Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

FAKULTÄT
FÜR WIRTSCHAFTS- UND
SOZIALWISSENSCHAFTEN

Identifying causal effects: Essays on empirical economics

* * * * * * * * * * * * * * * *

Universität Hamburg
Fakultät für Wirtschafts- und Sozialwissenschaften

Dissertation
Zur Erlangung eines Doktors der
Wirtschafts- und Sozialwissenschaften
"Dr. rer. pol."
(gem. PromO vom 18. Januar 2017)

* * * * * * * * * * * *

vorgelegt von

Jascha Tutt
aus Bergisch Gladbach

Hamburg, den 25.1.2018

* * * * * * * * * * * * * * * *

Vorsitzender: Prof. Dr. Ulrich Fritsche
Erstgutachter: Prof. Dr. Grischa Perino
Zweitgutachter: Prof. Dr. Andreas Lange
Zweitbetreuer: Prof. Dr. Michael Berlemann

Datum der Disputation: 14.05.2018

Acknowledgements

nanos gigantum humeris insidentes

First and foremost I am grateful to my supervisors and co-authors Michael Berlemann and Grischa Perino for all their valuable support, their guidance, and encouragement to pursue my own ideas. Sharing their experience and insights with me, greatly helped my to determine my own path through my PhD, academia, and teaching.
I am also thankful for the chance to meet so many inspiring and supportive people, colleagues, and friends at Hamburg University, Helmut-Schmidt University, and University of California Berkeley. With their intellectual exchange, coffee breaks, and occasional leisure-time activities they made my “research endeavour” more than worthwhile. Thank you Jan, Johannes, Marc, and Philipp. It is nice to have great colleagues, but it is even better to have colleagues that become friends. Special thanks go to my co-author Max Steinhardt from who I learned a lot about statistical inference.
Last but not least, I want to thank my family: my parents, Ulla, Michael, and Heinz, as well as my siblings. Without their support I would have hardly got into the position to ever write this dissertation. I am grateful to all of them as well as to my friends, for all the fruitful discussions and for giving me strength when I was struggling. Most importantly, I would like to thank Stephanie for always being able to find the right words and supporting me during the last years of my dissertation.
Finally, I would like to note that all the people and institutions that supported my research papers are acknowledged at the beginning of each respective section in this dissertation.

"I cannot live without brain-work. What else is there to live for?"

(Sherlock Holmes in: The Sign of Four)

Contents

Foreword

"From a drop of water", said the writer, "a logician could infer the possibility of an Atlantic or a Niagara without having seen or heard of one or the other. So all life is a great chain, the nature of which is known whenever we are shown a single link of it. Like all other arts, the Science of Deduction and Analysis is one which can only be acquired by long and patient study nor is life long enough to allow any mortal to attain the highest possible perfection in it."

(Sherlock Holmes, A Study in Scarlet)

Becoming an empirical economist, i.e. a person that is interested in drawing useful inference from data on people (Angrist and Pischke, 2009), is a comprehensive journey. On the one hand, an empirical economist is like a statistician applying the tools and techniques of statistical inference. On the other hand, an empirical economist is like a detective identifying probable culprits and collecting evidence for conviction. While I initially devoted much of my time as a PhD student to the acquisition of statistical tools and techniques, my first research project soon taught me the importance of detectives' work in empirical research. Society is a complex system and identifying the cause of an observed effect can be very challenging. Usually, this impedes the identification of a single explanation because multiple alternatives could be responsible for the observed effect. Moreover, the identified explanations are often intertwined like the strings of a cobweb. Therefore, the first challenge of empirical research is to untangle the paths of probable causal chains. During my time as a PhD student I learned that this task is not statistical in nature, but more closely related to what I believe to be detective work. Reading about detective work, I rediscovered Sir Arthur Conan Doyle's famous private detective Sherlock Holmes and his Science of Deduction and Analysis. To my surprise I found that there is much to be learned from him.

