102	0.033	0.958	0.162
	I	M	0.194	0.295	0.036	0.037	0.009	0.099	0.066	0.006	0.088
	I|noS	M	0.546	0.365	0.106	0.115	0.028	0.174	0.212	0.021	0.143
	S|I	M	0.110	0.065	0.110	0.080	0.674	0.004	0.428	0.076	0.035
	I&S	M	0.108	0.137	0.019	0.016	0.019	0.009	0.071	0.003	0.048
Depression	I	F	0.299	0.211	0.196	0.197	0.236	0.125	0.199	0.732	0.068
	I|noS	F	0.347	0.299	0.503	0.505	0.182	0.133	0.177	0.748	0.095
	S|I	F	0.078	0.011	0.086	0.107	0.001	0.000	0.000	0.000	0.000
	I&S	F	0.151	0.042	0.098	0.109	0.019	0.002	0.023	0.062	0.000
	I	M	0.767	0.944	0.578	0.694	0.185	0.252	0.007	0.452	0.093
	I|noS	M	0.856	0.934	0.478	0.581	0.209	0.125	0.007	0.414	0.041
	S|I	M	0.302	0.065	0.003	0.002	0.005	0.016	0.000	0.007	0.000
	I&S	M	0.695	0.753	0.112	0.148	0.023	0.062	0.000	0.222	0.002

BMI	I	F	0.419	0.260	0.192	0.192	0.113	0.008	0.039	0.038	0.001
	I|noS	F	0.330	0.251	0.096	0.104	0.127	0.025	0.018	0.027	0.000
	S|I	F	0.738	0.673	0.932	0.953	0.352	0.083	0.503	0.001	0.011
	I&S	F	0.531	0.345	0.356	0.368	0.115	0.003	0.046	0.006	0.000
	I	M	0.010	0.009	0.011	0.015	0.000	0.353	0.000	0.000	0.000
	I|noS	M	0.002	0.002	0.014	0.017	0.001	0.618	0.000	0.000	0.000
	S|I	M	0.249	0.664	0.664	0.623	0.642	0.248	0.931	0.316	0.487
	I&S	M	0.009	0.019	0.022	0.028	0.000	0.290	0.000	0.000	0.000
Smoke now	I	F	0.509	0.420	0.151	0.158	0.267	0.760	0.085	0.421	0.031
	I|noS	F	0.242	0.214	0.063	0.063	0.382	0.763	0.062	0.655	0.046
	S|I	F	0.838	0.625	0.637	0.635	0.034	0.082	0.140	0.309	0.172
	I&S	F	0.650	0.492	0.210	0.219	0.097	0.525	0.060	0.396	0.024
	I	M	0.366	0.710	0.598	0.579	0.001	0.128	0.000	0.577	0.221
	I|noS	M	0.146	0.504	0.390	0.420	0.068	0.196	0.679	0.016	0.374
	S|I	M	0.064	0.129	0.217	0.424	0.539	0.721	0.459	0.400	0.305
	I&S	M	0.182	0.528	0.488	0.570	0.001	0.200	0.000	0.569	0.208
ADL impairment	I	F	0.010	0.004	0.020	0.019	0.518	0.133	0.046	0.193	0.000
	I|noS	F	0.018	0.022	0.047	0.048	0.343	0.156	0.031	0.181	0.000
	S|I	F	0.818	0.831	0.859	0.850	0.011	0.049	0.238	0.000	0.000
	I&S	F	0.027	0.011	0.049	0.046	0.170	0.052	0.039	0.042	0.000
	I	M	0.004	0.067	0.016	0.020	0.339	0.229	0.026	0.397	0.018
	I|noS	M	0.044	0.114	0.018	0.026	0.692	0.640	0.305	0.468	0.264
	S|I	M	0.370	0.352	0.093	0.131	0.232	0.068	0.453	0.639	0.081
	I&S	M	0.005	0.069	0.008	0.012	0.275	0.111	0.029	0.433	0.011
IADL impairment	I	F	0.673	0.724	0.615	0.598	0.207	0.276	0.467	0.108	0.442
	I|noS	F	0.514	0.525	0.555	0.544	0.150	0.296	0.312	0.106	0.337
	S|I	F	0.368	0.230	0.303	0.281	0.021	0.668	0.114	0.110	0.053
	I&S	F	0.636	0.625	0.555	0.529	0.061	0.353	0.364	0.075	0.335
	I	M	0.016	0.024	0.038	0.022	0.658	0.202	0.009	0.137	0.053
	I|noS	M	0.003	0.011	0.024	0.014	0.633	0.069	0.003	0.084	0.010
	S|I	M	0.006	0.009	0.006	0.004	0.127	0.009	0.023	0.001	0.000
	I&S	M	0.002	0.003	0.004	0.002	0.486	0.041	0.003	0.031	0.002

Self-rated	I	F	0.151	0.376	0.006	0.008	0.021	0.493	0.004	0.000	0.000	0.000
health	I|noS	F	0.111	0.341	0.018	0.019	0.078	0.358	0.024	0.000	0.000	0.000
	S|I	F	0.001	0.001	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000
	I&S	F	0.009	0.044	0.000	0.000	0.004	0.001	0.000	0.000	0.000	0.000
	I	M	0.581	0.634	0.733	0.650	0.033	0.062	0.052	0.039	0.039	0.004
	I|noS	M	0.570	0.607	0.610	0.574	0.008	0.033	0.019	0.009	0.009	0.006
	S|I	M	0.034	0.020	0.004	0.003	0.002	0.000	0.000	0.000	0.000	0.000
	I&S	M	0.282	0.278	0.226	0.169	0.002	0.000	0.001	0.000	0.000	0.000

Notes: Results are for white females (F) and males (M). Abbreviations of tests are as follows: I=Unconditional invariance; I|noS=Invariance, conditional on non-causality; S|I=Non-causality, conditional on invariance; I&S=Joint invariance and non-causality. A "n.a." entry reflects the fact that information on certain health conditions is not available for the pre-Medicare population.

Table 1.8. *Tests for invariance and non-causality (All Ages)*
Empirical significance levels

Health condition	Test	Sex	Significance levels of tests											
			No refilling								Refilling			
			W2–6				W4–9				W2–9			
			70+	65+	50+	50–64	70+	65+	50+	50–64	70+	65+	50+	50–64
			(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(J)	(K)	(L)	(M)
Cancer	I	F	0.683	0.666	0.889	0.126	0.777	0.506	0.265	0.388	0.591	0.250	0.497	0.410
	I|noS	F	0.628	0.454	0.720	0.125	0.788	0.513	0.238	0.306	0.489	0.078	0.281	0.318
	S|I	F	0.350	0.367	0.659	0.552	0.020	0.007	0.064	0.349	0.048	0.151	0.077	0.301
	I&S	F	0.653	0.642	0.909	0.151	0.546	0.229	0.159	0.370	0.441	0.198	0.375	0.383
	I	M	0.349	0.269	0.160	0.021	0.585	0.623	0.136	0.108	0.156	0.024	0.116	0.068
	I|noS	M	0.187	0.096	0.245	0.056	0.419	0.647	0.119	0.142	0.263	0.034	0.112	0.055
	S|I	M	0.047	0.082	0.179	0.993	0.294	0.438	0.315	0.689	0.030	0.109	0.110	0.917
	I&S	M	0.194	0.166	0.120	0.058	0.545	0.620	0.127	0.143	0.081	0.015	0.080	0.106
Heart disease	I	F	0.017	0.012	0.043	0.113	0.874	0.867	0.853	0.397	0.051	0.057	0.211	0.546
	I|noS	F	0.010	0.007	0.028	0.144	0.946	0.864	0.912	0.749	0.084	0.039	0.112	0.520
	S|I	F	0.301	0.192	0.035	0.679	0.406	0.053	0.041	0.709	0.020	0.020	0.001	0.115
	I&S	F	0.016	0.009	0.016	0.153	0.863	0.733	0.689	0.460	0.021	0.023	0.053	0.448
	I	M	0.484	0.380	0.405	0.601	0.262	0.336	0.558	0.523	0.185	0.109	0.480	0.117
	I|noS	M	0.384	0.241	0.149	0.376	0.702	0.569	0.485	0.474	0.343	0.200	0.343	0.056
	S|I	M	0.216	0.316	0.071	0.045	0.077	0.092	0.025	0.043	0.098	0.084	0.007	0.119
	I&S	M	0.413	0.352	0.255	0.384	0.168	0.234	0.333	0.336	0.130	0.070	0.245	0.083
Stroke	I	F	0.023	0.008	0.018	0.073	0.710	0.690	0.319	0.199	0.206	0.152	0.201	0.008
	I|noS	F	0.035	0.043	0.052	0.046	0.791	0.730	0.377	0.138	0.313	0.248	0.286	0.037
	S|I	F	0.062	0.167	0.007	0.038	0.311	0.195	0.051	0.252	0.065	0.027	0.007	0.332
	I&S	F	0.010	0.005	0.003	0.029	0.675	0.616	0.189	0.172	0.135	0.078	0.078	0.008
	I	M	0.036	0.042	0.001	0.063	0.745	0.398	0.025	0.011	0.015	0.010	0.005	0.096
	I|noS	M	0.004	0.006	0.000	0.046	0.437	0.151	0.012	0.010	0.014	0.008	0.001	0.113
	S|I	M	0.215	0.181	0.426	0.775	0.529	0.673	0.486	0.958	0.682	0.570	0.082	0.630
	I&S	M	0.029	0.031	0.001	0.099	0.757	0.450	0.030	0.025	0.020	0.013	0.003	0.115

Mortality	I	F	0.165	0.044	0.111	0.079	0.491	0.317	0.305	0.280	0.011	0.002	0.051	0.019
	I|noS	F	0.321	0.096	0.324	0.105	0.355	0.288	0.221	0.199	0.010	0.005	0.130	0.033
	S|I	F	0.248	0.423	0.033	0.029	0.092	0.111	0.005	0.269	0.029	0.030	0.000	0.060
	I&S	F	0.140	0.048	0.045	0.029	0.368	0.231	0.100	0.249	0.005	0.001	0.006	0.010
	I	M	0.491	0.218	0.071	0.049	0.156	0.400	0.424	0.022	0.077	0.118	0.037	0.008
	I|noS	M	0.318	0.138	0.093	0.045	0.064	0.254	0.178	0.005	0.072	0.081	0.007	0.001
	S|I	M	0.331	0.372	0.338	0.279	0.002	0.001	0.000	0.013	0.009	0.006	0.000	0.002
	I&S	M	0.462	0.213	0.069	0.044	0.033	0.096	0.078	0.006	0.027	0.041	0.003	0.001
Lung disease	I	F	0.421	0.511	0.267	0.537	0.253	0.159	0.153	0.216	0.496	0.516	0.347	0.606
	I|noS	F	0.435	0.455	0.175	0.298	0.273	0.556	0.220	0.118	0.732	0.600	0.351	0.564
	S|I	F	0.185	0.283	0.291	0.331	0.219	0.262	0.189	0.616	0.329	0.412	0.432	0.442
	I&S	F	0.343	0.465	0.240	0.504	0.213	0.140	0.120	0.253	0.474	0.512	0.349	0.605
	I	M	0.102	0.040	0.017	0.254	0.069	0.102	0.591	0.361	0.241	0.346	0.574	0.952
	I|noS	M	0.082	0.053	0.020	0.361	0.074	0.082	0.601	0.311	0.192	0.321	0.643	0.903
	S|I	M	0.161	0.210	0.166	0.893	0.584	0.464	0.014	0.021	0.255	0.026	0.004	0.096
	I&S	M	0.073	0.031	0.012	0.363	0.085	0.111	0.330	0.179	0.215	0.207	0.303	0.909
Diabetes	I	F	0.158	0.305	0.032	0.027	0.282	0.700	0.427	0.248	0.384	0.241	0.115	0.220
	I|noS	F	0.098	0.232	0.037	0.020	0.143	0.529	0.387	0.679	0.593	0.295	0.110	0.218
	S|I	F	0.483	0.434	0.000	0.000	0.134	0.116	0.000	0.000	0.071	0.004	0.000	0.000
	I&S	F	0.174	0.311	0.001	0.001	0.211	0.590	0.036	0.018	0.280	0.091	0.001	0.018
	I	M	0.029	0.092	0.391	0.729	0.103	0.242	0.067	0.338	0.063	0.047	0.029	0.021
	I|noS	M	0.021	0.067	0.266	0.684	0.061	0.135	0.110	0.569	0.068	0.016	0.016	0.115
	S|I	M	0.794	0.723	0.009	0.029	0.419	0.506	0.185	0.621	0.115	0.039	0.329	0.061
	I&S	M	0.049	0.129	0.145	0.481	0.107	0.259	0.051	0.381	0.044	0.023	0.027	0.011
High blood pressure	I	F	0.006	0.003	0.249	0.568	0.683	0.450	0.167	0.037	0.099	0.003	0.028	0.116
	I|noS	F	0.018	0.008	0.215	0.547	0.416	0.255	0.236	0.080	0.074	0.012	0.079	0.202
	S|I	F	0.835	0.690	0.440	0.708	0.084	0.111	0.134	0.232	0.228	0.202	0.025	0.061
	I&S	F	0.012	0.006	0.258	0.633	0.550	0.345	0.120	0.031	0.084	0.377	0.011	0.069
	I	M	0.049	0.077	0.115	0.297	0.111	0.686	0.506	0.264	0.186	0.633	0.111	0.401
	I|noS	M	0.029	0.048	0.103	0.456	0.041	0.427	0.591	0.338	0.052	0.235	0.167	0.709
	S|I	M	0.327	0.257	0.201	0.085	0.035	0.168	0.260	0.082	0.091	0.041	0.335	0.253
	I&S	M	0.047	0.066	0.090	0.188	0.052	0.604	0.458	0.173	0.130	0.001	0.106	0.364

Arthritis	I	F	0.363	0.198	0.895	0.656	0.399	0.592	0.629	0.271	0.053	0.153	0.502	0.565
	I|noS	F	0.234	0.107	0.974	0.697	0.352	0.388	0.459	0.676	0.015	0.040	0.199	0.668
	S|I	F	0.011	0.001	0.110	0.489	0.035	0.016	0.434	0.161	0.000	0.000	0.027	0.671
	I&S	F	0.145	0.032	0.800	0.664	0.240	0.356	0.626	0.211	0.005	0.029	0.333	0.605
	I	M	0.125	0.470	0.857	0.705	0.581	0.912	0.790	0.168	0.884	0.979	0.833	0.022
	I|noS	M	0.112	0.399	0.869	0.764	0.325	0.806	0.706	0.129	0.690	0.945	0.695	0.001
	S|I	M	0.404	0.168	0.020	0.040	0.276	0.058	0.017	0.269	0.009	0.001	0.000	0.053
	I&S	M	0.128	0.382	0.622	0.488	0.538	0.814	0.560	0.149	0.718	0.850	0.432	0.011
Incontinence	I	F	0.195	0.256	0.186	0.231	0.540	0.758	0.241	0.824	0.185	0.321	0.081	0.109
	I|noS	F	0.199	0.294	0.096	0.071	0.727	0.918	0.394	0.596	0.402	0.651	0.079	0.010
	S|I	F	0.004	0.002	0.000	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	I&S	F	0.048	0.059	0.006	0.116	0.134	0.066	0.000	0.320	0.000	0.001	0.000	0.014
	I	M	0.553	0.465	0.644	0.093	0.022	0.257	0.067	0.390	0.393	0.235	0.051	0.027
	I|noS	M	0.730	0.650	0.579	0.043	0.312	0.523	0.171	0.228	0.616	0.391	0.112	0.005
	S|I	M	0.175	0.023	0.013	0.190	0.048	0.018	0.002	0.028	0.002	0.000	0.000	0.055
	I&S	M	0.464	0.249	0.356	0.071	0.010	0.121	0.012	0.220	0.169	0.055	0.004	0.014
Fall	I	F	0.089	0.066	n.a.	n.a.	0.444	0.544	n.a.	n.a.	0.359	0.071	n.a.	n.a.
	I|noS	F	0.037	0.040	n.a.	n.a.	0.525	0.654	n.a.	n.a.	0.478	0.194	n.a.	n.a.
	S|I	F	0.296	0.347	n.a.	n.a.	0.261	0.063	n.a.	n.a.	0.015	0.039	n.a.	n.a.
	I&S	F	0.081	0.064	n.a.	n.a.	0.402	0.399	n.a.	n.a.	0.205	0.038	n.a.	n.a.
	I	M	0.013	0.004	n.a.	n.a.	0.275	0.399	n.a.	n.a.	0.301	0.156	n.a.	n.a.
	I|noS	M	0.005	0.001	n.a.	n.a.	0.111	0.159	n.a.	n.a.	0.073	0.030	n.a.	n.a.
	S|I	M	0.009	0.049	n.a.	n.a.	0.079	0.013	n.a.	n.a.	0.015	0.047	n.a.	n.a.
	I&S	M	0.003	0.002	n.a.	n.a.	0.185	0.203	n.a.	n.a.	0.164	0.094	n.a.	n.a.
Hip fracture	I	F	0.083	0.134	n.a.	n.a.	0.569	0.392	n.a.	n.a.	0.143	0.116	n.a.	n.a.
	I|noS	F	0.086	0.097	n.a.	n.a.	0.302	0.241	n.a.	n.a.	0.043	0.040	n.a.	n.a.
	S|I	F	0.195	0.206	n.a.	n.a.	0.821	0.713	n.a.	n.a.	0.689	0.325	n.a.	n.a.
	I&S	F	0.064	0.108	n.a.	n.a.	0.643	0.447	n.a.	n.a.	0.170	0.110	n.a.	n.a.
	I	M	0.000	0.155	n.a.	n.a.	0.097	0.175	n.a.	n.a.	0.006	0.026	n.a.	n.a.
	I|noS	M	0.036	0.384	n.a.	n.a.	0.155	0.202	n.a.	n.a.	0.006	0.034	n.a.	n.a.
	S|I	M	0.011	0.032	n.a.	n.a.	0.048	0.028	n.a.	n.a.	0.203	0.382	n.a.	n.a.
	I&S	M	0.000	0.070	n.a.	n.a.	0.051	0.087	n.a.	n.a.	0.005	0.026	n.a.	n.a.

Proxy	I	F	0.953	0.792	0.820	0.519	0.482	0.512	0.059	0.209	0.444	0.202	0.130	0.093	
interview	I|noS	F	0.998	0.975	0.964	0.480	0.782	0.745	0.094	0.130	0.652	0.376	0.145	0.021	
	S|I	F	0.009	0.007	0.000	0.577	0.182	0.317	0.000	0.026	0.001	0.001	0.000	0.073	
	I&S	F	0.784	0.490	0.278	0.553	0.414	0.484	0.001	0.101	0.185	0.054	0.001	0.059	
	I	M	0.000	0.000	0.003	0.071	0.001	0.005	0.001	0.105	0.001	0.000	0.000	0.018	
	I|noS	M	0.000	0.000	0.001	0.184	0.001	0.020	0.005	0.216	0.010	0.007	0.000	0.018	
	S|I	M	0.000	0.000	0.000	0.000	0.004	0.001	0.000	0.000	0.000	0.221	0.000	0.000	
	I&S	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	
Cognitive	I	F	0.445	0.356	n.a.	n.a.	0.053	0.353	n.a.	n.a.	0.090	0.092	n.a.	n.a.	
impairment	I|noS	F	0.430	0.395	n.a.	n.a.	0.013	0.234	n.a.	n.a.	0.127	0.154	n.a.	n.a.	
	S|I	F	0.052	0.018	n.a.	n.a.	0.000	0.000	n.a.	n.a.	0.000	0.000	n.a.	n.a.	
	I&S	F	0.285	0.168	n.a.	n.a.	0.004	0.004	n.a.	n.a.	0.001	0.000	n.a.	n.a.	
	I	M	0.219	0.218	n.a.	n.a.	0.034	0.244	n.a.	n.a.	0.157	0.225	n.a.	n.a.	
	I|noS	M	0.166	0.344	n.a.	n.a.	0.052	0.347	n.a.	n.a.	0.228	0.217	n.a.	n.a.	
	S|I	M	0.036	0.010	n.a.	n.a.	0.001	0.000	n.a.	n.a.	0.000	0.000	n.a.	n.a.	
	I&S	M	0.109	0.076	n.a.	n.a.	0.005	0.022	n.a.	n.a.	0.003	0.000	n.a.	n.a.	
Psychiatric	I	F	0.074	0.034	0.858	0.756	0.945	0.757	0.615	0.236	0.175	0.024	0.488	0.823	
disease	I|noS	F	0.063	0.029	0.748	0.741	0.950	0.794	0.531	0.167	0.269	0.033	0.340	0.885	
	S|I	F	0.041	0.004	0.000	0.000	0.737	0.208	0.028	0.027	0.299	0.196	0.000	0.000	
	I&S	F	0.033	0.006	0.260	0.397	0.958	0.701	0.410	0.118	0.162	0.019	0.044	0.362	
	I	M	0.066	0.204	0.506	0.790	0.006	0.038	0.357	0.043	0.088	0.013	0.260	0.747	
	I|noS	M	0.212	0.174	0.496	0.851	0.021	0.056	0.541	0.159	0.143	0.020	0.276	0.798	
	S|I	M	0.428	0.280	0.248	0.870	0.076	0.018	0.049	0.009	0.035	0.036	0.022	0.354	
	I&S	M	0.071	0.183	0.450	0.856	0.003	0.013	0.224	0.011	0.048	0.006	0.144	0.730	
Depression	I	F	0.199	0.224	0.224	0.438	0.732	0.207	0.292	0.708	0.068	0.175	0.293	0.827	
	I|noS	F	0.177	0.196	0.457	0.547	0.748	0.217	0.222	0.802	0.095	0.167	0.456	0.746	
	S|I	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I&S	F	0.023	0.018	0.000	0.001	0.062	0.001	0.000	0.004	0.000	0.000	0.000	0.000	
	I	M	0.007	0.069	0.207	0.158	0.452	0.052	0.285	0.233	0.093	0.138	0.529	0.065	
	I|noS	M	0.007	0.039	0.064	0.078	0.414	0.045	0.109	0.143	0.041	0.045	0.133	0.029	
	S|I	M	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I&S	M	0.000	0.001	0.000	0.000	0.222	0.004	0.000	0.003	0.002	0.001	0.000	0.000	