Sherlock Holmes is undeniably someone proficient in the Science of Deduction and Analysis. In his cases he always convicts the culprit and most often his analysis seems more like magic than science. However, it is his reasoning that makes him a good detective and much can be learned form it. For instance, Uchii (2010) argues that Holmes is a logician whose analytic thinking is based on probabilistic reasoning. While this suggests that he has knowledge about probability theory, statistics, and logic, it does not explain how he identifies a single cause of an observed outcome. According to Anderson, Schum and Twining (2005) or Carson (2009), Holmes uses abductive reasoning to solve his cases.[1] The idea behind abductive reasoning is to infer from a single outcome all probable effects that might have caused it. Although this sounds challenging, Sherlock Holmes often points out to his companion Dr. Watson how to learn the technique. The first lesson is to train one's imagination. After all, having to come up with

[1]The Stanford Encyclopedia describes abductive reasoning as inference to the best explanation (last accessed: 17.1.2018).

a number of alternative explanations does require some possession of ingenuity. The second lesson is to carefully observe all available evidence and assign probabilities to all imagined explanations accordingly. Over time and with more information one explanation should become more probable than the others presenting itself as the solution. Indeed, it is this technique that let Sherlock Holmes derive solutions that initially seem very unlikely but turn out to be correct.

Now, how does Sherlock Holmes' abductive reasoning help to do good empirical research? First and foremost it trained me to carefully study the surrounding environment in which the outcome of interest is nested. I soon realized that such careful observation can be very helpful in selecting appropriate research design, data, and statistical technique. Moreover, trying to imagine all probable causes of an outcome inspired me to think about counterfactuals already at the beginning of my research inquiry. It turned out that having an idea about counterfactuals makes it easier to choose and formulate a hypothesis for later statistical testing. Finally, abductive reasoning helped me to remember that the tested hypothesis only represents one probable (causal) path. This is especially helpful as it trains a certain humbleness when it came to interpreting results.

As the reader might imagine, it is a cumbersome and sometimes frustrating task to observe surroundings and collect evidence. Yet, in the end it is worth the time and I often found that "the little things are infinitely the most important" (Holmes in: Case of Identity).

Empirical research does, however, not stop with the identification and selection of a probable cause. Rather it marks the beginning of the statistician's work. Actually, one of the most important goals for an empirical researcher is to establish a causal link between two events. While in the case of Sherlock Holmes his most probable solution equals the "truths", in reality inferring causality from available data is far more challenging. With respect to statistical inference, causality is not a binary status but a gradation of different shades of gray. That is, causality can only be inferred when an empirical study shows a high degree of internal validity. Shadish, Cook and Campbell (2002) argue that a high degree of internal validity can be inferred if (i) the cause precedes the effect, (ii) the cause and the effect are statistically related, and (iii) there are no plausible alternative explanations for the statistical relation.

Whether all of the proposed conditions for causal inference are met in practice critically depends on available data and research design. For instance, a randomized experiment with a clearly defined treatment and control group usually meets the conditions for causal inference. In this ideal case all of the measured difference between the two groups (the observed effect) can be attributed to the treatment (the cause). However, it is the nature of economics, as study of society, that makes many urgent questions difficult to answer through experiments or respective experiments difficult to implement. Assume you want to run an experiment testing whether a certain event causes a financial crisis, like the one of 2007/08. As you probably agree, the experimenter faces the difficult task to convince a government treating its economy with an event that is likely to cause a financial crisis. Moreover, the experimenter would have to find a twin economy that can serve as a control group. Even if the experimenter would succeed in her task, she would only have two observations and with the high degree of interconnectedness of the current world economy it is very doubtful that they are independent from each other. Hence, the empirical economist often has to rely on non-experimental data. In many circumstances this poses a serious threat to internal validity making rigourous causal inference impossible because there are just too many alternative causal paths that could yield the outcome under study (see condition (iii)). Due to the drawback of non-experimental data, some authors plead for careful research design and the use of natural experiments to strengthen the degree of internal validity when using non-experimental data (Angrist and Pischke, 2010). But experiments – designed or natural – also have drawbacks. Usually, the generalizability (i.e. external validity) of a derived result is weak even if the estimated effect is indeed causal. A critical person might question the use of results that are not transferable to other settings.[2]

In this dissertation the reader will experience how diverse the art and science of causal inference can be. It will become clear that there is no blueprint procedure to conduct empirical economic research. For each setting, the researcher must start from scratch, successively reconstruct intertwined (causal) paths, derive and formulate clear hypotheses, chose appropriate research designs, select adequate statistical tools, and carefully interpret the results. To sum up, empirical research is cumbersome work, but also very rewarding. With careful observation and the diligent use of statistical inference, there is much that can be learned. Be it the discovery of new (causal) mechanisms or the evaluation of policy interventions.

[2] Campbell and Stanley (1966) proposed that there is a trade-off between external and internal validity. While experiments are strong on internal validity, it is difficult to extrapolate their results to different (social) conditions. The opposite applies to surveys. Surveys are usually able to pick up much of the forces in the surrounding (social) environment but can only rarely yield results that are able to isolate a single effect providing a high degree of internal validity.