BMI	I	F	0.039	0.010	0.019	0.002	0.038	0.000	0.000	0.000	0.001	0.000	0.000	0.000	
	I|noS	F	0.018	0.005	0.008	0.004	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	S|I	F	0.503	0.328	0.221	0.437	0.001	0.004	0.004	0.116	0.011	0.008	0.002	0.042	
	I&S	F	0.046	0.010	0.015	0.002	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I|noS	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	S|I	M	0.931	0.828	0.260	0.388	0.316	0.182	0.006	0.102	0.487	0.007	0.004	0.147	
	I&S	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Smoke now	I	F	0.085	0.039	0.065	0.171	0.421	0.231	0.073	0.004	0.031	0.049	0.015	0.038	
	I|noS	F	0.062	0.061	0.031	0.161	0.655	0.343	0.228	0.113	0.046	0.031	0.007	0.261	
	S|I	F	0.140	0.046	0.000	0.000	0.309	0.084	0.000	0.000	0.172	0.166	0.000	0.000	
	I&S	F	0.060	0.018	0.010	0.006	0.396	0.159	0.005	0.000	0.024	0.038	0.000	0.000	
	I	M	0.000	0.904	0.103	0.351	0.577	0.087	0.700	0.170	0.221	0.415	0.087	0.003	
	I|noS	M	0.679	0.000	0.173	0.564	0.016	0.076	0.543	0.315	0.374	0.382	0.055	0.001	
	S|I	M	0.459	0.053	0.002	0.016	0.400	0.302	0.000	0.003	0.305	0.079	0.000	0.000	
	I&S	M	0.000	0.794	0.019	0.162	0.569	0.081	0.287	0.046	0.208	0.325	0.002	0.000	
ADL impairment	I	F	0.046	0.101	0.066	0.058	0.193	0.210	0.024	0.068	0.000	0.000	0.000	0.051	
	I|noS	F	0.031	0.072	0.079	0.020	0.181	0.244	0.036	0.030	0.000	0.000	0.000	0.028	
	S|I	F	0.238	0.295	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I&S	F	0.039	0.093	0.002	0.004	0.042	0.012	0.000	0.000	0.000	0.000	0.000	0.000	
	I	M	0.026	0.011	0.003	0.059	0.397	0.414	0.675	0.254	0.018	0.180	0.324	0.423	
	I|noS	M	0.305	0.264	0.038	0.019	0.468	0.688	0.850	0.443	0.264	0.386	0.381	0.208	
	S|I	M	0.453	0.197	0.000	0.000	0.639	0.034	0.000	0.000	0.081	0.000	0.000	0.000	
	I&S	M	0.029	0.009	0.000	0.003	0.433	0.273	0.003	0.001	0.011	0.036	0.000	0.004	
IADL impairment	I	F	0.467	0.602	0.577	0.546	0.108	0.008	0.006	0.344	0.442	0.243	0.131	0.052	
	I|noS	F	0.312	0.395	0.507	0.600	0.106	0.021	0.004	0.154	0.337	0.103	0.080	0.074	
	S|I	F	0.114	0.056	0.001	0.007	0.110	0.030	0.000	0.003	0.053	0.011	0.000	0.003	
	I&S	F	0.364	0.443	0.199	0.261	0.075	0.003	0.000	0.081	0.335	0.129	0.003	0.014	
	I	M	0.009	0.166	0.097	0.201	0.137	0.105	0.005	0.004	0.053	0.007	0.014	0.010	
	I|noS	M	0.003	0.092	0.071	0.251	0.084	0.082	0.021	0.015	0.010	0.008	0.005	0.014	
	S|I	M	0.023	0.027	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	I&S	M	0.003	0.078	0.000	0.006	0.031	0.007	0.000	0.000	0.002	0.000	0.000	0.000	

Self-rated	I	F	0.004	0.001	0.003	0.084	0.000	0.004	0.004	0.011	0.000	0.000	0.000	0.034		
health	I|noS	F	0.024	0.003	0.014	0.056	0.000	0.003	0.005	0.005	0.000	0.000	0.000	0.009		
	S|I	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	I&S	F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	I	M	0.052	0.017	0.020	0.216	0.039	0.016	0.021	0.063	0.004	0.001	0.001	0.011		
	I|noS	M	0.019	0.009	0.016	0.468	0.009	0.010	0.004	0.037	0.006	0.003	0.001	0.018		
	S|I	M	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
	I&S	M	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		

Notes: Results are for white females (F) and males (M). Abbreviations of tests are as follows: I=Unconditional invariance; I|noS=Invariance, conditional on non-causality; S|I=Non-causality, conditional on invariance; I&S=Joint invariance and non-causality. A "n.a." entry reflects the fact that information on certain health conditions is not available for the pre-Medicare population.

Chapter 2

Understanding the SES gradient in health: The roles of childhood circumstances and disease management among the elderly

2.1 Introduction

It is the health economic version of the classic "chicken and egg" problem: We know that people with high socio-economic status (SES) tend to be in better health and live longer than their economically disadvantaged counterparts but we are not sure which came first. Do economic resources determine health (hypothesis A)? Does health influence economic success (hypothesis B)? Or, are both health and wealth dependent on some third unaccounted factor (hypothesis C)? The body of literature dealing with this so-called socio-economic gradient in health is impressive (for an overview see Smith, 1999; Cutler *et al.*, 2011; and Stowasser *et al.*, 2012).[1] The traditional view that causality flows from SES to health is especially common among – but not exclusive to – epidemiologists. Often-cited causal pathways are the affordability of health services, better health knowledge and lifestyles among the higher educated, environmental hazards associated with poorly paying occupations and low-income living conditions, or the mere psychological burden that comes with a life of constant economic struggle. Economists were among the first to argue that causality may also work its way from health to economic outcomes, the most important channel being the development of human capital: Physical frailty is likely to have adverse effects on educational attainment, occupational productivity and, consequently, the accumulation of wealth. Finally, the statistical literature such as Heckman (1981b) stresses the point that the persistent

[1] Note that Stowasser *et al.* (2012) largely coincides with chapter 1 of this dissertation.

correlation between morbidity and SES may in fact be spurious and due to unobserved individual heterogeneity with a common influence on both health and wealth. Prime candidates for such hidden third factors are genetic disposition and other family effects with an impact on preferences and health-relevant behaviors.

The discrimination among these rivalling hypotheses is paramount since policy recommendations will critically depend on the nature and the sources of the gradient. Methodologically, the estimation of credible causal effects in population data requires addressing the challenges of simultaneity (hypothesis A vs. hypothesis B) and unobserved common effects (hypotheses A/B vs. hypothesis C).[2] The conventional solution to both of these problems is to exploit natural experiments that provide instruments for either health or SES. While this strategy of isolating exogenous variation certainly works well on paper, it is not always persuasive in practice. The main caveat is that convincing instruments are generally in short supply. As discussed by Stowasser *et al.* (2012), even the availability of instruments that are clearly exogenous and that have an impact on the endogenous regressor they seek to replace, may cause problems if the variation they reflect is not all that *relevant* for the dependent variable of interest. Moreover, since IV strategies usually rely on rather case-specific events, any uncovered effects may well be causal in nature but of questionable external validity.

For these reasons, Adams *et al.* (2003) (HWW henceforth) propose an alternative approach of uncovering causal links that makes use of the entire variation in health and economic variables. Using panel data, they test for Granger non-causality of SES for *innovations* in health, which deals with the econometric challenge of distinguishing hypotheses A and B.[3] Their purely statistical causality concept deviates from "true" causality in a structural sense, as their approach does not specifically address the issue of unobserved individual heterogeneity. As a consequence, the detection of Granger causality would not necessarily imply the validity of hypothesis A, since unobserved third factors may be at work instead. However, a finding that economic status is *not* Granger causal for health and that the relationship is invariant across a wide range of SES and health histories would be informative, as this would rule out true causality as well.[4] Applying their

[2] For a detailed discussion, see Stowasser *et al.* (2012).

[3] While HWW studied both wealth-to-health and health-to-wealth causation, this study concentrates on the question whether hypothesis A is correct.

[4] The rationale for this reasoning is that Granger causality – or conditional dependence across time – is thought of as a necessary but insufficient condition for causality in a more structural sense.

framework to a representative sample of US Americans over the age of 70, HWW are unable to reject the hypothesis that economic status has no causal effect on mortality and most health innovations, once health history is controlled for. Despite the fact that this result may not be overly surprising in light of the subgroup's quasi-universal access to Medicare and considering that causal links may well have been active in the past[5], their study stimulated much controversy in the literature, which scrutinized methodological issues as well as the external validity of HWW's findings.

On this account, Stowasser *et al.* (2012) (denoted HWWR for the remainder of this paper) revisit HWW's approach and investigate whether the original findings are confirmed when their methodology is applied to a more encompassing set of data that covers health histories of different lengths and varying age compositions.[6] In stark contrast to the original study, they find that it is much harder to reject the existence – or the activity – of causal links in more comprehensive samples. Importantly, this result is not exclusively driven by the inclusion of younger individuals, as the mere growth in sample size already leads to higher rejection rates of Granger non-causality, hinting at HWW's results being partly driven by low test power. In light of their findings, HWWR discuss three important methodological weaknesses of HWW's approach that merit attention in future research. First, the underlying notion of health dynamics, with health being modelled as a first-order Markov process, falls short of reflecting the stock characteristics of latent health capital as envisioned by Grossman (1972). Second, the failure to account for individual heterogeneity makes it impossible to distinguish between true causal links and third-factor effects in case Granger causality is detected. Third, even if common effects were convincingly controlled for, HWW's approach is only informative of the mere *presence* of causality but not of the exact mechanisms through which SES influences health. Although knowledge of this general link is important in its own right, the identification of specific pathways is equally critical from a policy perspective.

The present study aims at addressing all three of these drawbacks and gauges whether HWWR's main conclusion, that it is impossible to statistically reject SES-to-health causality even in a retired population aged 65 and older, is robust to adequate methodological refinements. The research

[5]Indeed, HWW find a steep gradient in the initial cross section, suggesting that a great deal of the relationship between health and wealth has already been determined during the (unexplained) first seven decades of respondent's life courses.

[6]It should be noted, that one of the authors of HWW, Daniel McFadden, is also part of the research team that forms HWWR.

strategy rests on the increasing availability of retrospective life-history data within large panel studies that link economic and health data, such as the US American *Health and Retirement Study (HRS)* used for this analysis.[7] These data innovations are the response to the rapidly growing literature on childhood health that makes the point that a meaningful analysis of the gradient should incorporate respondent's early-life information (for an overview, see Smith, 2009; Almond and Currie, 2011; and Currie, 2011). For instance, Case *et al.* (2002) suggest that part of the adult SES gradient in health originates in early childhood, as they find a strong relationship between parental economic status and childhood health that accumulates as children age. In another cohort study, Case *et al.* (2005) document that these early conditions have a lasting impact on adult health and – in line with hypothesis B – other outcomes such as education, labor supply and income. As Currie (2009) notes, these findings are supported by many, albeit not all, of the myriad of studies that complement the literature by exploiting data from natural experiments.

Not only does this evidence suggest the use of available information on childhood circumstances, to avoid bias from omitted variables when studying causal pathways in adulthood – the retrospective look at the beginning of life additionally has the potential to alleviate all three of the aforementioned problems of HWW's framework: First, it provides an opportunity to incorporate longer health histories and, thus, a more realistic model of health dynamics. Second, to the extent that retrospective data also covers information on family backgrounds and parental SES, it will be possible to proxy-control for some of the individual heterogeneity that is suspect of exerting a common influence on health and wealth. Third, controlling for both historic and contemporary variables may elucidate *when* the association between SES and health is established, which has important policy implications: If future outcomes are predetermined during childhood, resources spent on policies that aim at improving access to health care for adults and retirees may in fact be more wisely invested into educative and financial measures for young families.

In addition to these refinements, I pursue an alternative strategy – unrelated to the early-life literature – to provide further microfoundation of the causal pathways from SES to health. By decomposing the overall gradient into direct and indirect effects, this final part of the analysis sheds light on the question of *how* wealth, income, and education influence latent health

[7] Comparable data projects that follow the same trend include the *Survey of Health, Ageing and Retirement in Europe* (SHARE), the *English Longitudinal Study of Aging* (ELSA), or the *China Health and Retirement Longitudinal Study* (CHARLS).

capital: Does causation work its way through diverging rates of accumulation of health conditions? Are asymmetries in disease management – given the occurrence of adverse health events – the main driving force? Or, are other, non-disease specific channels responsible for the fact that general health status monotonically increases with economic well-being?

In summary, the results of this study suggest that HWWR's findings are largely insensitive to varying models of health histories. While SES is unlikely Granger causal for innovations in acute health insults, Granger non-causality can be statistically rejected for mental health conditions, mortality and changes in overall health. Evidence for chronic diseases and functional health is a bit more inconclusive. However, since the detection of Granger causality for these health conditions is adversely related to sample size, it is possible that we merely observe the statistical artifact – as already reported by HWWR – that test power suffers considerably in small datasets. The fact that results are also quite robust to the introduction of proxy controls for individual heterogeneity, lends support to a causal interpretation of the observed gradient. In line with the literature on early life circumstance, I find that childhood health has lasting predictive power for adult health. This, however, does not render contemporary factors unimportant. As to the specific pathways between SES and general health, it seems that economic status exerts its influence both via differentials in disease incidence as well as through more direct channels. Notably, there is no evidence that ample financial resources can act as a shield from the health consequences of illnesses – at least not in a post-retirement population that enjoys near-universal health care coverage. However, the prevalence of chronic conditions appears to have stronger adverse effects on overall health among the lower educated, suggesting that they have considerable difficulty in successful disease management, necessary to diminish the negative health consequences from illnesses such as diabetes or respiratory diseases.

The rest of this paper is structured as follows: Section 2.2 presents the data used for analysis. This is followed by a brief description of the methodological framework – which closely resembles that of HWW and HWWR – in section 2.3. The empirical analysis is presented in section 2.4. Section 2.5 concludes.

2.2 Data

As was the case for HWW and HWWR, the present paper uses data from the *Health and Retirement Study (HRS)*, which is a representative panel of

the US population aged 50 and older. Since the general characteristics of this survey study have already been discussed in chapter 1, this section is kept rather brief.[8] Due to substantial deviations in survey design, observations from the first panel wave are dropped. As a result, the main working sample consists of 8 biennial waves covering interviews conducted between 1993–2008. Close to the spirit of the original HWW study that focuses on the gradient in an elderly population, I restrict analysis to a mostly retired population of the age of 65 and above. On average, each wave contains roughly 11,400 individuals with usable records on health outcomes, SES variables and demographic information.[9] Attritors and members of refreshment cohorts are kept in the sample for as long as they participate in the survey. This corresponds to HWWR's "refilling" sampling scheme and ensures that sample size is kept high enough for precise estimation and that up to 8 waves can be used simultaneously.

In a notable deviation from HWWR, this study no longer estimates the incidence of 20 separate health conditions but combines some of them into disease clusters. As a result, health dimensionality is reduced to just 6 outcomes, which should considerably facilitate concise interpretability of results. The new list of health outcomes reads as follows: The number of acute – and immediately life-threatening – conditions (cancer, heart disease, and strokes), the number of chronic diseases (lung disease, diabetes, hypertension, and arthritis), the number of functional health limitations (incontinence, severe falls, hip fractures, ADL/IADL impairments, and an indicator for obesity), the number of mental illnesses (cognitive impairment, psychiatric disease, depression, and whether interviews were conducted with a proxy respondent), self-rated health status, and – finally – mortality. Summary statistics for these health indicators as well as for all SES variables used for analysis – namely wealth, income, education, dwelling condition, and neighborhood safety – are presented in appendix table 2.7.

This contemporary data is complemented with information from retrospective questionnaires on respondents' health, living conditions, and family backgrounds when they were children, that has subsequently become available within HRS. While this method of retrieving information about panel members' lives before the survey's baseline year provides advantages – in the form of low cost, speed, and reduced sample attrition – over lon-

[8] For further details on HRS, please refer to section 1.5.1.

[9] Just as in HWWR, I exclude individuals that generally failed to disclose information on their health. Gaps from insular item nonresponse are filled via simulation-based imputation. For missing wealth and income measures, I use imputations readily available in the public release files provided by the *RAND Corporation*.

gitudinal cohort studies that follow respondents from cradle to grave, one may express doubt about the accuracy of responses. After all, interviewees are asked to recall circumstances that date back at least 50 years.[10] Yet, the growing literature on the reliability of retrospective surveys, finds recall bias to be generally negligible (see Berney and Blane, 2010; and Garrouste and Paccagnella, 2010). For instance, while Smith (2009) reports some unsystematic recall error in retrospective HRS data, he finds no evidence for "coloring" – the selective recall of health histories induced by adverse health events late in life – of responses.

Retrospective information on childhood health has been introduced to HRS in two stages. A general index of self-rated health (SRH) before age 16 – which is constructed in the same way as the 5-point-scale measure for contemporary SRH – is already available since panel wave 4, hence covering a rather large share of the entire HRS population. On the other hand, effective sample sizes are considerably smaller for the multitude of detailed child-health measures introduced in wave 9, since these are only available for respondents, who were still sample members at this late stage. The latter list of variables includes 21 health conditions and whether respondents missed school for more than a month due to health problems. Once again, the individual health conditions are grouped to reduce complexity: I distinguish severe health problems (such as cancer or heart disease), less severe conditions (such as ear infections or allergies), mental health problems (such as depression or psychological problems), and classic child diseases (measles, chicken pox, and mumps).

HRS early-life data also covers the economic living conditions during childhood as well as family background measures and certain child behaviors. Again, some measures are available as early as wave 4. These include a 3-point index of self-assessed family SES, information on parental education, paternal unemployment, and whether the family ever solicitated financial help or had to move due to economic dire straits. Information on maternal labor-force participation and parental smoking were added in waves 8 and 9, respectively. In addition, starting with wave 9, HRS provides information on childhood smoking, drug and alcohol use, and whether the respondent experienced significant learning problems at school. Another pair of measures – already used by HWW and HWWR – that also capture family effects, but which are not considered part of HRS's retrospective module, are the ages at death (or just the ages, in case they are still alive) of the respondents' parents. Similarly, respondents' adult height is often used

[10]The HRS questionnaire defines childhood as life before the age of 16.

Table 2.1. *HRS early-life data*
Summary statistics

Variables	N	Mean	StDev.
Childhood health			
- Poor/fair self-rated health	25,266	0.654	0.247
- No. of severe conditions	11,624	0.243	0.526
- No. of less severe conditions	11,625	0.345	0.665
- No. of mental conditions	11,693	0.068	0.289
- No. of "child diseases"	10,565	2.228	0.982
- Missed school due to health problem	11,681	0.113	0.316
Family background			
- Self-rated family SES above average	25,389	0.066	0.249
- Self-rated family SES below average	25,389	0.317	0.465
- Family needed financial help	24,994	0.125	0.331
- Moved due to financial problems	25,246	0.180	0.384
- Father's Education (in years)	24,806	8.9	3.5
- Mother's Education (in years)	26,010	9.1	3.3
- Father ever unemployed	25,045	0.290	0.454
- Mother always worked	17,633	0.171	0.376
- Mother sometimes worked	17,633	0.327	0.469
- Any parent smoked	11,677	0.634	0.482
- Both parents smoked	11,677	0.169	0.375
- Smoked as child	15,219	0.185	0.389
- Drugs or alcohol as child	11,722	0.005	0.071
- Learning problems at school	15,218	0.027	0.162
- Father's age (at death) (in years)	29,482	71.6	14.4
- Mother's age (at death) (in years)	29,482	75.3	15.1
- Adult height (in meters)	29,482	1.69	0.10

Notes: N denotes the number of respondents for who information on the respective variable is available.

as a proxy for health at birth and is correlated with the uterine environment the family provides (see Case and Paxson, 2008; and Currie, 2011).

Summary statistics for all early-life data used for analysis are provided in table 2.1. As indicated, the number of available observations differs considerably among variables. This needs to be taken into account when deciding which of these information to use for analysis in section 2.4, as statistical power will certainly suffer in case of severe sample-size loss.

2.3 Methodology

Given that this study keeps modifications to HWW's methodological framework to a minimum and since their approach has been thoroughly discussed in chapter 1, the present section is kept fairly concise. Again, you should feel free to reconsider section 1.4 for further details.

The analysis in section 2.4 builds on the following dynamic model of health incidence:

$$f\left(HI_{it}^{j}|\mathbf{HI}_{it}^{k<j},\mathbf{H}_{it-\tau},\mathbf{S}_{it-1},\mathbf{X}_{it-1},I_i\right), \tag{2.1}$$

where i denotes the respondent and t indicates time. The dependent variable, HI_{it}^{j} measures a new incidence of a given health condition, where j stands for one of the six health clusters introduced above. As in HWW, health innovations are thought to be influenced by the following explanatory variables: Instantaneous causal effects from concurrent health shocks on HI_{it}^{j} – such as the development of cancer that is followed by death within the same inter-wave spell – are captured by $\mathbf{HI}_{it}^{k<j}$, containing the incidence variables for all health indicators $(1,...,k)$ that are causally arranged upstream of indicator j.[11] Furthermore, the model controls for health histories, $\mathbf{H}_{it-\tau}$, that capture state dependence and co-morbidities, respectively. The vector $\mathbf{X}_{it-1}$ includes demographic controls. The vector of main interest, $\mathbf{S}_{it-1}$, contains lagged levels of wealth, income, educational attainment, and indicators for subpar living environments. If SES is truly causal for health changes in an elderly population, we should expect significant coefficients for at least some of these variables. Moreover, the null hypothesis that

$$f\left(HI_{it}^{j}|\mathbf{HI}_{it}^{k<j},\mathbf{H}_{it-\tau},\mathbf{S}_{it-1},\mathbf{X}_{it-1},I_i\right)=f\left(HI_{it}^{j}|\mathbf{HI}_{it}^{k<j},\mathbf{H}_{it-\tau},\mathbf{X}_{it-1},I_i\right), \tag{2.2}$$

i.e. that past SES is not Granger causal for health deteriorations, should be rejected, while invariance tests, as described in section 1.4 , are expected to be confirmed.

Model 2.1 deviates from HWW's original specification – model 1.9 – in the following four aspects: First, health histories are no longer assumed to be first-order Markov, as τ may take on values larger than one, to better accommodate the stock characteristics of latent health capital. This part of the analysis, in which I estimate model 2.1 with alternative specifications for $\mathbf{H}_{it-\tau}$, is presented in section 2.4.1. Second, the model acknowledges the hypothetical presence of individual heterogeneity, I_i, that may induce spurious correlation between health and SES (see hypothesis C). The analysis in section 2.4.2 seeks to contain the confounding influence of such common effects by using proxy controls for family backgrounds and behavioral factors. Of main interest is whether HWWR's findings, that SES

[11] Similarly to HWW, the six health indicators are grouped in the order in which instantaneous causality is most likely to flow: Acute conditions are listed first, as they can have an immediate impact on mortality. The remaining indicators are stacked as follows: Acute conditions upstream of chronic conditions upstream of functional conditions upstream of mental conditions upstream of SRH.

Table 2.2. *Benchmark results*
Tests for Granger non-causality

Health indicator	Test results (65+) W2-9 (N=50,993)	
	F	M
Acute conditions		
Mortality	•	••
Chronic conditions	•••	•••
Functional conditions	••	•••
Mental conditions	•••	•••
Self-rated health status	•••	•••

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (•), rejected at 1% level (••), or rejected at 0.1% level (•••). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level. Blank cells indicate that Granger non-causality cannot be rejected. N denotes the number of respondent-year observations.

is Granger causal for innovations in health, even in an elderly population, survive when more realistic health dynamics and a richer set of control variables are incorporated. A confirmation of their results would lend support to a causal interpretation of the observed association.

In a third departure from HWW's model, I vary the way of controlling for SES when fitting the model for self-rated health status. Instead of merely including the most recent histories of economic variables, I additionally interact wealth, income, and education with the prevalence of health conditions. As argued in section 2.4.3, this will shed light on causal *pathways*, as it decomposes the overall influence of economic circumstances on general health status into three different effects: a direct effect on well-being, not transmitted by medical conditions; an indirect effect via higher incidence rates of diseases among the poor; and an interaction effect, which measures the additional impact of SES on health, given the detection of a health problem in the past. The last effect is of particular interest as it will provide information on the role of financial resources and educational status in limiting the negative health consequences of illnesses.

The final deviation from model 1.9 concerns the reduction in health dimensionality by grouping certain medical conditions together. As a consequence, model 2.1 is now fitted by ordered probit (except for mortality and the indicator for poor/fair SRH, which continue to be estimated with a

probit model). To ensure the innocuousness of this modeling choice and to provide a benchmark to which results from section 2.4 can be directly compared, I estimate model 2.1 with identical health histories and controls as in HWWR. Note that, since the present sample consists of individuals aged 65 and older, results in table 2.2 – which represent test results for Granger non-causality of SES, conditional on model invariance – need to be compared to the two final columns of table 1.3. Evidently, results are largely insensitive to the aggregation of health measures and mirror HWWR's finding that – with the exception of acute diseases – SES Granger causality cannot be rejected for medical events after the age of 65.[12] These results are at least significant at the 5% level, in many cases even at the 1% or 0.1% level, although model invariance across time is not always supported in a sample that spans over all 8 available panel waves.

2.4 Empirical analysis

2.4.1 Health dynamics

The notion of health being a latent capital stock that reflects the entire history of medically relevant events is not new. Ever since Grossman (1972) proposed his seminal health production framework, most health economists acknowledge the existence of "long memory effects" of the human body and mind. Heiss (2011) confirms that this feature characterizes the HRS population, too, as he detects a surprisingly high degree of state dependence in respondents' SRH: Studying the first seven panel waves, he finds that, even if the maximum number of six lags of SRH are included to predict SRH in the seventh wave, all historic variables have significant explanatory power on their own.

In light of this, and as discussed by HWWR (p. 494), HWW's treatment of health dynamics – which is viewed as a first-order Markov chain structure – will not provide an appropriate description of the evolution of health:

> Intuitively, this is because the Markov model assumes that all relevant information about the whole past is captured in the observed variables one

[12] While the aggregation of health conditions appears to occult HWWR's weak evidence of Granger causality for heart problems and strokes among women, robustness checks – not reported here due to space considerations – which repeat the analysis of section 2.4 with all 20 separate health conditions, suggest that this result was rather unstable to begin with and quickly disappears once modifications along the lines described above are implemented. In fact, this robustness analysis – available upon request – asserts that the findings of this study are not driven by the decision to use health-condition clusters.

> period ago. This is unrealistic since knowledge of longer histories would better capture the stock characteristics of health capital [...]. Taking functional limitations as an example, a respondent who reported difficulties with walking one year ago and no limitations previously has a different outlook than a respondent who consistently reported difficulties with walking for the last ten years.

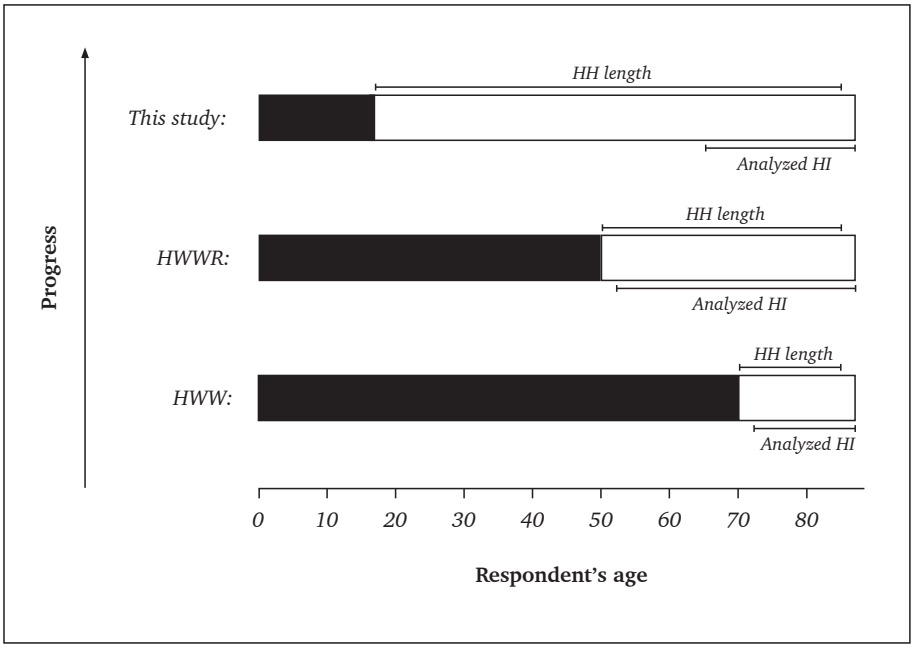
A straightforward way to improve HWW's model of health dynamics consists of increasing the length of health histories, model 2.1 controls for. While the performance of higher-order Markov models probably falls short of that of a fully-fledged *hidden* Markov model, such as Heiss (2011), they will likely pick up many of the same effects. More importantly, however, there are practical limits to this strategy: The more lags of health conditions are incorporated, the smaller the effective sample size that remains for analysis. On the one hand, it excludes all respondents that have been part of the sample for fewer waves than required by the desired history length. This may affect both sample attritors and members of refreshment cohorts, meant to keep the panel representative of the underlying population. On the other hand, the sample would even shrink if the panel was completely balanced, as each additional lag of control variables requires to drop one wave for the estimation of health innovations conditional on health histories.

As mentioned in section 2.1, such large drops in sample size constitute a severe problem for HWW's approach because of its susceptibility of being unable to reject Granger non-causality if test power drops too low. Given this apparent trade-off between richer health dynamics and the risk to obtain artifactual test results, the number of lags should only be increased with great care. On this account, the knowledge of health during childhood provides a promising alternative to control for even (much) longer histories without having to forego the potential scale limits in the data.

At the same time, the availability of information on child health alleviates the closely related problem of initial conditions – that is, life before respondents enter the panel (see Heckman, 1981a). As Smith (2009) (p. 388) notes,

> [k]nowing health or economic status beginning at [survey] baseline is not sufficient because the entire prior histories of health and economic trajectories may matter for current decision making. The absence of information on pre-baseline health histories, including childhood health, means that researchers have to rely on a key untestable assumption: baseline health conditions sufficiently summarize individuals' health histories. If they do

Figure 2.1. *Maximum health history lengths*
Comparison between studies



Notes: White boxes indicate known health histories. Black boxes depict unknown health histories. "HH length" denotes the maximum length of health histories that can be exploited for analysis. "Analyzed HI" stands for the age range used to analyze health incidence. HWW stands for the study by Adams *et al.* (2003). HWWR denotes Stowasser *et al.* (2012).

not, new health events unfolding during the panel may be the delayed (and perhaps predictable) consequence of some knowable part of an individual's health history. If so, health events within the panel cannot be used to measure effects of new exogenous, unanticipated events.

The extent to which retrospective data enables a look into the "black box" of early life, as compared to HWW and HWWR, is visualized in figure 2.1. Note that the effective health history length is depicted to be by one wave (or 2 years) shorter than panel length theoretically permits.

Given these considerations, I gauge the sensitivity of model 2.1 to varying representations of health history by gradually increasing the lag length of adult health prevalence, by the inclusion of child health, and by combinations of the two. As argued above and in section 2.2, these steps are associated with considerable reductions in effective sample size, which entails the risk of confounding any effect from longer health histories with the mere decline in test power that plagued the analysis of HWW. In order to separate these two effects, I also apply HWW's unaltered health history specification to these subsamples. These "dry runs" serve as the

benchmarks, to which results from models with more sophisticated health histories should be compared. The Granger non-causality test results for all of these specification are summarized in table 2.3.

The first alternative specification, models health histories as a second-order Markov process (i.e. the number of health-condition lags is increased to two), which reduces the size of the analyzable sample from 50,993 to 42,367 respondent-year observations. As is evident from comparing columns (C) and (D) with columns (A) and (B) of panel A in table 2.3, this has no significant impact on SES Granger causality tests. The same picture emerges when a third-order Markov model is used (see columns (G) and (H)). While with the latter specification, empirical p-values tend to be a bit higher than with the lower-ordered Markov model (as indicated by fewer dots), this is clearly not driven by the inclusion of the additional lag but by the reduction in sample size. To see this, consider that p-values also increase for the benchmark case – compare columns (E) and (F) with columns (A) and (B) – whereas the actual switch to a higher-order Markov model – compare columns (G) and (H) to columns (E) and (F) – has no systematic impact at all. Results for even higher-order Markov models are not presented here, as these imply sample sizes too low to conduct meaningful analysis that stratifies by gender.

Panel B of table 2.3 contains results for specifications that use child health to incorporate longer health histories. Recall from section 2.2 that the number of respondents with data on childhood SRH greatly exceeds that of individuals for who we have detailed information on early-life health conditions. For this reason, I add these variables in two sequential steps. Results in columns (L) and (M) are for model 2.1 when controlling for first-order Markov health histories – HWW's default – and self-rated health during childhood. Once again, Granger non-causality tests are not systematically influenced by the incorporation of longer health histories and suggest that, with the exception of acute diseases, causal links from SES to health cannot be statistically rejected. In the second step, I additionally include the more specific data on childhood health conditions, which roughly cuts the available sample size in half (49,962 to 25,175 respondent-year observations). The corresponding results in columns (P) and (Q) require some discussion: First of all, the effect of SES on mortality can no longer be studied because information on childhood health conditions is only available for respondents who were still alive in wave 9, which happens to be the most recent wave in the working sample. Furthermore, while it is true that Granger causality of SES is no longer supported for functional health conditions among women, this seems, once again, to be driven by the sub-

Table 2.3. *Results for varying health histories*
Tests for Granger non-causality

Panel A: Higher-order Markov models								
Health indicator	**Test results**							
	Sample for 2 lags (N=42,367)				Sample for 3 lags (N=38,886)			
	HWW		2 lags		HWW		3 lags	
	F	M	F	M	F	M	F	M
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Acute conditions								
Mortality	●●●	●●●	●●	●●●	●●	●●	●●	●●
Chronic conditions	●●●	●●	●●●	●●	●●●	●	●●	●
Funtional conditions	●●	●●●	●●●	●●●	●●	●●●	●●●	●●●
Mental conditions	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Self-rated health	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Panel B: Childhood health								
Health indicator	**Test results**							
	Sample for SRH (N=49,962)				Sample for Conditions (N=25,175)			
	HWW		SRH		HWW		HC	
	F	M	F	M	F	M	F	M
	(J)	(K)	(L)	(M)	(N)	(O)	(P)	(Q)
Acute conditions								
Mortality	●●	●●●	●●●	●●●	n.a.	n.a.	n.a.	n.a.
Chronic conditions	●●●	●●●	●●●	●●●	●●	●●●	●●●	●●●
Funtional conditions	●●	●●●	●●	●●●		●●		●
Mental conditions	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Self-rated health	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Panel C: Third-order Markov model and Childhood health								
Health indicator	**Test results**							
	Sample for SRH (N=34,136)				Sample for Conditions (N=19,527)			
	HWW		3L & SRH		HWW		3L & HC	
	F	M	F	M	F	M	F	M
	(R)	(S)	(T)	(U)	(V)	(W)	(X)	(Y)
Acute conditions								
Mortality	●●	●●	●●	●●	n.a.	n.a.	n.a.	n.a.
Chronic conditions	●●●	●	●	●	●	●		●
Funtional conditions	●●	●●●	●●●	●●●		●●●		●●
Mental conditions	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Self-rated health	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (●), rejected at 1% level (●●), or rejected at 0.1% level (●●●). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level. Blank cells indicate that Granger non-causality cannot be rejected. N denotes the number of respondent-year obsservations. SRH stands for self-rated health during childhood. HC denotes childhood health conditions. 3L abbreviates 3 lags. "n.a." indicates that lack in variation impedes estimation of mortality models.

stantial reduction in sample size. Also note that, while the change in results for functional conditions among men (when comparing columns (O) and (Q)) seems substantial at first sight, a look at the actual p-values reveals that the change – from 0.0089 to 0.0104 – is only marginal at best.

For results in panel C of table 2.3, I combine both ways of accommodating health histories, which should arguably provide the most comprehensive description of the long memory effects of latent health capital– although this comes at the cost of even greater sample-size loss. Test outcomes in columns (T) and (U) are from a model with third-order Markov health histories and childhood SRH. This specification is then amended with the data on childhood health conditions (see columns (X) and (Y)). Overall, test outcomes depicted here, corroborate the findings from panels A and B. If anything, evidence for SES being Granger causal for the development of chronic conditions becomes a little weaker, as the null hypothesis of non-causality is only rejected at the 5% level for men and the 5% to 15% level for women (the corresponding p-value in column (X) equals 0.141). Similarly, results for functional conditions among females do again become barely insignificant (p=0.120). While it is certainly possible to dismiss these observations as artifactual side effects of dwindling sample sizes, one should at least note that results are generally a bit less stable for chronic and functional conditions than for mental health and SRH.

Finally, a look at the coefficients of the underlying prediction model 2.1 – not reported here – confirms the earlier finding in the literature that even very long health histories have explanatory power for health innovations in an elderly population. For instance, Heiss' (2011) observation, that all lags of SRH have highly significant predictive power for current SRH, is confirmed even when controlling for SES and third-order Markov health-condition histories. The fact that the same holds true for SRH during childhood, hints at an astounding degree of state-dependence in latent health. In addition, to have been in subpar health as a child, significantly increases the risk of developing a chronic condition among women over the age of 65. Interestingly, the addition of childhood health conditions renders coefficients of childhood SES insignificant. Now, it is the number of severe health problems during childhood, which has explanatory power for innovations in acute, chronic, functional and mental health conditions. The number of less severe conditions only matters for the development of chronic conditions among men. Mental problems as a child are highly predictive of mental illness in adulthood as well as of functional health problems among females. As expected, the number of classic child diseases has no explanatory power for any future health outcomes. All of this evidence confirms

the long reach of childhood circumstances, established by the literature summarized in section 2.1. It is encouraging to observe that test results for Granger-non causality of SES are not significantly changed by accounting for these formerly omitted variables.

2.4.2 Common effects

As argued in section 2.1, HWW's Granger-causality framework cannot cleanly distinguish between hypotheses A and C – that is, between "true" causality and spurious correlation due to common effects. This identification problem arises because of unobserved individual heterogeneity – with respect to genetic endowment, family backgrounds, and early-life experiences – that influences both health and SES without there necessarily being a causal relationship between the two. Methodological solutions to this problem either require a set of valid instruments or the use of fixed-effects approaches. Since HWW, HWWR, and the present paper study whether HWW's framework can serve as a viable *alternative* to IV estimation, it would not make much sense to go down the first-mentioned route. Furthermore, while the HRS panel is certainly of sufficient length to estimate equations with individual fixed effects, it is not obvious that such models, which rely on the assumption that coefficients are constant over time, make sense when looking at health and wealth over a period spanning several decades.

For these reasons, this study follows a different strategy, which may well fall short of providing an outright solution to the problem, but which should alleviate the confounding influence of unobserved third factors. Acknowledging the fact that the underlying problem is one of omitted variables – namely unobserved individual heterogeneity – I add control variables that should provide reasonable proxies for characteristics of the family and the home environment, as the latter are likely to play a central role in shaping individual preferences, behaviors and genetic endowment. Naturally, the feasibility of this approach critically hinges on the data at hand. As extensively argued in the childhood-health literature, early-life data provides a number of variables that meet the above requirement (see – among several others – Case *et al.*, 2002; Berger *et al.*, 2005; Case *et al.*, 2005; Smith *et al.*, 2010; and Mazzonna, 2011). For instance, Case *et al.* (2005) (p.384)

> [...] include a large set of variables in [the control vector] C, and assume that this set of variables is rich enough to capture all individual heterogeneity. Indeed, our ability to control for a large set of childhood characteristics is an advantage over much of the previous literature that examines health and SES dynamics.

Table 2.4. *Results for varying family-background controls*
Tests for Granger non-causality

Health indicator	Test results							
	Sample for tier 1 (N=42,271)				Sample for tier 2 (N=21,250)			
	HWW		Tier 1		HWW		Tier 2	
	F	M	F	M	F	M	F	M
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Acute conditions								
Mortality	●●	●●●	●●	●●	n.a.	n.a.	n.a.	n.a.
Chronic conditions	●●●	●●	●●●	●●	●●●	●●●	●●	●●
Funtional conditions	●●	●●●	●	●●		●●		●
Mental conditions	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●
Self-rated health	●●●	●●●	●●●	●●●	●●●	●●●	●●●	●●●

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (●), rejected at 1% level (●●), or rejected at 0.1% level (●●●). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level. Blank cells indicate that Granger non-causality cannot be rejected. N denotes the number of respondent-year obsservations. For definitions of tier 1 and tier 2 see text. "n.a." indicates that lack in variation impedes estimation of mortality models.

The 15 family-background variables used to proxy-control for individual effects are listed in table 2.1. As was the case for childhood health conditions, the number of available observations differs substantially among variables, which is why they are also added in two sequential steps. The first tier of controls includes the four proxies for family SES, parental education, paternal unemployment status, parental age (of death) and respondents' adult height. The second tier consists of the aforementioned data to which maternal labor-force status, parental and own smoking behavior as a child, drug use, and information on learning problems in school are added. Again, I estimate benchmark dry-runs like those described in section 2.4.2 to distinguish the effects of adding the controls from those that are due to reductions in sample size. Results for Granger non-causality tests, conditional on model invariance, are summarized in table 2.4.

While p-values slightly increase across the board by the inclusion of both tier 1 and tier 2 variables, the changes in test results are not very substantial. Overall, the conclusion that Granger-non causality is statistically rejected for non-acute health events remains intact even after controlling for family backgrounds. The notable exception is functional health, for which results are a bit inconclusive. This underscores the earlier finding that the association between SES and this health dimension appears to be weaker than for other conditions.

Table 2.5. *Results for all controls*
Tests for Granger non-causality

Health indicator	Test results (N=16,335)			
	HWW		All controls	
	F	M	F	M
	(A)	(B)	(C)	(D)
Acute conditions				
Mortality	n.a.	n.a.	n.a.	n.a.
Chronic conditions	•	•		
Funtional conditions		••		•
Mental conditions	•••	•••	••	•••
Self-rated health	•••	•••	•••	•••

Notes: Results are for white females (F) and males (M). Abbreviations are as follows: Granger non-causality rejected at 5% level (•), rejected at 1% level (••), or rejected at 0.1% level (•••). Gray symbols indicate that the corresponding invariance test is rejected at the 5% level. Blank cells indicate that Granger non-causality cannot be rejected. N denotes the number of respondent-year obsservations. "n.a." indicates that lack in variation impedes estimation of mortality models.

In a final step, I estimate a version of model 2.1 that combines controls for family backgrounds with a more adequate model of health dynamics as developed in section 2.4.1. Note that, inasmuch as these longer histories capture the effect of latent health capital, they may also absorb some of the endogeneity imposed by genetic traits, with severe health problems in childhood being a signal for general frailty. To achieve the most conservative assessment for the presence of Granger causality, I model health histories as third-order Markov with controls for all available childhood health conditions and include the more encompassing second tier of early-life controls. Results are presented in table 2.5 and should be compared to columns (X) and (Y) of table 2.3 and columns (G) and (H) of table 2.4. Even in this most encompassing specification – that comes at the cost of an even smaller and less representative sample of just 16,335 respondent-year observations – SES Granger non-causality for mental health conditions and general health status is clearly rejected, which lends credibility to the interpretation that these associations do in fact reflect causal relationships. While results for chronic and functional health conditions are certainly less robust, it is not entirely clear how much of the increase in p-values is driven by the introduction of controls – which would in fact suggest the importance of third factors – and how much is due to dwindling test power that may occult the presence of true, albeit relatively weaker, causal links. A conclusive answer

to this question will have to wait for the addition of refreshment cohorts, which will eventually increase the number of available observations for early-life conditions as well.

2.4.3 Causal pathways

So far, the focus of this study has been the ability of HWW's approach to discriminate between true causality and the influence of third factors, in case Granger causality is detected. While this general distinction is certainly of interest in its own right, it is equally important to go beyond broad causality tests and investigate more narrowly focused questions about the *mechanisms* that connect specific health outcomes to specific dimensions in SES. This is especially true from the perspective of policy makers in search for economic policy levers that can improve the health of the general public. Yet, as discussed by HWWR (p. 512), one of the major downsides of HWW's approach lies exactly in the limited discrimination among causal pathways:

> Even if we were able to univocally confirm the presence of causal effects from wealth to health, we still would not know the channels through which they operate. Yet, the latter information is absolutely critical from a policy perspective: interventions to increase the affordability of health insurance would be warranted if channel A1 [affordability of care] were to be active, but would prove ineffective if the causal link were to work through, say, channel A3 (health knowledge) instead. To address this issue, we intend to specify and test more differentiated hypotheses that may facilitate the discrimination among these channels [in future research].