1 Initio

"I have devised seven different explanations, each of which would cover the facts as far as we know them. But which of these is correct can only be determined by the fresh information which we shall no doubt find waiting for us."

(Sherlock Holmes in: The Adventure of the Cooper Beeches)

Empirical economic research uses empirical evidence to test hypotheses and statistical inference to uncover general rules. However, very often several rules or causal mechanisms exist that can equally well explain the investigated outcome. This can be problematic whenever statistical inference does not yield convincing results, i.e. the degree of the study's internal validity is low. Although many different statistical tools and techniques have been developed to increase the degree of internal validity, in practice it remains difficult to claim causality. The quality of causal inference in many studies in empirical economics is heatedly discussed and no panacea is in sight. This dissertation includes several pieces of applied empirical work. In order to assist the reader's ability to judge empirical work, but also to better understand the power and pitfalls of empirical economics in general, I will start with a reproduction of a recent academic debate on the discipline's condition.

The Journal of Economic Perspectives (JEP) frequently holds symposia on urging topics in economic science. In its spring 2010 issue, the journal published a series of articles discussing the state of applied research in economics with special emphasis on causal inference. The title of the symposium "Con out of Econometrics" refers to Edward Leamer's prominent appeal for more credibility in applied economic research published 27 years earlier (Leamer, 1983). Leamer (1983), p.37 observed: "Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously." While his critique on the state of empirical research was adequate back then, empirical economics has since experienced a "credibility revolution" (Angrist and Pischke, 2010). Angrist and Pischke (2010) credit this revolution to the discipline's increasing focus on the quality of research designs. Exploiting natural and quasi experiments allows for powerful causal inference with very simple econometric techniques.

While most fields of applied microeconomics went a far distance in this revolution, other fields lag behind. According to Angrist and Pischke (2010) more work could be done in empirical macroeconomics and industrial organization. As it is suitable for a symposium, their claim does not remain unchallenged. For instance, Sims (2010), representing the field of macroeconomics, warns that "economics is not an experimental science and cannot be"(Sims, 2010, p.59). The narrow research design proposed by Angrist and Pischke (2010) might not be useful, or even harmful when it comes to the greater picture of resource allocation in an economy, as they usually only estimate an average local treatment effect. Moreover, simple econometric methods encourage empirical economists to devote less attention to the characteristics of the data such as heteroskedasticity, serial correlation and non-linearities. Sims (2010) worries that "there is a great deal going on in the data that our linear model is missing", suggesting that it would be better to account for characteristics more carefully. Additionally, Nevo and Whinston (2010) point out that estimating treatment effects based on actual or quasi experiments is not the panacea to all credibility issues in empirical work. They argue that structural modeling is much more important in the field of industrial organization. Specifically, the "attempt to use data to identify the parameters of an underlying economic model, based on models of individual choice or aggregate relations derived from them" (Nevo and Whinston, 2010, p. 69,70) also increases the degree of causal inference. Structural models are more important when addressing problems in industrial organization, as changes in the surrounding environment are usually believed to be heterogenous. For instance, data availability differs between policymakers and researchers, underlying models are more complicated, and strategic interaction between agents is of first-order importance. Eventually, Nevo and Whinston (2010) conclude that "any serious empirical researcher should build a toolkit consisting of different methods, to be used according to the specifics of the question being studied and the available data. That this should not be an either-or proposition seems quite obvious to us."

So, what can be learned from the discussion in the symposium? The rather short answer is: a lot! First and foremost, careful research design in combination with quasi-experimental variation can lead to powerful causal inference without the use of nontransparent and very complex econometric techniques. Secondly, not all economic questions – and maybe even the most important ones – can be answered with the help of quasi-experimental evidence. Moreover, the linear model might not always provide the ability to answer the most interesting questions about the data. Last but not least, different fields in economics require different methods. Applied work in industrial organization might require a higher degree of structural modeling while dynamics are more important in macroeconomics.

The discussion in the JEP's symposium "Con out of Econometrics" greatly influenced the intent of this dissertation. In my opinion, the statement by Nevo and Whinston (2010) that the right approach to credible empirical research is not an either-or question seems very sensible. Aiming to be a "serious" empirical economist, I have acquired a toolkit covering different tools and techniques suitable to address problems in many fields of empirical economics. The three different papers comprising this dissertation respect the caveats presented above. Topics in each respective study can roughly be allocated to the fields of empirical microeconomics, macroeconomics, and industrial organization. Hence, empirical settings and the use of econometric tools differ between studies. While this requires the reader to engage with each study anew, it also makes this dissertation more extensive in the coverage of topics and methods in empirical economics. Naturally, each study has its own challenges and the "degree" of causal inference varies. However, adhering credit to the discussion in the symposium, each study follows the proposed advice as best as possible to strengthen causal inference. Ultimately, it is the overall aim of this dissertation to present applied work that helps "Taking the Con out of Econometrics".