In this spirit, the present section explores the relative importance of pathways suggested in the literature. For this, I deconstruct the general gradient to study *how* wealth, income and education leave their mark on the general health of individuals. Specifically, I vary the way of controlling for SES when estimating model 2.1 for SRH: Instead of only including the most recent histories of economic variables, I additionally interact wealth, income, and education with the prevalence of health conditions. This allows the decomposition of the overall influence of SES into the following three effects:

Direct effect. This is the effect on well-being not transmitted by medical conditions. One possible pathway that could account for such an influence of SES is channel A2[13] – the psychological burden of being

[13]Note that the classification of causal channels used here, follows the nomenclature proposed by HWWR. For their discussion of potential pathways see section 1.2.1.

poor – where constant economic distress may cause general discomfort or frailty even in the absence of measurable disease. Similarly, low SES may be causal for health conditions not used for analysis – such as migraines or back pain – which would, however, represent an instance of omitted variables rather than a true direct effect.

Indirect effect. This is the effect on general health status that is due to the accumulation of health conditions. Here, low SES may be causal for the development of certain illnesses, for instance via channel A3 – environmental hazards – where the proximity of low-income housing to smokestack factories may induce lung disease among the poor. The severity of the gradient that works through this link also depends on the importance of the respective condition for overall health, which will be larger for, say, cancer than for hypertension.

Interaction effect. This effect measures the additional impact of SES on health, given the presence of a health problem. The effect is of particular interest as it provides information on the role of financial resources and educational status in limiting the negative health consequences of illnesses. Potential pathways, whose plausibility can be put to test, include channel A1 – the affordability of health care – where certain treatments may be too costly for the financially constrained as well as channels A4 and A5 – health knowledge and risk behaviors, respectively – that stress the prominent role of educational attainment in adhering to certain lifestyles or regimens if medically demanded.

Probit regression results presented in table 2.6 come from estimating model 2.1 with interactions between health-condition indexes and wealth, income, and education years, respectively. To make use of maximum sample size, I do not stratify by gender and use HWW's benchmark specification without longer health histories and controls for family effects. Results, not presented here, are qualitatively similar but less precise when the more sophisticated specifications are used. Boldfaced numbers indicate significance at the 1% level. As is evident from the first six rows, all SES measures enter the model with intuitive signs – note that the dependent variable is an indicator for poor or fair SRH – and are highly statistically significant. This suggests that direct effects of SES that are not transmitted through medical events play an important role in determining health. A similar statement can be made about indirect effects, since – unsurprisingly – the presence of any health problem increases the probability to rate ones health as subpar. Given that SES is (at least Granger) causal for mental conditions and

Table 2.6. *Results for interaction model*
Dependent variable: Indicator for poor/fair SRH

Explanatory variables	**Probit regression**	
	Coefficients	(SE)
SES variables		
- Low wealth	**0,2161**	**(0.0434)**
- Low income	**0,1864**	**(0.0375)**
- High school	**-0,1596**	**(0.0224)**
- College	**-0,0925**	**(0.0224)**
- Poor neighborhood	**0,1060**	**(0.0268)**
- Poor dwelling	**0,2359**	**(0.0241)**
Health conditions		
- Acute	**0,1188**	**(0.0428)**
- Chronic	**0,2998**	**(0.0325)**
- Functional	**0,0931**	**(0.0261)**
- Mental	**0,0953**	**(0.0336)**
Interactions		
- LowWealth∗Acute	**-0,0821**	**(0.0273)**
- LowWealth∗Chronic	0,0231	(0.0217)
- LowWealth∗Functional	-0,0295	(0.0167)
- LowWealth∗Mental	-0,0417	(0.0222)
- LowIncome∗Acute	-0,0291	(0.0245)
- LowIncome∗Chronic	-0,0147	(0.0198)
- LowIncome∗Functional	0,0103	(0.0155)
- LowIncome∗Mental	0,0435	(0.0209)
- EducationYrs∗Acute	0,0019	(0.0033)
- EducationYrs∗Chronic	**-0,0092**	**(0.0025)**
- EducationYrs∗Functional	-0,0027	(0.0020)
- EducationYrs∗Mental	0,0030	(0.0026)
N	50,993	
LL	-21,729.2	

Notes: Results are for the HWW benchmark specification. All SES variables are defined as dummies. Health conditions are count variables. Standard errors are in brackets. Boldfaced numbers indicate statistical significance at the 1% level.

also likely for chronic and functional health problems, this establishes an additional link between health and wealth.

Turning to the interaction effects, there is no evidence that financial resources alleviate the health consequences of diseases, once they have been diagnosed. This clearly speaks against the importance of channel A1, at least in a population – like the one under consideration – that is Medicare-eligible. In fact, the only statistically significant point estimate for an interaction with a monetary SES variable has the opposite sign than theory would predict. According to this result, being poor (as measured by being in the bottom quartile of the wealth distribution) has a positive effect on SRH, conditional on having developed an acute health condition in the past. However, this anomalous result is likely due to a level effect: Given

that respondents at the very bottom of the wealth distribution tend to be already in poor health, the diagnosis of an acute disease will not have any measurable additional impact on self-assessed health – which cannot be rated worse than poor. This interpretation is confirmed by estimations, not reported here, that stratify by wealth quartile: While the diagnosis of an acute condition has a strong positive impact on reporting a deterioration in health status among the top three quarters of the wealth distribution, the same effect is much smaller and insignificantly different from zero among the bottom quartile.[14]

While money does not seem to make a difference in alleviating the negative health consequences from diseases, there is some evidence that another important SES marker does: Conditional on suffering from a chronic illness, those with higher educational attainment report to be in better health than individuals with fewer years of completed schooling. A possible interpretation for this finding is that the latter group may have considerable difficulty in disease management that often requires abrupt changes in lifestyle and the adherence to complex treatments necessary to contain the adverse health repercussions associated with illnesses, such as emphysema or diabetes. Hence, this finding provides suggestive evidence for the relative importance of channels A4 and A5. Naturally, it is unclear whether the uncovered role of education is causal, in the sense that the extra years spent at college are responsible for changes in health knowledge and preferences, or if education merely serves as a marker for unobserved – and immutable – cognitive ability. To which degree personal traits and decision making about health behaviors are malleable during this pivotal gateway period between childhood and adulthood, is currently under study in the literature (see Conti *et al.*, 2010).

2.5 Conclusion

This study addresses three main methodological weaknesses of HWW's approach of inferring causality in the health-wealth nexus. The analysis is facilitated by the availability of retrospective data on early-life events that allows for improved control of initial conditions and individual heterogeneity. The first shortcoming – the inadequate description of health dynamics – is addressed by implementing higher-order Markov models and by controlling for information on childhood health to accommodate the long memory

[14]The point estimate for an acute condition equals 0.1294 (t=11.60) in the wealthier sample. Among those in the bottom quartile, the estimate equals 0.0342 (t=1.48).

effects of latent health capital. In line with the literature on early-life circumstance, I find that childhood health has lasting predictive power for adult health. This, however, does not render contemporary factors unimportant. The results of this analysis suggest that conclusions are largely insensitive to varying models of health histories. Furthermore, they confirm the findings by HWWR that SES is unlikely to be causal for the development of acute health conditions but that Granger non-causality can – even in an elderly population aged 65 and older – be statistically rejected for mental health conditions, mortality and changes in overall health. Evidence for chronic diseases and functional health is somewhat inconclusive. However, since the detection of Granger causality for these health conditions is adversely related to sample size, this may simply reflect HWWR's earlier finding that HWW's Granger-causality tests require relatively large sample sizes to obtain adequate power.

The second methodological drawback – the inability to distinguish between true causal links and common effects in case Granger causality is detected – is alleviated by conditioning on early-life events that may function as proxies for unobserved individual heterogeneity, with health problems in childhood being a signal for physical frailty, and parental SES and health-relevant behaviors capturing family effects. Results from this modification closely mirror those of accounting for longer health histories. The fact that results for mental health and overall health status are remarkably robust, lends support to a causal interpretation of the observed gradient for these health dimensions. Ultimately, however, the assessment of this issue will depend on how narrow one wishes to define "true" causality. In my opinion, it is fair to argue that SES may even have a causal effect – in a rather wide sense – on individual heterogeneity, rendering the distinction between HWWR's hypotheses A and C almost arbitrary. In fact, there is increasing evidence that personal characteristics are not as immutable as was once believed. For instance, part of the literature on the education-health gradient argues that the years spent in education may not only change health-relevant knowledge, but also preferences, behaviors, and the way people think about their future (see Cutler and Lleras-Muney, 2006). In a similar vein, Currie (2011) reports evidence that even the activation of genetic traits – once considered the holy grail of irrevocability – may depend on environmental factors as well.

The third weakness of HWW's approach – the lacking microfoundation of pathways between health and SES – is addressed by decomposing the overall influence of economic variables into direct effects, indirect effects that work through the accumulation of medical events, and an interaction

effect that measures the additional impact of SES, conditional of having been diagnosed with an illness. The evidence suggests that economic status exerts its influence both via differentials in disease incidence as well as through more direct channels. Notably, there is no evidence that ample financial resources can act as a shield from the health consequences of diseases – at least not in a post-retirement population that enjoys near-universal health care coverage. However, the prevalence of chronic conditions appears to have stronger adverse effects on overall health among the lower educated, suggesting that they have considerable difficulty in successful disease management necessary to diminish the negative health consequences from illnesses such as diabetes or respiratory diseases.

2.A Appendix: Additional tables

The following table contains summary statistics for the dataset used in my analysis. Due to its large dimensions, it is displayed on the next page.

Table 2.7. *Contemporary health and SES Variables used for analysis*
Summary statistics

Variable	Wave 2 (N=8,726)		Wave 3 (N=9,258)		Wave 4 (N=11,916)		Wave 5 (N=11,953)		Wave 6 (N=12,273)		Wave 7 (N=12,153)		Wave 8 (N=12,502)		Wave 9 (N=12,468)	
	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.	Mean	StDev.
Health prevalence																
No. of acute conditions	0.541	0.685	0.659	0.743	0.643	0.743	0.658	0.746	0.671	0.753	0.698	0.767	0.710	0.776	0.721	0.779
No. of chronic conditions	1.004	0.880	1.086	0.905	1.089	0.908	1.121	0.913	1.173	0.921	1.229	0.926	1.285	0.929	1.341	0.938
No. of functional conditions	1.316	1.120	1.535	1.264	1.562	1.278	1.654	1.299	1.171	1.303	1.768	1.311	1.851	1.316	1.905	1.308
No. of mental conditions	0.587	0.874	0.635	0.900	0.601	0.897	0.628	0.921	0.642	0.924	0.619	0.909	0.589	0.874	0.580	0.866
Poor/fair self-rated health	0.358	0.479	0.341	0.474	0.373	0.484	0.328	0.469	0.327	0.469	0.335	0.472	0.334	0.472	0.329	0.470
Health incidence																
No. of acute conditions			0.244	0.491	0.277	0.522	0.219	0.471	0.236	0.486	0.219	0.471	0.218	0.467	0.221	0.475
Died since last wave			0.104	0.306	0.101	0.301	0.108	0.311	0.115	0.319	0.098	0.297	0.100	0.300	0.101	0.301
No. of chronic conditions			0.198	0.453	0.206	0.441	0.205	0.444	0.199	0.436	0.227	0.458	0.224	0.461	0.215	0.447
No. of functional conditions			0.843	1.072	0.874	1.077	0.798	1.018	0.830	1.038	0.836	1.042	0.931	1.068	0.908	1.040
No. of mental conditions			0.260	0.561	0.236	0.530	0.199	0.488	0.205	0.498	0.169	0.447	0.163	0.432	0.156	0.425
SES variables																
1st quartile wealth indicator	0.255	0.436	0.230	0.421	0.227	0.419	0.219	0.413	0.214	0.410	0.215	0.411	0.217	0.412	0.212	0.409
4th quartile wealth indicator	0.183	0.387	0.221	0.415	0.255	0.436	0.275	0.447	0.291	0.454	0.306	0.461	0.330	0.470	0.328	0.470
1st quartile income indicator	0.368	0.482	0.335	0.472	0.291	0.454	0.292	0.455	0.277	0.447	0.278	0.448	0.272	0.445	0.274	0.446
4th quartile income indicator	0.117	0.321	0.144	0.351	0.161	0.368	0.167	0.373	0.169	0.375	0.176	0.381	0.177	0.382	0.193	0.394
Poor/fair housing condition	0.133	0.340	0.128	0.334	0.114	0.318	0.106	0.308	0.097	0.296	0.116	0.320	0.108	0.310	0.104	0.305
Poor/fair neighborhood safety	0.145	0.325	0.132	0.338	0.101	0.302	0.089	0.284	0.075	0.263	0.086	0.281	0.096	0.294	0.094	0.291
High school (educ. > 10 y.)	0.613	0.487	0.629	0.483	0.672	0.469	0.692	0.462	0.716	0.451	0.735	0.441	0.754	0.431	0.766	0.423
College (educ. > 14 y.)	0.147	0.354	0.156	0.363	0.172	0.378	0.180	0.385	0.193	0.195	0.199	0.399	0.206	0.404	0.218	0.413

Notes: Summary statistics are for the age-eligible sample (65+).

Chapter 3

Electoral cycles in savings bank lending

3.1 Introduction

Politicians regularly rank highly in polls of people's most mistrusted professions, ahead of usual suspects such as car salesmen or estate agents.[1] Arguably, the most important reason for this unflattering ranking is, according to modern political economics, that politicians are not necessarily interested in what is best for their constituency but rather in what they can get out of it for themselves. This may take on drastic forms such as outright corruption and illegal enrichment or more subtle ones like a preference for policies that benefit re-election prospects but are socially suboptimal. A prime example for the latter is the theory of (opportunistic) political business cycles (PBC) pioneered by Nordhaus (1975), and MacRae (1977), which describes politicians' incentives to enact expansionary fiscal policies shortly before elections to boost their own popularity, only to countermand them with contractionary policies afterwards. While this theory has been empirically put to test in numerous studies (e.g. Alesina *et al.*, 1997; Mitchell and Willett, 2006; and Schneider, 2010), the literature so far has concentrated on the behavior of politicians on the federal or state level. As a consequence, only little is known about electoral cycles on the municipal level, even though local office holders should have incentives to strategically affect the regional economy as well.

Though local politicians are not known for having control over macro policies like stimulus packages, the federal tax code, or interest rates, it may not take large-scale measures to have a significant impact on a local economy that is of small scale itself. It probably is enough to invest in long overdue road maintenance, to inaugurate a recreational park, or to vigor-

[1] To name just a few of these recurring surveys, consider the "Veracity Index" by *Ipsos*, *YouGov's* "KRC poll" and the "Voice of the People" study conducted by *Gallup International*.

ously – and, more importantly, visibly – fight for the preservation of jobs at a local business in dire straits, to shift the mood of the electorate in one's favor. Moreover, the literature on political connectedness (PC) makes the point that the ability to manipulate economic outcomes may not depend on statutory authority alone, but also on social relationships that create a more informal opportunity to achieve one's goals. Most contributions to the PC literature such as Fisman (2001), Johnson and Mitton (2003), Ramalho (2004), Sapienza (2004), Faccio (2006), and Faccio *et al.* (2006) focus on the advantages for *firms* from entertaining close ties with politicians. Yet, the reverse direction of these relationships is being documented by academic work as well, as evidenced by Shleifer and Vishny (1994) who develop the quid-pro-quo argument of favors being traded between (state-owned) corporations and the political sector.

A key takeaway from the PC literature is that the degree to which political connections can affect real outcomes greatly depends on a country's institutional environment: in general, the weaker the institutions, the easier it will be for political and business elites to form networks that allow for rent extraction. On this account, it is not surprising that most of the empirical evidence of harmful political ties comes from developing or emerging countries. A notable exception is the study by Bertrand *et al.* (2007) who find that French firms destroy less jobs in election years, when they are run by CEOs who are politically or socially connected with the head of the respective local government. They argue that the seeds for these networks are planted during higher education, as the vast majority of business and political leaders attend the same very limited set of elite schools, the so-called *Grandes Ecoles*. This example demonstrates that even in highly developed countries, politicians may find ways to extend their influence beyond the statutory powers that come with holding office – as long as the institutional setup gives them the opportunity to do so.

In Germany, local politicians have such an opportunity, opened up by an institutional peculiarity in the German savings bank system: Not only is the board of directors of each regional savings bank staffed with a number of politicians from said region. On top of that, the board's chairman is usually the elected official of that respective community. This, taken together with the vital role savings banks play in local lending to private households and small to medium-sized enterprizes (SMEs)[2], provides an alluring opening for municipal politicians to improve their electoral outlook. The main research hypothesis we put to test is that incumbents may

[2] See Section 3.2.2.

try to boost economic conditions – and voter satisfaction – by pushing for more lavish lending policies within the bank's board of directors, shortly before elections. We also expect negative fallouts of this policy intervention, taking the form of higher credit default rates, as some of these loans should probably not have been granted in the first place. Moreover, in order to be sure to not measure any general booms around elections, we expect neutral or overly prudent lending policies immediately after elections. Our identification strategy relies on the fact that we should only observe politically motivated lending around election years, only in municipalities in which elections are held at this point in time, and – importantly – only for politically connected savings banks. Econometrically, we conduct difference-in-difference (DD) as well as triple-difference (DDD) estimation embedded in a fixed-effects panel data setup.

To test our hypotheses we use a novel, largely hand-collected dataset that combines detailed information on German municipal elections, macroeconomic data on the district level, and balance-sheet information on bank lending. We find that savings banks systematically extend more credits in pre-election periods. This effect is robust to various specifications. Importantly, the effect is not present with cooperative banks that are very similar to savings banks, except for their lack of political connectedness. Our hypotheses are further strengthened by the fact that lending cycles only occur around municipal but not around state or federal elections. We find weak evidence for overly prudent lending policies after elections, consistent with a binding credit constraint that banks face so that they have to make up for excessive pre-election lending. Our evidence also suggests that the ability to induce political lending cycles depends on the degree of dominance of the incumbent party. Finally, we find significant differences in effect sizes across states, suggesting an interaction with other details of the institutional environment that we intend to further investigate in future research.

Our results complement the finance literature, which has long suspected that the behavior of government-controlled banks is rather different from that of private financial institutions. Caprio Jr. and Martinez Peria (2000), Barth *et al.* (2001), and La Porta *et al.* (2002) find that politicians use public banks to further their own political objectives and that government ownership in banks is associated with increased risk of banking crises, reduced financial development, and subpar economic growth, respectively. In related work, Khwaja and Mian (2005) find that politically connected firms in Pakistan have easier access to credit but that this preferential treatment is only granted by government banks. Once again, however, the evidence so far is limited to case-studies in the developing world. In fact, Dinç (2005)

studies in a framework similar to ours, whether the lending behavior of public banks depends on the timing of elections, and finds that – as opposed to emerging markets – no such wrongdoing seems to be occurring in developed economies. One potential reason for the discrepancy between our results and those of Dinç is likely to be found in our decision to concentrate on municipal (instead of general) elections, reflecting that in the German case, political connections are established on the local and not the federal level.

The remainder of this paper is organized as follows: The institutional background that facilitates the creation of connections between local politicians and savings banks is described in section 3.2 and followed by section 3.3, in which we specify our research hypotheses and predictions. Section 3.4 describes the data we use and discusses the problems that arise from having to rely on balance-sheet information. Methodological issues and our identification strategy are presented in section 3.5. Section 3.6 contains the empirical results while section 3.7 is reserved for robustness analysis. Section 3.8 concludes.

3.2 Institutional background

In this section we provide the institutional details relevant for understanding the German public banking sector. In doing so, we lay out the case why savings banks are a prime example for politically connected firms, how cooperative banks are an ideal control group and how the German electoral rules allow us to cleanly estimate causal effects of elections on banks' lending behavior.

3.2.1 Public guarantee obligation

German law installs public guarantee obligation (*Gewährträgerhaftung*) for public institutions. This rule provides that the creditor is going to be reimbursed by the government in case the public institution is not able to live up to its contractual obligations. German savings banks have been founded by the respective municipalities (see below), were considered public institutions, and were covered by a municipal public guarantee obligation.[3] Due to these governmental guarantees, essentially transforming savings-bank debt into Federal German AAA-rated bonds, savings banks had substantially improved access to refinancing in the capital market. Given this important

[3]The European Court of Justice deemed this an obstacle to competition in retail banking and savings banks were exempted from public guarantee obligation as of July 19, 2005.

institutional linkage, the municipality holds substantial sway over the savings bank's management, as will be detailed in section 3.2.2.

3.2.2 German banking system

Traditionally, the German banking systems relies on three pillars (*Drei-Säulen-Modell*): private banks, savings banks (*Sparkassen*), and cooperative banks (*Genossenschaftsbanken*). Whereas private banks are best described as profit-maximizing firms, savings banks and cooperative banks are legally bound to also pursue welfare enhancing policies, in particular within their region. According to the German Central Bank (*Deutsche Bundesbank*), in 2010 there were roughly 1,100 cooperative banks, 431 savings banks and 218 private banks in Germany.[4] Since savings banks and cooperative banks are the focus of our empirical analysis, we describe these two bank types in more detail.

Savings banks

The first "modern" savings banks in Germany were founded by local governments in the late 18th century in Northern Germany. Initially, the number of savings banks increased from 300 (in 1836) to more than 3,000 (in 1913). Gradually, this number was reduced when for efficiency reasons neighboring local institutions merged. Today there exist 431 savings banks, i.e., roughly in every municipality (electoral district) there is one savings bank.

The German savings-bank sector has a three-level structure: On the local level there are the individual savings banks. On the state level there are associations (*Sparkassen- und Giroverbände*) to realize economies of scale for operative tasks. On the federal level, a further association (*Deutscher Sparkassen- und Giroverband (DSGV)*) is primarily responsible for representing the interests of savings banks towards the federal government and international institutions. All relevant decisions regarding an individuals savings bank's business policy are autonomously made on the local level. Due to their local structure, and imposed by law, the savings banks' operational areas have a strong focus on their respective regions (*Regionalprinzip*). Their main clientele are private customers and local businesses and savings banks hold a large share in the retail banking markets. According to the DSGV, in 2009 they held more than 40% of all private deposits and, important for our paper, more than 40% of all credits to SMEs and

[4] Currently, our sample covers four states with a total of 268 savings banks and 722 cooperative banks.

more than 70% of all credits to handicraft businesses. The latter two are traditionally considered the backbone of the German economy.

Local politicians have strong influence over the banks' decisions, in particular their lending activities. Since savings banks are founded by the local governments and were covered by a public guarantee obligation, municipalities have the formal right to send representatives into the board of directors (*Sparkassenverwaltungsrat*) and the central credit committee (*Kreditausschuss*) of the local savings bank. As a result, their members are to a large degree composed of local parliament members, roughly reflecting the relative political powers in the electoral district. On top of that, the chairmen of both chambers is, as a rule, the executive representative of the respective district. By law, the directors are not bound by an imperative mandate but are supposed to only consider the greater good of the savings bank. Besides general guideline competence, board members also hold substantial sway over credit decisions that exceed the authority of the savings bank's management, as the board of directors or the central credit committee have to vote on those credits (that are either large in size or considered rather risky).

Cooperative banks

The first cooperative banks in Germany were founded by Franz Hermann Schulze-Delitzsch und Friedrich Wilhelm Raiffeisen in the middle of the 19th century. They are organized as cooperatives, i.e. every customer is also a "member". They are locally organized, with basically every municipality being the location of at least one cooperative bank and their main clientele are private customers and local businesses. In 2010 they had a market share of 16% for private deposits, 15% of all credits to SMEs and 28% of all credits to self-employed persons.

Most local cooperative banks are organized in a federal association of cooperative banks (*Bundesverband der Deutschen Volksbanken und Raiffeisenbanken*). Cooperative banks are not covered by the public guarantee obligation but their federal association provides an insurance fund to provide deposit guarantees. Since cooperative banks are independent from governmental institutions and are not protected by public guarantees, politicians have no formal way to influence cooperative banks' business policies.

Hence, cooperative banks are an ideal control group for our purposes: They share the regional structure and their clientele with savings banks[5],

[5]Comparing the regulating laws (our translation) describing the purposes of cooperative banks (here for *Volksbanken*) and savings banks (here for Baden-Württemberg) highlights that they share basically the same objectives:

but they are exempted from the direct control local politicians hold over savings banks' business policies.[6]

3.2.3 German electoral system

Germany has a federal system with three layers of government: the federal state, the 16 states (*Bundesländer*), and 399 municipalities (consisting of 292 counties (*Landkreise*) and 107 urban municipalities (*Kreisfreie Städte*)).[7] Each layer has specific powers and responsibilities as well as separate parliaments and elected executives. These are elected in regular intervals: every 4 years on the federal level, every 4-5 years on the state level and every 4-9 years in the respective municipalities (depending on the state and the form of the municipalities). We focus on the latter class of elections, as it is the *local* ruling politicians that have direct influence on the policies of their local savings bank.

There exist two types of municipal elections: elections of legislative bodies (*Kreistag* in case of counties and *Stadtrat* in case of urban districts) and of executive representatives (*Landrat* in case of counties and *Bürgermeister* in case of urban districts). The latter may occur in direct elections or indirect elections (via the aforementioned legislative bodies).

Elections of legislative bodies

Each municipality has its own parliament. The elections of these legislative bodies are coordinated on the state level, i.e., within a state they all take place on the same election day. These dates, however, generally differ from election dates of federal or state parliaments (*Bundestagswahlen* and *Landtagswahlen*, respectively); i.e., as a rule they are not held on the same day. Furthermore, municipal election dates differ across states with intervals between elections varying between 4 and 6 years. For the states used in our analysis the elections are held every 5 years in Baden-Württemberg, Hesse and Rhineland-Palatinate and every 6 years in Bavaria. The electoral

§1(1) Genossenschaftsgesetz: "[...] to foster the income or the enterprise of the members [...]"
§6(1) Sparkassengesetz Baden-Württemberg: "[...] to ensure the provision with money and credit in their region in particular for SMEs [...]"

[6] In contrast to this, private banks differ greatly from savings banks: First, their business model solely focusses on profit-maximization and is unrestricted by welfare considerations. Second, their outreach is usually not confined to a specific region. Third, and most importantly, their regional representation does not consist of independent regional units but of mere branches that are legally part of operational headquarters. For these reasons, private banks are not suitable as a control group for our purposes.

[7] Currently, our sample covers four states with a total of 202 electoral districts.

system is one of proportional representation with a minimum vote share requirement.

Elections of executive representatives

Executive representatives, on the other hand, are elected with majority rule and runoffs. There are substantial cross-state differences in electoral laws and various changes across time. Some states have direct elections of executive representatives, others indirect elections via municipal parliaments. In the last decades there has been a strong trend towards more direct representation, whereas up to the early 1990s indirect elections were the norm. States also differ with respect to whether they hold the municipal legislative and executive elections on the same time (minority of states) or on differing days (majority of states). In fact, executive elections are usually not held at the state level but dispersed across time. Finally, there are substantial differences regarding the intervals, from 4 to 9 years, between these elections. The details for the states used in our analysis are as follows.

Baden-Württemberg: Elections are held every 8 years and on separate dates from elections of legislative bodies. In counties the executive representatives are determined in indirect elections via municipal parliaments, whereas urban municipalities choose their mayors in direct elections.

Bavaria: All elections are direct and are held every 6 years. Election dates coincide with those of legislative bodies.

Hesse: Since 1993 all elections are direct (before 1993 all were indirect elections via municipal parliaments) and are held every 6 years. Election dates do not coincide with those of legislative bodies.

Rhineland-Palatinate: Since 1994 all elections are direct (before 1994 all were indirect elections via municipal parliaments) and are held every 6 years. Election dates do not coincide with those of legislative bodies.

3.3 Main hypothesis and testable predictions

The main hypothesis this paper seeks to test is whether local politicians take advantage of the institutional environment described in section 3.2 and artificially expand lending in their respective districts in the wake of elections, in hopes of swaying their prospects at the ballot box. The opportunities and incentives for doing so certainly exist. To begin with, there should

not be any real doubt about the connectedness between local politicians and savings bank, as their relation is legally manifest in board-of-directors membership or even chairmanship. Moreover, the board of directors does not have mere representational functions that consist of rubber-stamping the decisions made by the bank's management. On the contrary, the bank's large-scale lending activities are one operational area where directors can directly influence decisions and do not have to rely on taking indirect influence.

Our design that uses cooperative banks as control group allows us to distinguish the increased lending from a mere increase in demand for credit in response to real economic growth around election years, caused e.g. by traditional political spending cycles, as the latter would arguably affect cooperative bank lending equivalently.

The question remains whether increased lending will help politicians to tip the electoral scales in their favor. We argue that there is a number of reasons why this may indeed be the case. First, given the legally mandated local focus of savings banks, borrowers will almost certainly live – and vote – in the region represented by the incumbent, and it is save to assume that constituents will be more satisfied when they are not troubled by credit rationing. This argument becomes even more powerful when loans to SMEs are under consideration, as these may be paramount for the creation or preservation of employment in the politician's district. Second, in small municipalities where little goes unnoticed, the politician's role during credit negotiations may well become common knowledge sooner or later. In that case, hearsay about the mayor relentlessly fighting for her constituency will even send a signal to voters that are not directly affected by the approval of loans. Third, the option of affecting economic outcomes through lending may be attractive inasmuch as the potential costs of this intervention (e.g., higher default rates on the marginally granted credits) are deferred until the loans in question mature. As a consequence, the negative fallout is not instantly visible and may in fact never be traced back to the responsible politician.

The final argument in favor of our main hypothesis concerns the timing of bank lending distortions. If politicians truly exploit their ability to sway the credit operations of their local savings bank, we should expect a concentration of such behavior in times when it helps them most. Assuming that voters are myopic, political gain is maximal when the incumbent's fate is on the line: in the wake of elections. Hence, it seems fair to assume that politically motivated lending be focussed on election seasons rather than equally distributed throughout the legislative period.

As we do not want to rely on these theoretical arguments alone, we formulate five predictions we expect to survive empirical testing, if politicians truly behave as suggested above.

Prediction 1: Election effect. Before municipal elections, local savings-bank lending should increase, compared to a hypothetical situation without elections. Importantly, and facilitating identification, there should be no increased pre-election lending for cooperative banks that are similar in local organization and clientele but are not formally politically influenced.

Prediction 2: Election kind. Lending should react to *municipal* elections. Elections on the state or federal level should have no (or at least a much weaker) systematic impact, since politicians from these levels of government are not institutionally connected with local savings banks.

Prediction 3: Pre- and post-election lending. Politically motivated lending increases should exclusively occur in the ultimate wake of elections. Given voter myopia, they should not happen too far in advance as this would represent premature flexing of political muscle. For similar reasons, the election effect should not be permanent, since incentives to allure voters will instantly vanish once the polls are closed. Whether post-election lending should return to its steady-state level or whether the election increase should be compensated by overly prudent post-election lending is not entirely clear, a priori. Which of the two options is more plausible depends on the credit constraints savings banks face. In addition, confirmation of prediction 3 would reduce the risk of confounding *political* motives behind a surge in lending with the effect of any general booms around elections, as the latter would arguably create less abrupt patterns in the extension of credit.

Prediction 4: Electoral competition. Does politically induced lending depend on the contestedness of elections? Predicting the answer to this question on theoretical grounds alone is not trivial, as electoral competition may exert two offsetting effects. The first is a matter of efficiency considerations, according to which election effects should be more pronounced in districts with fierce *current* electoral competition. The assumption underlying this hypothesis is that – given prediction 5 below holds – politically motivated lending will be costly for

the savings bank and ultimately harm the politician's reputation. As a consequence, incumbents may not make much use of this distorting instrument unless they genuinely fear for their re-election. By contrast, the second effect of electoral competition may curb the politician's ability to influence savings-bank lending if the political process is *generally* contested and has led to close election results in the past. The rationale for this argument is one of dominance: A competitive electoral environment will be reflected in the partisan composition of the bank's board of directors, reducing the likelihood of collusion among board members who represent rivalling political parties. Similarly, the degree of sway over lending decisions may well depend on informal ties between politicians and the bank's management, which are likely to be stronger, the longer the incumbent has been in office. As a result, regular changes in power and slim majorities would limit the scope of electoral lending cycles.

Prediction 5: Increased default risks. Loans granted during the height of an election season should have a higher tendency to default, since some marginal loans would not have been been approved, if financial fundamentals alone had played a role in determining the credit rating.

Whether our hypothesis is verified by empirical testing, is investigated in section 3.6. Before turning to this analysis, however, we continue with the description of our data and discuss which of the above predictions are ultimately testable with the information at hand.

3.4 Data

We use a novel, in large parts hand-collected, dataset that combines information from multiple sources. The observational units are savings and cooperative banks in Southern Germany. This bank data is merged with information on municipal, state-level and federal elections as well as with macro-economic and demographic data on the district level.

3.4.1 Bank data

The source of our bank data is *Hoppenstedt*, a business data provider that hosts the largest commercial database for balance sheets and annual reports in Germany. The main advantage of *Hoppenstedt*, compared to other

commercial databases such as *Bankscope*, are the ample N and T dimensions their sample provides: It covers virtually all savings banks and a very large fraction of cooperative banks that operated in Germany between 1987 and 2009.[8] For the four Southern German states we have initially picked for our analysis (see section 3.4.2), this amounts to a total of 268 savings banks (4,568 bank-year observations) and 722 cooperative banks (6,181 bank-year observations). Note that these numbers include a sizeable amount of banks that exited or entered the sample due to bank mergers. The average time, savings banks remain in the sample is 17 years, whereas the average cooperative bank is only observable for roughly 9 consecutive years. This reflects that our panel is considerably less balanced for cooperative banks, as a large fraction is only covered by the sample since the early 2000s. To ensure that our results are not driven by these sample characteristics, we perform robustness checks by varying the degree of panel balancedness in section 3.7.2.

All information is taken from official balance sheets. The key variables are the bank's overall lending position, the amount of non-performing loans, total assets and the capital ratio. All monetary positions are deflated and measured in 1995 EUR. A look at the panel characteristics reveals that for all items between-variation is substantially greater than within-variation.

Note that the use of balance-sheet data has the disadvantage of rendering prediction 5 untestable within the scope of this paper, as we are unable to determine the time period in which non-performing loans were granted. If marginal credits from pre-election periods were truly more default-prone, this could still take its toll 5, 10, or 15 years after the loan was granted, making it empirically impossible to causally tie the eventual default to the election in question.[9]

3.4.2 Election data

A database that combines information on German municipal elections in any comprehensive way does not exist. Even on the state-level, the collection of local electoral data is the clear exception. For this reason, we have begun to create our own unique dataset by gathering all relevant infor-

[8] We ran several internal consistency checks to ensure that the *Hoppenstedt* data be of comparable quality to that of *Bankscope*.

[9] To test prediction 5, we additionally applied for access to officially kept loan-level bank data at the Federal Savings Banks Association and at the German *Bundesbank* that would have provided information on the exact timing and the eventual (non-)performance of credit contracts. While these data have been made available for academic research in the past, neither institution has granted us access, once we described our research project.

mation ourselves. To this end, we have contacted regional statistical offices, the respective communities, and historical archives all over Germany. Given the enormous labor intensity of this project, data collection is still ongoing. So far, we were able to complete our search for the following Southern German states: Baden-Württemberg, Bavaria, Hesse, and Rhineland-Palatinate. By focussing on neighboring states, we are confident that regional differences be kept to a minimum, improving the comparability of cross-state observations. Together, these four states account for more than half of all German counties and urban districts (202 out of 399). As data collection progresses, we will gradually extend our sample until all of Germany is covered.

The Southern German municipal election data already gathered, covers the years between 1970 and 2009. Yet, since this political data is merged with the aforementioned bank data, the maximum interval for our analysis is effectively reduced to 1987–2009 as well. During this time span, Bavaria held 4, and the other states 5 elections of legislative bodies. The pattern for executive elections is a lot more irregular: Districts in Bavaria and Hesse experienced an average of 4 elections, Rhineland-Palatinate and Baden-Württemberg's urban municipalities just 2 to 3, and finally counties in Baden-Württemberg – qua institution – none.

Our dataset contains information on election dates, election results (measured in vote shares), names and party affiliations of incumbents and winners, election types (direct vs. indirect), whether runoffs were necessary and whether there was a change in executive power. Overall, data quality is better for legislative elections than for executive elections if the latter were held in indirect fashion through municipal parliaments, since for these votes information is unavailable. In fact, we do not even know *when* such indirect elections of regional executives occur. We only observe if the identity of the person in power changes. Whether this is due to electoral defeat, retirement, or death, however, is unknown. For this reason, our main analysis focusses on *legislative* elections and their effect on savings bank lending. While there are good reasons to believe that this election type is in fact the more relevant one for the mechanism under study, this is ultimately an empirical question that we defer to section 3.7.1.

To enable empirical testing of prediction 2, we have also added dates and outcomes of state and federal elections.

3.4.3 District data

Finally, to warrant better control for confounding factors and to increase statistical precision, we augment our sample with macro-economic and demographic information at the district level, which are available at the German Federal Statistic Office (*Statistisches Bundesamt Deutschland*). These include population size, GDP, unemployment, public spending and expenditure, public debt, as well as firm creation, closures and bankruptcies. Once again, all monetary values are converted to 1995 EUR. Available time spans vary significantly among these variables so that the addition of certain control variables would result in significant loss of sample size. The longest time series are available for GDP, population size and unemployment, spanning from the early 1990s to 2009. The collection of the other variables by the Statistic Office sets in considerably later. As a result, the effective time-span covered by our preferred econometric specification presented in section 3.5 covers the years 1993 to 2009, whereas longer time spans are analyzed for robustness in section 3.7.

3.4.4 Descriptive statistics

Summary statistics of variables used in our analysis are presented in table 3.1. Overall, our data is substantially right-skewed, which is why our preferred empirical specification presented below makes use of log-transformed data. As is evident from panel A, savings banks are on average larger than their cooperative counterparts, which reflects the wave of mergers that took place among German savings banks throughout the years. Judging from the ratio of loans and total assets, both bank types clearly set their business focus on lending operations: The average loan position of savings banks makes up 70% of the entire balance sheet, while that number is even slightly higher for cooperative banks, which devote 75% of their operations to providing credit. Furthermore, the capital ratio seems to be mildly, but systematically, larger for cooperative banks.

A look at panel B reveals that municipalities in Baden-Württemberg and Bavaria are clearly dominated by conservative parties – Bavaria's *Christlich-Soziale Union (CSU)* and its sister party, *Christlich Demokratische Union (CDU)*, which competes in the rest of Germany – whereas the other two states see a closer gap between the main political rivals: For one, Germany's largest left-of-center party, *Sozialdemokratische Partei Deutschlands (SPD)*, generally fares very poorly in the two former states. In addition, incumbent dominance appears to be much stronger, suggesting a rather static political environment. As an illustration, consider that only about 6% of all

Table 3.1. *Variables used for analysis*
Summary statistics

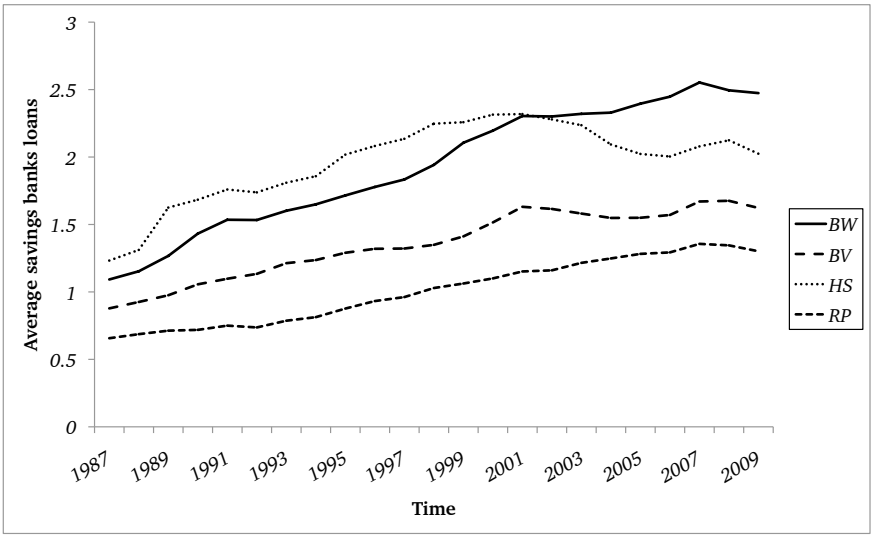
Variables	Total	BW	BV	HS	RP
	N=10,738	N=3,624	N=4,229	N=1,639	N=1,246
Panel A: Banks					
Savings banks					
- No. of banks	268	73	103	52	40
- Total assets	1.936	2.295	1.658	2.485	1.464
	(1.950)	(1.920)	(1.579)	(3.095)	(0.790)
- Loans	1.346	1.588	1.148	1.749	1.023
	(1.357)	(1.332)	(1.078)	(2.176)	(0.599)
- Capital ratio	0.046	0.043	0.048	0.044	0.046
	(0.010)	(0.008)	(0.012)	(0.009)	(0.010)
Cooperative banks					
- No. of banks	722	241	316	95	70
- Total assets	0.651	0.655	0.623	0.758	0.575
	(1.581)	(0.687)	(2.408)	(0.881)	(0.641)
- Loans	0.488	0.484	0.468	0.565	0.451
	(1.230)	(0.467)	(1.892)	(0.665)	(0.556)
- Capital ratio	0.057	0.056	0.058	0.058	0.056
	(0.017)	(0.012)	(0.017)	(0.028)	(0.014)
Panel B: Municipal elections					
No. of elections	19	5	4	5	5
Vote share *CDU/CSU*	38.97	36.83	42.45	35.35	40.64
	(7.39)	(7.07)	(6.00)	(7.02)	(7.75)
Vote share *SPD*	26.33	21.06	23.42	38.55	34.46
	(9.80)	(5.09)	(8.27)	(7.76)	(8.18)
Vote share swing	9.30	8.21	8.99	10.25	12.19
	(2.56)	(2.30)	(2.42)	(1.67)	(1.85)
Party change	0.105	0.059	0.065	0.245	0.171
	(0.160)	(0.115)	(0.158)	(0.153)	(0.172)
Panel C: Municipal districts					
No. of districts	199	44	94	26	35
Population	33,364	10,754	12,539	6,067	4,004
Real GDP	6.644	8.339	5.731	7.901	3.226
	(8.646)	(5.134)	(11.369)	(8.305)	(2.089)
Unemployment rate	7.01	6.11	6.81	8.35	8.48
	(2.56)	(1.84)	(2.58)	(2.74)	(2.72)

Notes: Reported are total numbers (for the state level) and means (for the district level) respectively. For the latter, standard deviations are in brackets. BW, BV, HS, and RP denote the states of Baden-Württemberg, Bavaria, Hesse, and Rhineland-Palatinate, respectively. N stands for the number of available bank-year observations. Election data refers to municipal elections of legislative bodies. *CDU/CSU* are the conservative parties and *SPD* the social-democratic party of Germany. "Vote share swing" denotes the average swing in vote shares (cumulated over all parties) that results from a given election. "Party change" indicates the share of elections that result in a change of the winning party. State population is measured in million habitants (as of 2010). All monetary values are measured in 1995 EUR billion.

municipal elections in Bavaria and Baden-Württemberg result in a change of the winning party, whereas Hesse and Rhineland-Palatinate experience such changes in power after 25% and 17% of all elections, respectively.