Whether a study's results are based on quasi-experimental evidence or on correlations, the degree of causal inference depends on the researcher's knowledge about the state of the world surrounding the event under study.[3] The story or sequence of events that determines the outcome has to be conclusive and at least the most prominent alternative has to be addressed and dismissed. All three papers in this dissertation devote a significant part of their overall length to these issues. In the first paper some limitations of quasi-experiments such as parallel treatment and external validity are encountered. The second paper faces the challenge to address different probable causal sequences in an econometrically demanding environment. Finally, the third paper tests a theoretically very plausible causal mechanism against the backdrop of many intertwined connections. The rest of this section provides a short description of each paper. For the sake of brevity, most references are left out but can be found in the respective chapter.

The first paper, coauthored with Michael Berlemann and Max Steinhardt, can be allotted to the field of empirical microeconomics (section 2). The aim of the paper is to investigate whether individuals react to natural disasters (flooding) by adjusting their saving behavior. The research question is of great economic and social interest, as several macroeconomic studies show that natural disasters, such as flooding, are linked to lower economic growth at the national level. Relying on the tools prominently discussed in Angrist and Pischke (2009, 2010), the paper focuses on the 2002 Elbe flooding in Saxony Germany exploiting the catastrophe as a quasi-experiment. From an econometric point of view the empirical setting provides a textbook-like example to employ a differences-in-differences [DiD] estimator. Identification of affected and non-affected individuals is based on the geo-coordinates of surveyed households and flooded areas. The precision of the geo-referenced data in combination with a detailed panel survey makes the paper one of the first of its kind. Comparing pre- and post-disaster differences in treatment and control group, the paper finds that the flooding had a negative effect on individual saving volumes. However, the paper does not test individual behavior derived from a specific structural model.[4] Therefore, several possible reasons for the negative effect are analyzed and discussed. Ultimately, the most probable reason for the observed behavior are aid payments.

[3] See remarks in the Foreword to this dissertation as well as the discussion in Heckman and Vytlacil (2007).

[4] Heckman and Vytlacil (2007) warn that evidence of natural experiments does not need to be causal. They advise to either provide theory or evidence of causal/plausible mechanisms.

Unusually high amounts of governmental financial aid allotted to affected individuals induced moral hazard as predicted by the Samaritan's Dilemma. The paper makes a case for policy makers to carefully design post disaster aid payments in order to not cause detrimental reductions in individual precautionary efforts.

In the second paper (section 3), which is my single-authored paper, I use aggregate country level data to construct a panel data set. The empirical aim of the study is not to exploit an experimental setting but to use an estimation strategy that uncovers the dynamics of the data.[5] The study can be assigned to the field of empirical macroeconomics and uses aggregate data from various sources. The research question focusses on the link between foreign education and domestic productivity. Specifically, the paper empirically tests whether cross-country skill-set alignment accelerates domestic productivity growth. The research question is of general relevance, as many governments around the world allocate public funds to support student exchange programs. However, to this date very little empirical evidence exists regarding the aggregate effects of foreign education. Given the lack of a natural or quasi-experiment, causal inference in this setting is more difficult. The identification strategy will be based on exploiting temporal as well as cross-country variation of foreign students at U.S. universities. Estimations rely on different (dynamic) panel estimators. From an econometric perspective this requires to account for more characteristics of the data than simple ordinary least squares can handle. Eventually, the paper finds that the more students a country sends to the U.S., the higher subsequent domestic rate of productivity growth will be. Moreover, the results show that this effect is driven by developing countries. The findings suggest that skill transfer through foreign education supports productivity growth in developing countries. Thus, foreign education poses a viable additional strategy for economic development.