Note that these summary statistics are for pooled data and represent an average over time. To better assess the dynamics of German bank lending,

Figure 3.1. *Time trends in bank lending 1*
Savings bank lending across states



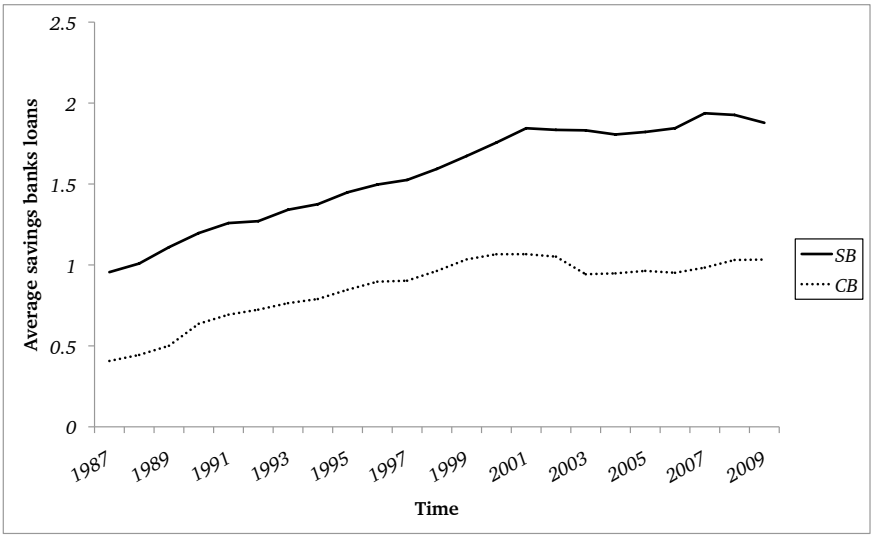
Notes: Depicted are time series from a balanced panel of average savings bank lending for Baden-Württemberg (BW), Bavaria (BV), Hesse (HS) and Rhineland-Palatinate (RP). Loans are measured in 1995 EUR billion.

figure 3.1 plots the time series of average savings bank lending, stratified by state. Clearly, our loan data is subject to an upward trend. For this reason, section 3.5 proposes two alternative approaches to account for the effect of time: a year dummies specification and quadratic time trends. Overall, savings banks in all four states appear to be on similar time trends which provides good news for a DD identification strategy such as ours (see Angrist and Pischke, 2009). If anything, Hesse's trend becomes a bit idiosyncratic after the turn of the century, which is why results that seem exclusively driven by Hesse may be taken with a grain of salt. Finally, figure 3.2 shows that time trends are also comparable for both bank types (averaged over all states in our sample), which provides further evidence that cooperative banks are indeed a valid control group for savings banks.

3.5 Methodology

Our strategy to identify any causal effect of elections on savings-bank lending, relies on the fact that we should only observe politically motivated lending before election years, only in municipalities in which elections are held at this point in time, and – importantly – only for politically connected savings banks. Identification is facilitated by the different timing of elections across states and the existence of a control group of cooperative banks that operate in the same electoral districts as savings banks. Furthermore,

Figure 3.2. *Time trends in bank lending 2*
Savings bank versus cooperative bank lending



Notes: Depicted are time series from a balanced panel of savings bank (SB) and cooperative bank (CB) lending, averaged over all four states in our sample. Loans are measured in 1995 EUR billion.

given the statutory nature of legislative elections at the municipal level, for which early elections are largely non-existent, we certainly need not worry about any endogeneity in (the timing) of our key regressor. Econometrically, we conduct difference-in-difference (DD) as well as triple-difference (DDD) estimation embedded in a fixed-effects panel data setup.

Testing prediction 1: Election effect

A natural starting point for testing prediction 1 that savings-bank lending increases in the wake of elections, is the following empirical specification:

$$Y_{ist} = \mathbf{X'_{ist}}\boldsymbol{\beta_1} + \mathbf{S'_s}\boldsymbol{\gamma_1} + \mathbf{T'_t}\boldsymbol{\lambda_1} + \delta_1 ELEC^M_{st} + \epsilon_{ist}, \tag{3.1}$$

where Y_{ist} is a measure for loans from savings bank i, operating in state s at time t. The parameter of interest, δ_1, estimates the causal effect of municipal election seasons, which are indicated by the dummy variable $ELEC^M_{st}$.[10] To ensure identification of δ_1, we control for the following covariates and fixed effects: $\mathbf{S_s}$ donates a full vector of state effects to control for secular lending differences across states. Similarly, time effects, $\mathbf{T_t}$, are included to capture any national trends or year shocks. Note that we use two al-

[10]Note that the way model 3.1 is written down, corresponds to the case of legislative elections, $ELEC^M_{st}$, as these vary at the state level. For election of executive representatives (dealt with in chapter 3.7.1), $ELEC^M_{ist}$ would be permitted to also vary at the district (and, hence, individual) level.

ternative specifications to capture the effect of time: The first consists of introducing a full set of year dummies, whereas the second – with the purpose of reducing the number of parameters to be estimated – controls for quadratic time trends instead. Finally, $\mathbf{X_{ist}}$ is a vector of bank- and district-specific variables that may directly influence the outcome variable. The inclusion of these covariates should considerably improve the predictability of Y_{ist}, which will in turn reduce the sample variance of our estimates.

Estimation of model 3.1 by OLS ensures that both T- and N-variation are exploited. The former compares the same banks across time, as each bank will be subject to recurring election "treatments". The latter contrasts different banks at a given time, as municipal elections dates vary across states. Yet, while a positive estimate of δ_1 would speak in favor of prediction 1 and suggest that savings banks are indeed systematically easing credit before elections, we cannot yet say anything about the reasons for this behavior. For instance, it could well be that banks simply react to increased demand for credit if – for some reason – local GDP were positively correlated with the timing of elections. Even though we are able to shield ourselves from such confounds by adding macro-economic district variables to the list of controls $\mathbf{X_{ist}}$, a fundamental problem remains: So far, we have no way of gauging whether the responsibility for the credit boost lies with local politicians and their involvement with the savings banks' board of directors. The latter hypothesis, however, would be substantially strengthened if the detected pattern were to apply to savings banks alone.

The control group of cooperative banks allows us to identify this: If our hypothesis is true, we should not observe increased cooperative-bank lending before elections, as politicians have no institutional sway over their credit policies.[11] Therefore, a more convincing way of testing prediction 1 consists of applying the following DD model to a sample that includes both types of banks:

$$Y_{isbt} = \mathbf{X'_{isbt}}\boldsymbol{\beta_2} + \mathbf{S'_s}\boldsymbol{\gamma_2} + \mathbf{T'_t}\boldsymbol{\lambda_2} + \mu_2 B_b + \theta_2 ELEC^M_{st} + \delta_2 ELEC^M_{st} * B_b + \epsilon_{isbt}. \quad (3.2)$$

Model 3.2, which is also estimated by OLS, differs from specification 3.1 inasmuch as bank-type effects, B_b, are needed to control for perpetual differences between savings and cooperative banks. B_b is defined as a dummy variable that takes on the value of 1 if the individual unit is a savings bank. In addition, we interact the election dummy with the bank-type indicator

[11] This, of course, can – and will – be individually tested by applying specification 3.1 to the sample of cooperative banks. This time, we would expect δ_1 to *not* be significantly different from zero for our prediction to hold.

such that $ELEC^M_{st} * B_b$ switches on if and only if Y_{isbt} measures lending activity of a savings bank during election season.

The parameter of interest is denoted δ_2 and arguably provides improved identification over δ_1, since the representation of counterfactual lending in the absence of election is more encompassing. The estimate from model 3.1 measures the difference in lending behavior between savings banks in election periods and savings banks in non-election periods. The DD estimate from model 3.2, on the other hand, captures the difference between election-induced increase in savings-bank lending (which is expected to be positive after controlling for time trends) and election-induced increase in cooperative-bank lending (which is expected to be zero after controlling for time trends).

Testing prediction 2: Election kind

Our main hypothesis would be further strengthened if prediction 2 – that only municipal elections have a systematic impact on savings-bank lending – were to survive empirical testing as well. Recall that it is local politicians who are granted membership in the bank's board of directors. While a few exceptions from this rule (with members of state parliaments being granted access as well) certainly exist, any potential effect should at least be considerably weaker than that of municipal elections.

Empirical testing of prediction 2 is straightforward if state elections are concerned, as model 3.2 can be applied almost verbatim since both, legislative municipal elections and state elections (which are always for the state *legislature*) vary at the state level. The only difference to the specification used for prediction 1 is that $ELEC^M_{st}$ is replaced with an indicator for state election seasons, $ELEC^S_{st}$. The case of federal elections (which are also parliamentary), however, requires a slightly altered specification, as there is no more cross-state variation in treatment:

$$Y_{isbt} = \mathbf{X'_{isbt}}\boldsymbol{\beta_3} + \mathbf{S'_s}\boldsymbol{\gamma_3} + \mathbf{T'_t}\boldsymbol{\lambda_3} + \mu_3 B_b + \theta_3 ELEC^F_t + \delta_3 ELEC^F_t * B_b + \epsilon_{isbt}. \quad (3.3)$$

Note that $ELEC^F_t$, which indicates federal election periods, only varies in the time dimension, taking the value of 1 every 4 years. As a consequence, the effect of federal elections will not be identified in case year dummies are used to control for time effects. For this reason, $\mathbf{T_t}$ will have to be represented by quadratic time trends when specification 3.3 is used.

Testing prediction 3: Pre- and post-election lending

Another way of solidifying support for our hypothesis is to look at pre- and post-election periods, as the increase in lending should be confined to the

immediate election season. Particularly, we expect lending policies to immediately return to their steady-state level once ballots are cast. In case higher default rates on the marginally granted credits need to be compensated, we may even expect overly prudent lending behavior in the aftermath of elections. Prediction 3 can be tested with the following specification to be estimated with OLS:

$$Y_{isbt} = \mathbf{X'_{isbt}}\boldsymbol{\beta_4} + \mathbf{S'_s}\boldsymbol{\gamma_4} + \mathbf{T'_t}\boldsymbol{\lambda_4} + \mu_4 B_b + \theta_4 ELEC^M_{st-\tau} + \delta_4 ELEC^M_{st-\tau} * B_b + \epsilon_{isbt}, \tag{3.4}$$

and $\tau = (1,2,3,4)$.[12] To study post-election periods, the dummy variable $ELEC^M_{st-\tau}$ indicates whether there was an election in state s, τ years ago. We expect the estimate of δ_4 to be either close to zero or, in case of binding credit constraints, negative. For the analysis of pre-election periods, $ELEC^M_{st-\tau}$ will be replaced with a corresponding $ELEC^M_{st+\tau}$ indicator variable. Once again, we would interpret a zero estimate of δ_4 as supportive evidence for prediction 3.

Testing prediction 4: Electoral competition

The test for prediction 4 can be implemented with the following DDD model, estimated with OLS:

$$\begin{aligned} Y_{isbt} = {} & \mathbf{X'_{isbt}}\boldsymbol{\beta_5} + \mathbf{S'_s}\boldsymbol{\gamma_5} + \mathbf{T'_t}\boldsymbol{\lambda_5} + \mu_5 B_b + \psi_5 I_{it} + \theta_5 ELEC^M_{st} + \dots \\ & + \phi^1_5 B_b * I_{it} + \phi^2_5 B_b * ELEC^M_{st} + \phi^3_5 I_{it} * ELEC^M_{st} + \dots \\ & + \delta_5 ELEC^M_{st} * B_b * I_{it} + \epsilon_{isbt}, \end{aligned} \tag{3.5}$$

where I_{it} is the respective indicator variable of interest: In case *current* electoral competition is investigated, $I_{it} = C_{it}$ is an indicator for whether the present election is contested. The ruling party's dominance (or alternatively: the lack of electoral competition *in general*) is measured with $I_{it} = D_{it}$.[13] In line with our predictions in section 3.3, the former indicator switches on if the election is competed, while the latter takes the value of one in case the political process is *not* contested. The first line of model 3.5 contains the usual controls as well as all main fixed effects. Line 2 contains the full set of first-order interactions which are necessary to identify the causal effect of interest, captured by the DDD estimate of δ_5 in line 3 (see Gruber, 1994).

[12] Higher-order lags exceeding 4 years should not be used, since this would blur the line between post-election periods of the past and pre-election periods of the next campaign.

[13] Note that we use several alternative measures for electoral contestedness and party dominance (see section 3.6). While in general D_{it} may change its value depending on the competitiveness of the election under consideration, D^G_i is defined to be time-constant in case we want to capture the *general* electoral competition of the district, bank i operates in.

Testing prediction 5: Increased default rates

As previewed in section 3.4.1, prediction 5 is untestable with the balance-sheet data at hand, as it shows the entire *stock* of non-performing loans at year's end that also reflects lending activities outside of election seasons. Since we do not know when a given non-performing loan was granted, we will, for the scope of this paper, have to leave the question about the welfare cost of politically induced lending unanswered and adjourn this analysis to future research, in hope that access to loan-level data will eventually be granted.

3.6 Results

All results presented here are estimates from an unbalanced panel to which we apply our preferred empirical specification with the following properties: The dependent variable, Y_{it}, is defined as the natural logarithm of the loan sum of bank i as reported in the balance sheet for year t. This facilitates interpretation of coefficients – which represent (semi-)elasticities – and accounts for the right-skewedness of our data. The pre-election indicator, $ELEC_{st}$, is defined as follows: It takes on the value of 1 if there is an election in either the final two quarters of the same year, or the first two quarters of the following year.[14] The vector of control variables, $\mathbf{X_{ist}}$, includes bank-specific (total assets and capital ratio, to account for bank size and degree of capitalization, respectively) and district-specific (population size and real GDP) covariates. To account for the possibility that the bank variables are only sequentially exogenous, we use their lagged values instead (see Dinç, 2005). All elements of $\mathbf{X_{ist}}$ are log-transformed. Finally, standard errors are clustered on the bank level (as opposed to the bank-year level), to correct for substantial serial correlation. Note that our results are not driven by these modeling choices. As section 3.7 demonstrates, the main conclusions are insensitive to varying definitions of key variables, sets of controls, sample compositions, and assumptions regarding the error-term structure.

[14] This definition ensures that election-induced lending is reflected in the balance sheet of the actually relevant year: If an election takes place in, say, January, pre-election lending will arguably leave its mark in the balance sheet of the previous year, which is why the latter will switch on $ELEC_{st}$, whereas $ELEC_{st} = 0$ for the actual election year. By contrast, if the election is held around year's end, the balance sheet of the preceding year is probably less informative than that of the election year, for which reason the pre-election indicator would then coincide with the year of the election. Note that results, not reported here, based on alternative definitions of $ELEC_{st}$, are comparable to those presented in this paper.

In a nutshell, all of our testable predictions withstand empirical scrutiny, which corroborates our hypothesis that local politicians exploit their membership in savings banks' directing boards to sway their electoral fortunes. Not only do all estimated effects have the correct sign, they are also statistically significant at the 5% level and, in the majority of cases, even at the 1% level. Effect sizes are estimated to be in the range of 1%–3% election-induced increase in the stock of lending. Note that this increase is relative to the *total* stock in bank lending. If we were to model the extension of *new* credit contracts alone, relative effect sizes would certainly be larger. To provide a better sense for the actual magnitude of the effect, consider that its absolute size amounts to an average of EUR 30.6 million extra stock in lending per bank, when our preferred empirical specification is used.

Prediction 1: Do savings banks expand lending prior to elections?

The empirical tests of prediction 1 are summarized in table 3.2, which contains OLS estimates of the key parameters from models 3.1 and 3.2, as well as regression coefficients of control variables. The first four columns display results from variants of model 3.1 being applied to savings banks. Estimates of the key parameter, δ_1, can be found in the first row. Results in column (A) suggest that savings banks increase lending by about 1.6% in the wake of municipal elections. This result is largely unchanged by the inclusion of state fixed effects in column (B). As the comparison of results in columns (C) and (D) shows, estimates are somewhat sensitive to the way, time effects are accounted for: If quadratic time trends are used, effect sizes are roughly twice as large as those coming from a year-dummy specification.

While these results are certainly encouraging, our hypothesis would be greatly invigorated if the election effect were exclusively present for savings banks. This enhanced prediction is under study in columns (E) through (H) of table 3.2. The first two of those repeat the analysis of columns (C) and (D) for the subsample of cooperative banks. As predicted, there is no evidence for these politically unconnected banks adjusting their lending in response to elections. The last two columns contain the DD results of model 3.2 being applied to a sample that includes both bank types. Apparently, the presence of a control group strengthens results considerably, suggesting that savings banks' lending increases by 2% to 3% in the wake of elections. These δ_2 estimates are statistically significant at the 1% level. Moreover, the sensitivity of results to the way time effects are incorporated is no longer present, as both controlling for time trends and including year dummies, yield almost identical results.

Since we believe that DD model 3.2 provides improved identification of the causal election effect over model 3.1, the remainder of this section describes results that only make use of the former specification. Yet, since the superiority of the DD model critically depends on the assumption that cooperative banks are indeed an appropriate control group, for robustness we also tested predictions 2, through 4 with model 3.1 that only uses differences between savings banks. As results in appendix 3.A demonstrate, our main conclusions are largely insensitive to which of the two specifications we use.

Besides these causal effects of interest, there is only one additional covariate with a statistically significant impact on lending: a bank's (lagged) total assets. In fact, the relationship between these two variables proves to be extremely strong. Not only are estimated effects very sizeable, the inclusion of total assets to the set of regressors leaves barely any variation in the data unexplained, as is evidenced by R^2 exceeding 0.95. Note that this is not an indication for overfitting. Much rather, this tight connection is not surprising since German financial regulation *mandates* that a bank's lending position be backed by equivalent net equity. Given this quasi-mechanical relationship between these variables, we repeat our analysis for robustness without total assets in section 3.7.1.

Prediction 2: Does lending react to other elections?

To further back up our theory of *local* politicians being the driving force behind electoral cycles in bank lending, we now turn to the second prediction that credit policy should react only to municipal elections. A look at table 3.3 suggests that this seems to indeed be the case. This table contains estimates for δ_2 and δ_3 when DD models 3.2 and 3.3, respectively, are applied to our full sample. In line with our premise, we find no evidence that lending reacts in any systematic way to either state (see columns (A) and (B)) or general (see column (C)) elections. These findings are corroborated by an additional specification which jointly regresses on all three election types. As results in column (D) show, savings bank lending is in fact only responsive to elections at the municipal level, whereas elections at higher government levels have practically no influence.

Prediction 3: What happens to lending before and after election seasons?

Prediction 3 suggests that the increase in lending should be limited to the immediate election season. Particularly, the positive effect should instantly disappear, or even become negative, once the election was held. We have

Table 3.2. *Results for prediction 1*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)							
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Key regressors								
- $ELEC^M_{st}$	**0.016** (**0.001**)	**0.017** (**0.000**)	**0.010** (**0.032**)	**0.018** (**0.000**)	-0.001 (0.795)	0.006 (0.310)	-0.008 (0.137)	0.001 (0.913)
- $ELEC^M_{st} * B_b$	–	–	–	–	–	–	**0.027** (**0.001**)	**0.024** (**0.004**)
Bank controls								
- Total assets	**0.988** (**0.000**)	**0.988** (**0.000**)	**0.999** (**0.000**)	**0.991** (**0.000**)	**0.959** (**0.000**)	**0.959** (**0.000**)	**0.977** (**0.000**)	**0.976** (**0.000**)
- Capital ratio	-0.015 (0.599)	-0.024 (0.389)	0.011 (0.758)	0.013 (0.719)	0.020 (0.512)	0.023 (0.466)	0.037 (0.102)	0.039 (0.083)
District controls								
- Population	-0.003 (0.890)	0.005 (0.825)	0.004 (0.865)	0.004 (0.858)	-0.028 (0.504)	-0.031 (0.452)	-0.016 (0.548)	-0.018 (0.502)
- Real GDP	0.025 (0.195)	0.025 (0.203)	0.024 (0.222)	0.024 (0.219)	0.053 (0.145)	0.056 (0.122)	0.043 (0.080)	0.045 (0.067)
State FE		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE			Yes		Yes		Yes	
Time trends				Yes		Yes		Yes
Bank type FE							Yes	Yes
Banks in sample	SavB	SavB	SavB	SavB	CoopB	CoopB	All	All
N	3,083	3,083	3,083	3,083	4,033	4,033	7,116	7,116
R^2	0.969	0.969	0.970	0.969	0.966	0.966	0.978	0.978

Notes: Results are for our preferred empirical specification (see text). Key regressors are $ELEC^M_{st}$ for columns (A) to (F), and $ELEC^M_{st} * B_b$ for columns (G) and (H), respectively. The index M denotes a municipal election. Boldfaced numbers indicate statistical significance at the 5% level.

tested this hypothesis by estimating the effect of municipal elections on savings-bank lending in the four years preceding and following a municipal election. The results are depicted in figure 3.3, which provides a visual representation of the electoral lending cycle. The solid line represents DD estimates of the effect elections have on lending in the eight years surrounding said election. The dotted lines indicate 95% confidence intervals. Note that these δ_4 estimates come from 9 distinct regressions, for which τ is accordingly varied in model 3.4.[15]

As is evident from the graph, lending stays relatively flat before election season, only to spike upwards in the immediate wake of an election year. This increase is statistically significant at the 1% level (and identical to the estimate in column (G) of table 3.2). As expected, the effect quickly dissipates after the election, returning to its steady state level or even slightly below. Three years after the election, the election effect dips

[15]The empirial specification used for this graph controls for time via year fixed effects. Notably, the depicted pattern also holds if DD model 3.4 is combined with quadratic time trends instead.

Table 3.3. *Results for prediction 2*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)			
	(A)	(B)	(C)	(D)
Key regressors				
- $ELEC^{M}_{st}$	–	–	–	0.002 (0.836)
- $ELEC^{M}_{st} * B_b$	–	–	–	**0.027 (0.006)**
- $ELEC^{S}_{st}$	-0.005 (0.358)	0.002 (0.643)	–	0.004 (0.491)
- $ELEC^{S}_{st} * B_b$	0.007 (0.245)	0.003 (0.670)	–	0.011 (0.142)
- $ELEC^{F}_{t}$	–	–	-0.005 (0.344)	-0.006 (0.285)
- $ELEC^{F}_{t} * B_b$	–	–	-0.005 (0.357)	-0.004 (0.486)
Bank controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Time FE	Yes			
Time trends		Yes	Yes	Yes
Bank type FE	Yes	Yes	Yes	Yes
Banks in sample	All	All	All	All
N	7,116	7,116	7,116	7,116
R^2	0.977	0.977	0.977	0.977

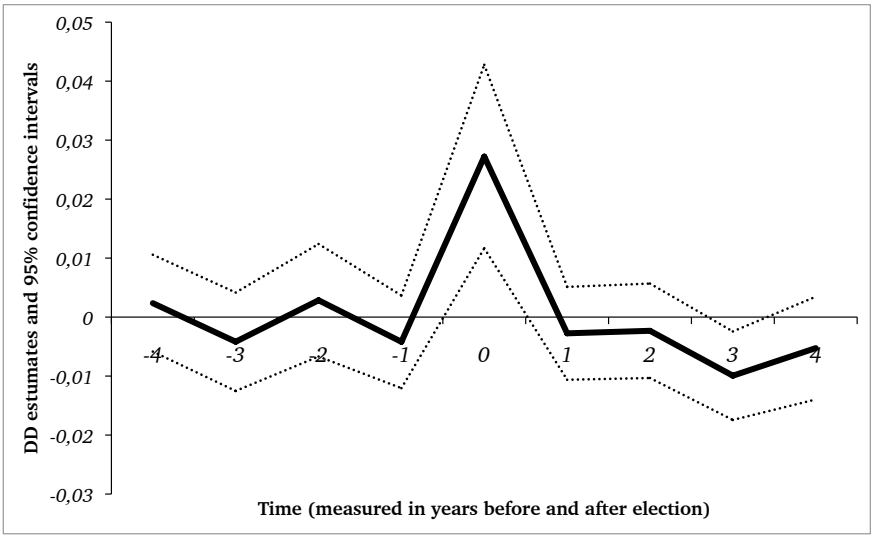
Notes: Results are for our preferred empirical specification (see text). Key regressors are $ELEC^{M}_{st} * B_b$, $ELEC^{S}_{st} * B_b$, and $ELEC^{F}_{t} * B_b$, respectively. The indexes M, S, and F denote municipal, state, and federal elections, respectively. Boldfaced numbers indicate statistical significance at the 5% level.

into statistically significant negative territory (at the 5% level). We take this as weak evidence for overly prudent lending policies after elections, consistent with a binding credit constraint that banks face as they have to make up for excessive pre-election lending. Importantly, the spike in lending appears to be too abrupt and short-lived to reflect an increase in demand for credit in response to real economic growth around election years, as the latter would arguably result in a much smoother pattern. Thus, figure 3.3 provides strong evidence for the political nature of the increase in loan extensions.

Prediction 4: What is the role of electoral competition?

Finally, prediction 4 suggests that electoral competition may have two partly offsetting effects. For one, the increase in lending may depend on the *ability* to manipulate the bank's policies and hence be more pronounced, the clearer the general (or historical) dominance of the incumbent party. This

Figure 3.3. *Results for prediction 3*
Visualization of the lending cycle



Notes: Results are for our preferred empirical specification (see text). The solid line depicts DD estimates of δ_4 coming from model 3.4 with year fixed effects. Dottet lines indicate the corresponding 95% confidence intervals. Time is measured on the abscissa: A value of zero denotes an election season. Negative and positive values stand for years before and after an election, respectively.

hypothesis is under consideration in table 3.4. We use two alternative measures to capture the strength of the ruling party. The first, D_i^G, is an indicator variable that switches on if the electoral district, bank i operates in, experiences relatively few changes in party power.[16] As column (A) of table 3.4 indicates, there is some evidence that stability of incumbency may in fact be a precondition for electoral cycles in lending: Despite being barely imprecisely estimated, savings banks in politically stable areas increase lending by 2.6% in the wake of elections relative to districts that see more frequent changes in power. For the latter, this election effect is – albeit positive and amounting to 1.3% – statistically insignificantly different from zero.

A similar picture emerges for our second measure of incumbent dominance, D_{it-1}, that reflects results of the *preceding* election, hence capturing

[16]To construct this measure, we create a normalized index that counts the number of times the strongest party has changed within a district. D_i^G indicates whether the electoral district under consideration ranks in the bottom quartile of the distribution of said index. Note that a change in relative party strength may not necessarily translate into a change in power, as the party with the plurality of votes may fail to reach an outright majority, in which case it may have to accept opposition status if the other parties agree to form a coalition government. This notwithstanding, D_i^G should provide a reasonable approximation to the general stability of incumbency, we are ultimately interested in.

Table 3.4. *Results for prediction 4*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)				
	(A)	(B)	(C)	(D)	(E)
DDD Interaction					
- $ELEC^M_{st} * B_b * D^G_i$	0.026 (0.082)	–	–	**0.026** **(0.049)**	–
- $ELEC^M_{st} * B_b * D_{it-1}$	–	0.028 (0.088)	–	–	**0.035** **(0.048)**
- $ELEC^M_{st} * B_b * C_{it}$	–	–	0.007 (0.583)	0.016 (0.234)	0.015 (0.318)
DD Main effect					
- $ELEC^M_{st} * B_b$	0.013 (0.385)	**0.023** **(0.010)**	**0.024** **(0.004)**	0.003 (0.831)	0.015 (0.190)
Bank controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Bank type FE	Yes	Yes	Yes	Yes	Yes
First-order interactions	Yes	Yes	Yes	Yes	Yes
Banks in sample	All	All	All	All	All
N	7,116	7,084	7,116	7,116	7,084
R^2	0.977	0.977	0.977	0.977	0.977

Notes: Results are for our preferred empirical specification (see text). D^G_i indicates whether political contestedness is *generally* low in the district, bank i operates in. D_{it-1} indicates whether the *preceding* election was close. C_{it} measures the contestedness of the *current* election. Boldfaced numbers indicate statistical significance at the 5% level.

the relative strength of the sitting government during its expiring term.[17] Results in column (B) indicate that lending increases by – again, barely imprecise – 2.8% in districts with clear majorities relative to all other districts, for which the electoral effect already amounts to 2.3%.

The converse effect of current electoral competition, which may increase *incentives* to induce a lending cycle, is under study in the last three columns of table 3.4. Since pre-election polling is generally unavailable for municipal elections, we have to rely on an ex-post measure when assessing the contestedness of the electoral campaign: the actual election outcome. Given rational expectations and a reasonable feeling of local politicians for the mood of their electorate, we argue that the closeness of the final result should provide a reasonable proxy for the perceived closeness of the contest itself. On this account, our indicator of current electoral competition,

[17] Similar to the first measure, D_{it-1} is a bottom-quartile indicator for a normalized index that measures both, the absolute vote share of the winning party, as well as its margin of victory.

C_{it}, takes on the value of 1 if the winner's final vote share is either below 40% or if the winning margin is less than 5%.[18]

According to estimates in column (C), there is no compelling evidence of current electoral competition exerting any systematic influence on the strength of the election effect, which is estimated to be 2.4%. Yet, given that our measure of present contestedness may in part capture the diametric effect of party dominance as well, we refine our empirical specification by additionally controlling for general and historical contestedness, respectively. Results are displayed in columns (D) and (E) and they suggest that, given overall party dominance, there is at least some indication that lending cycles may indeed be more likely if the upcoming election promises to be close. According to our point estimates, savings banks in districts that fall into this category, increase their lending by roughly 1.5% relative to a situation with little competition. However, given the apparent lack of precision of these estimates, we refrain from interpreting these results as anything more than suggestive evidence. Notably, earlier results that incumbent dominance can be viewed as a precondition for politically induced lending, are soundly reconfirmed, as statistically significant election effects appear to be exclusively present in districts with high degrees of political stability and current incumbent strength, respectively.

3.7 Robustness

As mentioned above, results presented in section 3.6 are based on our preferred empirical specification being applied to the full data sample. To ensure that conclusions are not driven by these choices, we present a number of robustness checks that demonstrate that the election effect, as measured by δ_2, is immune to varying definitions of key variables, sets of controls, sample compositions, and assumptions regarding the error-term structure. Note that the following results should be compared to the baseline estimates in columns (G) and (H) of table 3.2, which suggest politically induced lending increases of 2.7% and 2.4% (relative to the bank's total loan sum), respectively.

[18] Note that we have also experimented with alternative combinations of cutoff points – namely vote shares of 37% and 43% as well as winning margins of 3%, 7% and 10% – but that results are largely insensitive to these variations. The same holds true when instead applying a quartile indicator that follows the same logic as the index-based measure of past incumbent dominance, D_{it-1}.

Table 3.5. *Alternative dependent variables*
Dependent variables: See table notes

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)			
	(A)	(B)	(C)	(D)
Key regressor				
- $ELEC^{M}_{st} * B_b$	**30.593** (0.001)	**26.057** (0.006)	**0.019** (0.000)	**0.016** (0.000)
Bank controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Time FE	Yes		Yes	
Time trends		Yes		Yes
Bank type FE	Yes	Yes	Yes	Yes
N	7,116	7,116	7,116	7,116
R^2	0.977	0.977	0.135	0.131

Notes: Dependent variables for columns (A) and (B) are bank's loans in 1995 EUR and logs of (loans/total assets) for columns (C) and (D). Boldfaced numbers indicate statistical significance at the 5% level.

3.7.1 Alternative choices of variables

Alternative dependent variables

We start by gauging the robustness of the election effect to the choice of the dependent variable. Columns (A) and (B) of table 3.5 contain estimates of model 3.2, when applied to loan data in real values instead of log-transformed data. Apparently, our results in section 3.6 are not driven by the log-transformation of variables, as sign and significance of estimates are very similar to those in corresponding columns (G) and (H) of table 3.2. As a side effect, these estimates provide a better feel for the average size (in absolute terms) of the election effect, which is estimated to lie in the range of EUR 26.0 million and 30.6 million per bank, depending on the way time effects are controlled for.

Yet another specification can be found in tables (C) and (D) of table 3.5. Here, loans are normalized by total assets before being transformed into logs, so that loan sum sizes are put into perspective with the size of the respective bank. Once again, the election effect is positive and highly significant at the 0.1% level.

Alternative control variables

The following two tables provide evidence that the election effect is also robust to variations in the set of covariates used for the analysis. Table 3.6 displays results for specifications that drop certain variables from the list

Table 3.6. *Alternative control variables 1: Fewer covariates*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Key regressor							
- $ELEC^{M}_{st} * B_b$	**0.187** (**0.000**)	**0.027** (**0.001**)	**0.027** (**0.001**)	**0.027** (**0.001**)	**0.116** (**0.000**)	**0.015** (**0.012**)	**0.069** (**0.000**)
Bank controls							
- Total assets	–	**0.974** (**0.000**)	**0.978** (**0.000**)	**0.983** (**0.000**)	–	**0.990** (**0.000**)	–
- Capital ratio	**-1.023** (**0.000**)	–	0.038 (0.089)	0.042 (0.062)	–	**0.048** (**0.036**)	–
District controls							
- Population	**-0.691** (**0.000**)	-0.017 (0.511)	–	**0.028** (**0.000**)	**-0.751** (**0.000**)	–	–
- Real GDP	**0.936** (**0.000**)	0.045 (0.070)	**0.031** (**0.000**)	–	**1.002** (**0.000**)	–	–
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7,116	7,116	7,116	7,116	7,714	8,880	9,934
R^2	0.714	0.977	0.977	0.977	0.524	0.977	0.411

Notes: Combinations of explanatory variables are excluded from the vector of controls as indicated by "–". Boldfaced numbers indicate statistical significance at the 5% level.

of regressors, whereas specifications in table 3.7 are augmented with additional control variables, not used in our preferred empirical model.

As is evident from columns (A) through (D) of table 3.6, the election effect remains significant at the 1% level if any of the four control variables is individually excluded from the set of regressors. In fact, this robustness check suggests that results presented in section 3.6 represent rather conservative estimates of the true effect. As previewed above, once we exclude a bank's total assets – which have a quasi-mechanical relationship with loans and leave hardly any variation in the data unexplained – R^2 drops significantly to 0.714 and the estimate of the election effect increases to 18.7%. A similar picture emerges if all bank controls are excluded from our regression (column (E)) and if model 3.2 is implemented as pure DD without any additional controls (column (G)), with estimates for δ_2 ranging within a 7% and 12% increase in lending. On the other hand, results are slightly weakened if we fail to control for district controls (column (F)), as the election effect drops to 1.5%, which still represents an increase that is statistically significant at the 5% level. Given that our analysis certainly benefits from controlling for variables that may impact lending decisions irrespective of electoral timing, we attach higher credibility to specifications that account for both bank-specific factors and district-level macroeconomic fac-

tors. We are nonetheless pleased that the election effect is found in all of the aforementioned specifications and apparently not just the artificial result of omitted-variable bias or bad control.

To examine whether the further addition of covariates has a dampening effect on our δ_2 estimates, we include a multitude of district-level control variables to the set of regressors. As results in table 3.7 demonstrate, neither information on local public debt, (un)employment, real earnings, real wages, nor firm creation have a notable impact on the election effect, with the latter remaining in a narrow interval of 2.5% to 2.7%. Note that, while firm creation seems to have a statistically significant effect on banks' lending behavior, we decided to exclude this variable from our preferred specification, since it is unavailable for the time before 1998, which would needlessly reduce sample size and preclude the analysis of bank lending for most of the 1990s.

Alternative municipal election types

As argued in section 3.4.2, our main analysis concentrates on municipal elections of *legislative* bodies since our data provides more informational detail for this type of election than for municipal elections of *executive* representatives. This notwithstanding, we additionally test whether savings bank lending also reacts to direct elections of mayors and county representatives, which, as a rule, do not coincide with the election of municipal parliaments.[19] As results in columns (A) and (B) of table 3.8 show, this is indeed the case, even though the effect is a bit smaller – ranging between 2% and 2.5% – than that generated by legislative elections.

Since direct municipal elections were not the norm until the mid-1990s and are still not implemented in Baden-Württemberg's rural counties, these results are based on a lower number of electoral events than was the case in section 3.6. Moreover, they may not be based on the most useful measure of elections of executive representatives because a substantial fraction of the latter is indirectly determined by the regional parliament. As a consequence, the fate of this kind of executive depends on legislative elections, as well. For this reason, we create an alternative variable, $ELEC_{ist}^{M_{XP}}$, that indicates both the occurrence of a legislative election if the executive is appointed by the legislative body, and the occurrence of a direct executive election in case the respective district stipulates this electoral rule. We expect this indicator to be a more appropriate measure of *politically relevant* elections

[19] Note that, since executive elections are usually not held at the state level, the respective indicator variable, $ELEC_{ist}^{M_X}$, is allowed to vary at the district level, which was not the case for the legislative election dummy, $ELEC_{st}^{M}$.

Table 3.7. *Alternative control variables 2: Additional covariates*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Key regressor							
- $ELEC^{M}_{st} * B_b$	**0.027** (**0.001**)	**0.027** (**0.001**)	**0.027** (**0.001**)	**0.027** (**0.001**)	**0.027** (**0.003**)	**0.025** (**0.006**)	**0.025** (**0.007**)
Bank controls							
- Total assets	**0.976** (**0.000**)	**0.977** (**0.000**)	**0.977** (**0.000**)	**0.974** (**0.000**)	**0.974** (**0.000**)	**0.970** (**0.000**)	**0.968** (**0.000**)
- Capital ratio	0.034 (0.134)	0.037 (0.101)	0.036 (0.111)	0.030 (0.184)	0.033 (0.154)	0.022 (0.352)	0.018 (0.435)
District controls							
- Population	-0.001 (0.773)	-0.016 (0.584)	-0.010 (0.773)	**-0.161** (**0.003**)	-0.015 (0.599)	**-0.138** (**0.003**)	-0.236 (0.072)
- Real GDP	**0.049** (**0.037**)	0.040 (0.267)	**0.047** (**0.030**)	0.039 (0.135)	0.021 (0.663)	0.028 (0.328)	-0.059 (0.252)
- Public debt	-0.012 (0.356)	–	–	–	–	–	-0.014 (0.258)
- Employment	–	0.003 (0.948)	–	–	–	–	0.042 (0.680)
- Unemployment	–	–	-0.011 (0.567)	–	–	–	0.020 (0.464)
- Real earnings	–	–	–	**0.137** (**0.000**)	–	–	0.122 (0.144)
- Real wages	–	–	–	–	0.023 (0.585)	–	0.047 (0.501)
- Firm creation	–	–	–	–	–	**0.126** (**0.000**)	**0.092** (**0.006**)
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6,435	7,116	7,116	6,435	6,087	5,394	5,394
R^2	0.978	0.977	0.977	0.978	0.978	0.977	0.978

Notes: Combinations of explanatory variables are excluded from the vector of controls as indicated by "–". Boldfaced numbers indicate statistical significance at the 5% level.

of local executives. According to results displayed in the two final columns of table 3.8, our earlier conclusion that executive politicians also induce electoral lending cycles (that may be mildly less pronounced than those by their legislative counterparts) is confirmed.

3.7.2 Alternative sample compositions

Alternative panel balancedness

Our main results from section 3.6 come from the entire available data sample. As mentioned in section 3.4.1, however, our bank data is quite unbalanced since many banks entered the sample some time after 1993. At the same time, the sample is subject to mild attrition that is mainly due to mergers of savings banks. To ensure that our results are not driven by

Table 3.8. *Alternative municipal election types*
Dependent variables: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)			
	(A)	(B)	(C)	(D)
Key regressor				
- $ELEC_{ist}^{M_X} * B_b$	**0.022** **(0.014)**	**0.024** **(0.010)**	–	–
- $ELEC_{ist}^{M_{XP}} * B_b$	–	–	**0.020** **(0.011)**	**0.021** **(0.017)**
Bank controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Time FE	Yes		Yes	
Time trends		Yes		Yes
Bank type FE	Yes	Yes	Yes	Yes
N	7,116	7,116	7,116	7,116
R^2	0.977	0.977	0.977	0.976

Notes: The index M_X stands for direct municipal elections of the executive, and M_{XP} denotes politically relevant elections for executive representatives. Boldfaced numbers indicate statistical significance at the 5% level.

these data characteristics, we re-estimate DD model 3.2 on a completely balanced panel. This alternative sample consists of 159 savings banks (2,703 bank years) and 104 cooperative banks (1,678 bank years) and represents roughly 25% of the original sample.[20]

As is evident from column (A) of table 3.9, the election effect proves to be immune to even such extreme reductions in sample size: Based on banks that remained in the sample from 1993 to 2009, the estimated increase in savings bank lending amounts to 3% and is statistically significant at the 1% level.

Alternative time intervals

To investigate the stability of the election effect across time, we divide the whole sample into two panel sets of equal length, with the first covering the years between 1993 and 2001 and the second covering the time between 2001 and 2009. As can be seen in columns (B) and (C) of table 3.9, lending increases of roughly 2.5% occur in both the 1990s and the 2000s and are, hence, unlikely to be driven by any temporal anomalies not captured by

[20]Note that in additional robustness checks not reported here, we have also experimented with earlier and later cut-off points than 1993 to create balanced panels. Estimations based on these, however, have not yielded any different conclusions.

Table 3.9. *Alternative sample compositions*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Key regressor							
- $ELEC^{M}_{st} * B_b$	**0.030** (0.001)	**0.024** (0.036)	**0.027** (0.009)	**0.020** (0.033)	0.014 (0.079)	**0.044** (0.000)	**0.028** (0.001)
Bank controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank type FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balanced panel	Yes	No	No	No	No	No	No
Year range	93-09	93-01	01-09	93-09	93-09	93-09	93-09
Dropped state	None	None	None	BW	BY	HS	RP
N	3,945	2,767	4,353	4,730	4,077	6,002	6,304
R^2	0.970	0.969	0.978	0.975	0.984	0.975	0.977

Notes: Abbreviations are as follows: Years: 93=1993; 01=2001; 09=2009. States: BW=Baden-Württemberg; BY=Bavaria; HS=Hesse; RP=Rhineland-Palatinate. Boldfaced numbers indicate statistical significance at the 5% level.

our set of covariates and time dummies. The fact that the election effect for the later time interval is estimated with increased precision, is most likely attributable to the higher number of bank years: Recall that the representativeness of Hoppenstedt's cooperative bank data greatly improves during the early 2000s, as many smaller banks whose balance sheets were not collected before, are added to the sample around this time. On this account, it is encouraging that the election effect is robust to this kind of sample selectivity as well.

Alternative regional compositions

Finally, to assess the generality of results, we apply model 3.2 to four subsamples that individually exclude one of the four states our main sample consists of. As columns (D) through (G) of table 3.9 show, the strength of the election effect appears to be sensitive to the choice of states used for analysis. While results are broadly unchanged by the exclusion of Rhineland-Palatinate, the election effect is weakened if either Baden-Württemberg or Bavaria are dropped from our sample, suggesting that politically induced lending is especially pronounced in the two latter states. On the other hand, the estimate of δ_2 is considerably higher if Hesse is omitted from the working sample.[21]

[21] While this may reflect the absence of electoral lending cycles in Hesse, it is also possible that this observation is simply an artefact of Hesse's idiosyncratic time trends mentioned in section 3.4.4, which may disqualify Hesse as a viable control state.

Finding an explanation for these cross-state differences is certainly desirable, as this may further our understanding of the interplay of institutional features in limiting the extent of politically induced lending. One potential reason for the observed pattern could be found in the role of incumbent dominance, which – as argued in section 3.6 – may be a precondition for the ability of politicians to manipulate bank policies. Considering that electoral contests in Bavaria and Baden-Württemberg are historically much less contested than those in Hesse, the differences among states may well be driven by this feature.[22] Other possible explanations for this regional heterogeneity may, for instance, be found in the aforementioned differences in electoral laws, minor variations in states' savings bank regulation, partisan effects, or – as is the case for Hesse – the concentrated presence of private banks that changes the operational environment for savings and cooperative banks.

A meaningful cross-state analysis that discriminates among such hypotheses, however, proves to be infeasible with just 4 states at our disposal. It is not only the low number of states that poses problems, but also the limited cross-state variability of certain features that is cause for concern: Recall that the states chosen for analysis are all from the southern part of Germany. While this regional proximity improves the identification of the election effect since DD relies on states being on reasonably similar time trends, it is unlikely to provide sufficient variation in institutional and other state-specific characteristics. For example, it would be impossible to tell whether the differences across states are driven by electoral competition or, say, partisan effects because the two states with the highest incumbent dominance – Bavaria and Baden-Württemberg – are also historically dominated by the same party.[23] Hence, distinguishing between these alternative hypotheses would require the addition of states which are dominated by parties that represent the opposite of the political spectrum.[24] For these reasons, we are currently augmenting our dataset with several states from

[22] This pattern holds for all indicators we constructed to capture the degree of electoral competition in a district. To give one example, 19% of all municipal elections of legislative bodies in Hesse are lost by the incumbent party. While this number is only a bit lower for Rhineland-Palatinate (17%), Baden-Württemberg (7%) and Bavaria (4%) are characterized by considerably higher party dominance.

[23] Recall from section 3.4.4 that, while Bavaria's ruling party, *CSU*, is legally independent from its sister party, *CDU*, which dominates politics in Baden-Württemberg, they are generally considered the same organizational entity, covering the conservative spectrum of German politics.

[24] An obvious candidate appears to be the state of *North Rhine-Westphalia* whose political system is largely dominated by left-of-center *SPD*.

other parts of Germany to obtain a more representative sample than currently available. Besides enabling us to study cross-state heterogeneity, this step will also determine whether electoral lending cycles are a general phenomenon or if our results are instead driven by regional outliers.

3.7.3 Alternative assumptions on the error structure

The final robustness analysis presented here, assesses the stability of results to varying modes of statistical inference. The growing literature on cluster-robust inference (see Bertrand *et al.*, 2004; Angrist and Pischke, 2009; and Cameron and Miller, 2010 for an overview) highlights the importance of accounting for potential serial correlation and regional clustering in panel data. Both phenomena implicate a violation of one of the main assumptions traditionally imposed when working with cross-sectional data: the independence of observations. While OLS will still be consistent, precision is likely overestimated if these issues are ignored.

Serial correlation

Serially correlated errors, ϵ_{ist}, are a typical problem of panel data applications and there is little reason to believe that the present study is an exception. Formally, $\text{Cor}(\epsilon_{ist}, \epsilon_{isu}) = \rho_\epsilon \neq 0$, for $t \neq u$, where ρ_ϵ denotes the intraclass correlation of the error. That is, the individual (here: bank i) is thought of as a cluster whose observations over time are not independent of one another. A rough estimate of ρ_ϵ – the average autocorrelation over 5 lags of OLS residuals coming from model 3.2 – equals 0.412 and suggests that our data is indeed subject to substantial serial correlation. For this reason, and in line with Bertrand *et al.* (2004), Khwaja and Mian (2005), and Cameron and Trivedi (2010), our preferred empirical specification already corrects for serial correlation by clustering standard errors on the bank level instead of the bank-year level.

An alternative way of dealing with autocorrelated errors consists of estimating model 3.2 with a random-effects (RE) specification (see Cameron and Trivedi, 2010). The individual-effects model provides the following rationale for serial correlation: If the error $\epsilon_{ist} = \alpha_i + r_{ist}$, then the presence of a bank-specific effect, α_i, induces correlation over time, even if the idiosyncratic, r_{ist}, is iid. If these assumption on the error structure are correct and as long as α_i is truly random, RE is more efficient than OLS, which is why we present results for regression 3.2, fitted by FGLS in column (B) of table 3.10. When compared to our baseline specification in column (A), the estimated election effect is virtually unchanged by this alternative approach

Table 3.10. *Alternative error assumptions*
Dependent variable: Log loans

Explanatory variables	**Regression coefficients** (Empirical p-values in brackets)			
	(A)	(B)	(C)	(D)
Key regressor				
- $ELEC^{M}_{st} * B_b$	**0.027** **(0.001)**	**0.028** **(0.000)**	**0.022** **(0.000)**	**0.027** **(0.002)**
Bank controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	n.a.	Yes
Time FE	Yes	Yes	Yes	Yes
Bank type FE	Yes	Yes	n.a.	Yes
Estimator	OLS	RE	FE	OLS
Cluster	Bank	Bank	Bank	District
N	7,116	7,116	7,116	7,116
R^2	0.978	0.976	0.954	0.978

Notes: Abbreviations are as follows: OLS=Ordinary Least Square; RE=Random Effects; FE=Fixed Effects (Within estimator). "Cluster" indicates whether standard errors are clustered on the bank or the district level. "n.a." denotes that time-constant variables are omitted. Boldfaced numbers indicate statistical significance at the 5% level.

of correcting for serial correlation. The estimated standard deviation of the individual effect, $\hat{\sigma}_\alpha$, equals 0.104 and is roughly as large as that of the idiosyncratic error, $\hat{\sigma}_r = 0.120$. Furthermore, intraclass correlation is estimated to equal 0.427, which is in line with our ad-hoc estimate, mentioned above.

Of course, the RE estimator is only consistent if α_i is uncorrelated with regressors. If we wish to relax this assumption, the individual effect needs to be eliminated with a fixed-effects (FE) specification that only relies on variation over time. Even though we are not particularly worried about correlated effects, results for model 3.2 when fitted by a within estimator are presented in column (C) of table 3.10. While the election effect appears to be somewhat smaller than before (amounting to 2.2%), it is still very precisely estimated. We take this as encouraging evidence that our results survive, even when identification is based on within-variation alone.

Regional clustering

If data has a group structure, independence may not only be violated for observations of one individual bank over time, but also across banks that are part of the same regional cluster. In this case, ϵ_{ist} will contain some variation that is likely to be common to banks in the same geographical

area and year, for instance, a regional business cycle. An obvious solution to this problem is to correct standard errors for clustering on the geographical level, these region-year shocks are most likely to occur.

In our context, there are two candidates for such regional clusters: the municipal district (which typically contains one savings bank and one to four cooperative banks) and the state. While clustering standard errors on the district level is straightforward, this methodological fix is infeasible for the state level, since robust inference requires a larger number of clusters than is available in our data.[25] As the non-trivial issue of inference with few clusters is still under study in the literature, we content ourselves with implementing what is methodologically feasible at present and provide estimates of model 3.2 with standard errors clustered at the district level in column (D) of table 3.10. Evidently, precision is only marginally reduced – with standard errors rising from 0.0079 to 0.0087 – suggesting that district-level clustering is not much reason for concern.

3.8 Conclusion

We exploit a particularity in the German public banking system where local politicians are by law actively involved in the management of savings banks' lending decisions to test the hypothesis that incumbent politicians pursue policies that benefit their re-election probability but might be socially sub-optimal. Our identification strategy relies on the fact that we should only observe politically motivated lending around election years, only in municipalities in which elections are held at this point in time, and – importantly – only for savings banks that are by law politically connected. Econometrically, we conduct difference-in-difference (DD) as well as triple-difference (DDD) estimation embedded in a fixed-effects panel data setup.

We use a unique, largely hand-collected dataset that combines detailed information on German municipal elections, macro-economic data on the district level, and balance-sheet information on bank lending. We find that savings banks systematically extend more credits in pre-election periods. This effect is not only statistically significant but also economically relevant. Given that savings banks are holding the largest market share in the private customer deposit market and they are the most important lender to the *Mittelstand* (SMEs) that is considered the backbone of the German

[25] Note that this problem cannot by solved by simply increasing the number of states in our dataset, since the natural limit will be the total number of 16 German states, whereas the literature suggests a minimum number of around 40 to 50 groups (see Angrist and Pischke, 2009 and Cameron and Miller, 2010).

economy, it is potentially worrisome to find their policies distorted. The pre-election excess-lending effect is robust to various specifications. Importantly, this effect is not present with cooperative banks that are very similar to savings banks but that lack the political connectedness of savings banks. Furthermore, and in line with our hypothesis, lending cycles only occur with *municipal* but not with state or federal elections. In addition, we find weak evidence for overly prudent lending policies after elections, consistent with a binding credit constraint that banks face as they have to make up for excessive pre-election lending.

Our evidence suggests that the *ability* of politicians to influence the lending policies of their respective savings bank depends on the dominance of the incumbent party, since electoral cycles are less common in districts with frequent changes in power and historically close elections. Whether *incentives* to induce socially inefficient lending increases are influenced by the contestedness of the current electoral campaign, is inconclusive and may become answerable once test power increases in the course of gradually extending our dataset for further research: On the one hand, we are currently adding municipal election results for additional German states to better understand the interplay of various institutional features in limiting the extent of politically induced lending cycles. On the other hand, we are hopeful to eventually be granted access to more detailed credit-contract data that will allow us to test more nuanced hypotheses such as prediction 5, as this will shed more light on the social costs of interfering with bank policies for political gain.

3.A Appendix: Predictions 2 to 4 with model 3.1

As argued in section 3.6, we believe DD model 3.2 to provide improved identification of the causal election effect over model 3.1. However, for robustness, this appendix also tests predictions 2 through 4 with model 3.1 that only exploits differences between savings banks. Overall, our main results are confirmed, even though effect sizes appear to be a bit smaller.

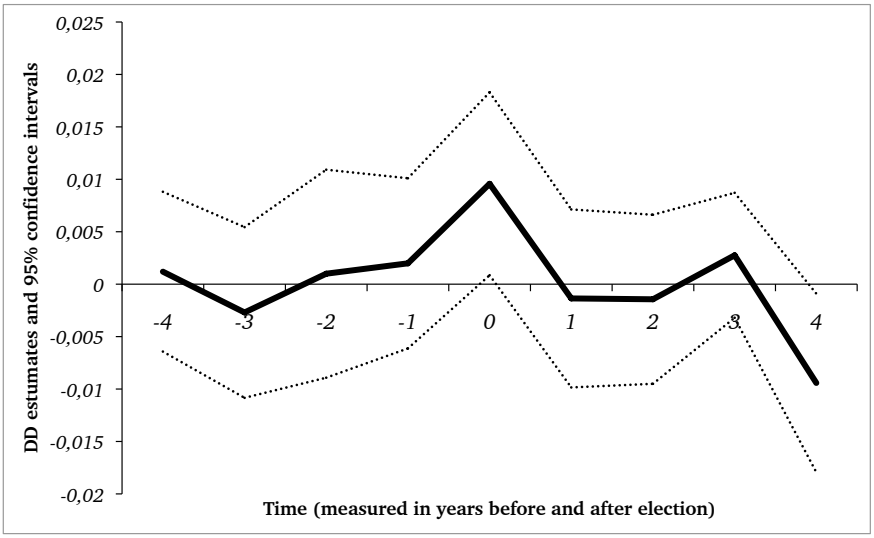
As is evident from table 3.11, municipal elections are once again the only type of election with a statistically significant impact on savings bank lending, which confirms prediction 2, that state-level and federal elections should play less of a role. The electoral lending cycle, as estimated without a control group of politically unconnected banks, is visualized in figure 3.4 and roughly follows the pattern depicted in figure 3.3, with a significant spike just before an election and weak evidence for credit crunching in the years thereafter. The evidence for the effects of electoral competition is a little more mixed (see table 3.12). As before, current contestedness does not seem to systematically influence politically induced lending. Now, this result even holds if incumbent dominance is controlled for. As far as the latter is concerned, there is still evidence that the ability to induce lending cycles increases with the decisiveness of the past electoral victory, even though the significance level has increased to roughly 10%. Note that, while results in columns (A) and (D) of table 3.12 do no longer suggest that *general* incumbent stability facilitates electoral cycles, this result does indeed survive if we run separate regressions – not presented here – on two subsamples that are stratified by overall competitiveness of the electoral district.

Table 3.11. *Results for prediction 2 (Model 1)*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)			
	(A)	(B)	(C)	(D)
Key regressors				
- $ELEC^M_{st}$	–	–	–	**0.021** **(0.000)**
- $ELEC^S_{st}$	-0.000 (0.942)	0.003 (0.362)	–	0.010 (0.102)
- $ELEC^F_t$	–	–	-0.003 (0.294)	-0.003 (0.350)
Bank controls	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Time FE	Yes			
Time trends		Yes	Yes	Yes
Banks in sample	SavB	SavB	SavB	SavB
N	3,083	3,083	3,083	3,083
R^2	0.970	0.969	0.969	0.969

Notes: Results are for our preferred empirical specification (see text). Key regressors are $ELEC^M_{st} * B_b$, $ELEC^S_{st} * B_b$, and $ELEC^F_t * B_b$, respectively. The indexes M, S, and F denote municipal, state, and federal elections, respectively. Boldfaced numbers indicate statistical significance at the 5% level.

Figure 3.4. *Results for prediction 3 (Model 1)*
Visualization of the lending cycle



Notes: Results are for our preferred empirical specification (see text). The solid line depicts estimates of δ_1 coming from model 3.1 with year fixed effects. Dottet lines indicate the corresponding 95% confidence intervals. Time is measured on the abscissa: A value of zero denotes an election season. Negative and positive values stand for years before and after an election, respectively.

Table 3.12. *Results for prediction 4 (Model 1)*
Dependent variable: Log loans

Explanatory variables	**OLS regression coefficients** (Empirical p-values in brackets)				
	(A)	(B)	(C)	(D)	(E)
DD Interaction					
- $ELEC^M_{st} * D^G_i$	0.009 (0.376)	–	–	0.008 (0.434)	–
- $ELEC^M_{st} * D_{it-1}$	–	0.022 (0.093)	–	–	0.022 (0.108)
- $ELEC^M_{st} * C_{it}$	–	–	-0.003 (0.767)	0.000 (0.969)	0.001 (0.963)
Main effect					
- $ELEC^M_{st}$	0.004 (0.605)	0.007 (0.196)	0.011 (0.071)	0.004 (0.674)	0.006 (0.375)
Bank controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Banks in sample	SavB	SavB	SavB	SavB	SavB
N	3,067	3,059	3,067	3,067	3,059
R^2	0.970	0.970	0.970	0.970	0.970

Notes: Results are for our preferred empirical specification (see text). D^G_i indicates whether political contestedness is *generally* low in the district, bank i operates in. D_{it-1} indicates whether the *preceding* election was close. C_{it} measures the contestedness of the *current* election. Boldfaced numbers indicate statistical significance at the 5% level.

Bibliography

Adams, P., Hurd, M. D., McFadden, D., Merrill, A. and Ribeiro, T. (2003). Healthy, wealthy, and wise? tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, **112**, 3–56. 2, 6, 60, 71

Adda, J., Chandola, T. and Marmot, M. (2003). Comment: Socio-economic status and health: Causality and pathways. *Journal of Econometrics*, **112**, 57–63. 7, 24

Adler, N. E. and Ostrove, J. M. (1999). Socioeconomic status and health: What we know and what we don't. In N. E. A. et al. (ed.), *Socioeconomic Status and Health in Industrialized Nations*, New York Academy of Sciences, pp. 3–15. 8

Alesina, A., Roubini, N. and Cohen, G. D. (1997). *Political cycles and the macroeconomy*. MIT Press. 86

Almond, D. and Currie, J. (2011). Human capital development before age five. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, vol. 4b, *15*, Elsevier. 3, 62

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press. 101, 121, 123

Arellano, M. and Bond, S. R. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, **58** (2), 277–297. 41

Barth, J. R., Caprio Jr., G. and Levine, R. (2001). Banking systems around the globe: Do regulation and ownership affect performance and stability? In F. S. Mishkin (ed.), *Prudential supervision: What works and what doesn't*, University of Chicago Press, pp. 31–96. 88

Berger, L. M., Paxson, C. and Waldfogel, J. (2005). Income and child development. Princeton University Center for Research on Child Wellbeing Working Paper 05-16-FF. 75

Berney, L. and Blane, D. (1997). Collecting retrospective data: Accuracy of recall after 50 years judged against historical records. *Social Science & Medicine*, **45** (10), 1519–1525. 65

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, **119** (1), 249–275. 121

—, Kramarz, F., Schoar, A. and Thesmar, D. (2007). Politicians, firms and the political business cycle: Evidence from France. *mimeo*. 87

Cameron, A. C. and Miller, D. L. (2010). Robust inference with clustered data. *mimeo*. 121, 123

— and Trivedi, P. K. (2010). *Microeconometrics using Stata*. Stata Press. 121

Caprio Jr., G. and Martinez Peria, M. S. (2000). Avoiding disaster: policies to reduce the risk of banking crises. In E. Cardoso and A. Galal (eds.), *Monetary policy and exchange rate regimes: Options for the Middle East*, The Egyptian Center for Economic Studies, pp. 193–230. 88

Case, A., Fertig, A. and Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, **24**, 365–389. 62, 75

—, Lubotsky, D. and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, **92** (5), 1308–1334. 62, 75

— and Paxson, C. (2008). Height, health, and cognitive function at older ages. *American Economic Review: Papers & Proceedings*, **98** (2), 463–467. 66

Conti, G., Heckman, J. and Urzua, S. (2010). The education-health gradient. *American Economic Review*, **100** (2), 234–238. 81

Contoyannis, P., Jones, A. M. and Rice, N. (2004). The dynamics of health in the british household panel survey. *Journal of Applied Econometrics*, **19** (4), 473–503. 41

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, **47** (1), 87–122. 62

— (2011). Inequality at birth: Some causes and consequences. *American Economic Review: Papers & Proceedings*, **101** (3), 1–22. 62, 66, 82

Cutler, D. M. and Lleras-Muney, A. (2006). Education and health: Evaluating theories and evidence. NBER Working Paper 12352. 82

—, — and Vogl, T. (2011). Socioeconomic status and health: Dimensions and mechanisms. In S. Glied and P. C. Smith (eds.), *The Oxford Handbook of Health Economics*, Oxford University Press, p. forthcoming. 1, 8, 59

Dinç, I. S. (2005). Politicians and banks: Political influences on government-owned banks in emerging markets. *Journal of Financial Economics*, **77** (2), 453–479. 4, 88, 89, 106

Faccio, M. (2006). Politically connected firms. *American Economic Review*, **96** (1), 369–386. 87

—, Masulis, R. W. and McConnell, J. J. (2006). Political connections and corporate bailouts. *Journal of Finance*, **61** (6), 2597–2635. 87

Fisman, R. (2001). Estimating the value of political connections. *American Economic Review*, **91** (4), 1095–1102. 87

Florens, J.-P. (2003). Comment: Some technical issues in defining causality. *Journal of Econometrics*, **112**, 127–128. 7, 26

Garrouste, C. and Paccagnella, O. (2010). Data quality: Three examples of consistency across SHARE and SHARELIFE data. In M. Schröder (ed.), *Retrospective data collection in the Survey of Health, Ageing amd Retirement in Europe*, Mannheim Research Institute for the Economics of Aging (MEA). 65

Geweke, J. (2003). Comment: Econometric issues in using the ahead panel. *Journal of Econometrics*, **112**, 115–120. 7, 26

Goldsmith, N. (2001). Social inequalities in health: Disentangling the underlying mechanisms. In M. Weinstein and A. Hermalin (eds.), *Strengthening the dialogue between epidemiology and demography*, vol. 954, New York Academy of Sciences, pp. 118–139. 5, 8

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37** (3), 424–38. 19

— (2003). Comment: Some aspects of causal relationships. *Journal of Econometrics*, **112**, 69–71. 7, 25

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, **80** (2), 223–255. 2, 10, 16, 25, 61, 69

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, **84** (3), 622–641. 105

Gustman, A. L., Steinmeier, T. L. and Tabatabai, N. (2010). What the stock market decline means for the financial security and retirement choices of the near-retirement population. *Journal of Economic Perspectives*, **24** (1), 161–182. 40

Hausman, J. A. (2003). Comment: Triangular structural model specification and estimation with application to causality. *Journal of Econometrics*, **112**, 107–113. 7, 20, 25

Heckman, J. J. (1981a). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C. F. Manski and D. McFadden (eds.), *Structural Analysis of discrete data and econometric applications*, MIT Press. 70

— (1981b). Statistical models for discrete panel data. In C. F. Manski and D. McFadden (eds.), *Structural Analysis of discrete data and econometric applications*, MIT Press. 59

— (2000). Causal parameters and policy analysis in economics: A twentieth century perspective. *Quarterly Journal of Economics*, **114** (1), 45–97. 20

— (2003). Comment: Conditioning, causality and policy analysis. *Journal of Econometrics*, **112** (73–78). 7, 24, 25, 26

Heiss, F. (2011). Dynamics of self-rated health and selective mortality. *Empirical Economics*, **40** (1), 119–140. 43, 69, 70, 74

Hoover, K. D. (2001). *Causality in macroeconomics*. Cambridge: Cambridge University Press. 20

— (2003). Comment: Some causal lessons from macroeconomics. *Journal of Econometrics*, **112**, 121–125. 7

Hurd, M. and Kapteyn, A. (2003). Health, wealth, and the role of institutions. *Journal of Human Resources*, **38** (2), 386–415. 42

Johnson, S. and Mitton, T. (2003). Cronyism and capital controls: Evidence from Malaysia. *Journal of Financial Economics*, **67** (2), 351–382. 87

Khwaja, A. I. and Mian, A. (2005). Do lenders favor politically connected firms? rent provision in an emerging financial market. *Quarterly Journal of Economics*, **120** (4), 1371–1411. 88, 121

La Porta, R., Lopez de Silanes, F. and Shleifer, A. (2002). Government ownership of banks. *Journal of Finance*, **57** (1), 265–301. 88

MacRae, D. C. (1977). A political model of the business cycle. *Journal of Political Economy*, **85** (2), 239–263. 3, 86

Mazzonna, F. (2011). The long-lasting effects of family background: A European cross-country comparison. Munich Center for the Economics of Aging Working Paper 245-2011. 75

Meer, J., Miller, D. L. and Rosen, H. S. (2003). Exploring the health–wealth nexus. *Journal of Health Economics*, **22** (5), 713–730. 15

Michaud, P.-C. and van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics*, **27**, 1312–1325. 5, 16, 41

Mitchell, D. M. and Willett, K. (2006). Local economic performance and election outcomes. *Atlantic Economic Journal*, **34** (2), 219–232. 86

Nordhaus, W. D. (1975). The political business cycle. *Review of Economic Studies*, **42** (2), 169–190. 3, 86

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press. 20

Poterba, J. M. (2003). Comment: Some observations on health status and economic status. *Journal of Econometrics*, **112** (65–67). 7, 24, 25

Ramalho, R. (2004). The effects of anti-corruption campaign: Evidence from the 1992 presidential impeachment in Brazil. *mimeo*. 87

Robins, J. M. (2003). Comment: General methodological considerations. *Journal of Econometrics*, **112** (89–106). 7

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. journal estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66** (5), 668–701. 20

— and Mealli, F. (2003). Comment: Assumptions allowing the estimation of direct causal effects. *Journal of Econometrics*, **112** (79–87). 7

Sapienza, P. (2004). The effects of government ownership on bank lending. *Journal of Financial Economics*, **72** (2), 357–384. 87

Schneider, C. (2010). Fighting with one hand tied behind the back: Political budget cycles in the West German states. *Public Choice*, **142** (1), 125–150. 86

Shleifer, A. and Vishny, R. W. (1994). Politicians and firms. *Quarterly Journal of Economics*, **109** (4), 995–1025. 87

Sims, C. A. (1972). Money, income, and causality. *American Economic Review*, **62** (4), 540–552. 19

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relationship between health and economic status. *Journal of Economic Perspectives*, **13** (2), 145–166. 5, 8, 59

— (2005). Consequences and predictors of new health events. In D. A. Wise (ed.), *Analyses in the Economics of Aging*, National Bureau of Economic Research, pp. 213–240. 16

— (2009). Reconstructing childhood health histories. *Demography*, **46** (2), 387–403. 62, 65, 70

—, Shen, Y., Strauss, J., Zhe, Y. and Zhao, Y. (2010). The effects of childhood health on adult health and SES in China. RAND Working Paper 809. 75

Stowasser, T., Heiss, F., McFadden, D. and Winter, J. (2012). "Healthy, wealthy and wise?" revisited: An analysis of the causal pathways from socio-economic status to health. In D. A. Wise (ed.), *Investigations in the Economics of Aging*, University of Chicago Press, forthcoming. 59, 60, 61, 71