Finally, the third paper, coauthored with Grischa Perino, can be attributed to the field of empirical industrial organization. The research design of the paper is based on combining a structural model with variations in environmental policy. The paper tests whether prices for an abatement technology are influenced by the type of environmental regulation of polluting sources. This is an important relationship to be investigated as price discrimination can hamper the diffusion of adequate abatement technology. Whenever environmental regulation targets only a subset of otherwise (almost) identical firms in an industry, the willingness-to-pay [WTP] for abatement technology of the firms covered by the policy can be affected. Vendors of the abatement devices (here flue-gas desulfurization [FGD] devices) can then use the variation in environmental regulation to identify firms with a higher WTP and charge higher prices accordingly. The paper makes use of sulfur-dioxide [SO_2] regulation of coal-fired power plants in the U.S. The data

[5] Actually, it is very difficult to find quasi-experiments that help to uncover causal relationships answering research questions as the one presented here.

come from annual surveys conducted by the U.S. Energy Information Agency. Identification is based on quasi-experimental variation generated by the partial introduction of a SO_2 allowance trading scheme (the Acid Rain Program). The applied estimation strategy is rather simple and statistical (causal) inference relies on the quasi-experimental nature of the data. To strengthen causal inference, behavioral incentives of regulated coal-fired power plants are derived from a structural model. Underlying assumption of the structural model, such as the existence of market power, are tested and discussed. In the end, the paper finds that scrubber prices are significantly higher for plants participating in permit trading. This suggests that vendors of FGD devices apply third-degree price discrimination based on the regulatory instrument. Policy makers should therefore consider that regulatory instruments can have unintended side-effects hampering the diffusion and adoption of abatement technology.

The rest of the dissertation includes the three papers outlined above and a short concluding statement in section 5. The appendix includes all appendices of the included papers (sections A to C) as well as some additional formalities (section D).

2 Do Natural Disasters Affect Individual Saving? Evidence from a Natural Experiment in a Highly Developed Country?[6]

"There is nothing more deceptive than an obvious fact."
(Sherlock Holmes in: The Boscombe Valley Mystery)

2.1 Introduction

Climate change is often seen as one of the most challenging problems of our time. The United Nations Human Development Report 2007/2008 declares: "In the long run climate change is a massive threat to human development [...]" (page v). Against this background it does not come as a surprise that many scientific disciplines are dealing with the causes and consequences of climate change. This also holds true for economics, as climate-related economic research has intensified considerably throughout the last decades. Early research focused on the question of how regulation might contribute to a slowdown of carbon-dioxide emissions. However, more recently the focus has changed towards forecasting the likely economic consequences of climate change and to appropriate adaptation policies.

One consequence of climate change is the increased frequency and/or severity of certain types of natural disasters and extreme weather events (UNISDR - United Nations International Strategy for Disaster Reduction Secretariat, 2009; Thomas, 2014). Against this backdrop, there has been an increasing interest in the question of whether and how natural disasters affect economic growth. Since the first systematic analysis of this question was conducted by Skidmore and Toya (2002), a growing body of empirical literature studying the growth effects of natural disasters has evolved. Most of the existing empirical evidence concerns the short-term effects of natural disasters (see, e.g. Kahn, 2005; Anbarci, Escaleras and Register, 2005; Bluedorn, 2005; Raddatz, 2009; Noy, 2009; Mechler, 2009; Hochrainer, 2009; Loayza et al., 2012; Strobl, 2012). The existing body of research tends to find negative short-term growth effects of natural disasters. These negative short-term effects are more pronounced in less developed than in high income countries. As Noy (2009) argues, this might be due to financial constraints for reconstruction, less developed insurance markets, and limited possibilities to run counter-cyclical

[6]Acknowledgements: This paper has been co-authored with Michael Berlemann and Max Steinhardt. We would like to thank Bernd Fitzenberger, Albrecht Glitz, Alkis Otto, Grischa Perino, Erik Plug, Marcel Thum, and seminar participants of the Workshop"Climate Shocks and Household Behavior" at German Institute of Economic Research (DIW Berlin), the 2014 conference of the Verein fr Socialpolitik in Hamburg, the 2014 Spring Meeting of Young Economists in Vienna, the 3rd workshop on the "Economy of Climate Change" at ifo Dresden, the workshop of the Committee of Environmental and Resource Economics of the Verein fr Socialpolitik, and the Research Seminar of University of Hamburg for useful comments. We also would like to thank Jan Goebel and Christine Kurka (DIW Berlin) for their data support. This work is part of the disasterEcon project, funded by the German Ministry of Education and Research (BMBF) as part of the program "Economics of Climate Change".

fiscal policies. Much less empirical evidence is available on the long-term growth effects of natural disasters. In their cross-sectional study of 89 countries, the pioneering paper by Skidmore and Toya (2002) finds different results for climatic and geologic disasters. Whereas the frequency of climatic natural disasters turns out to have a positive effect on economic growth, geologic disasters tend to have a negative although insignificant impact on economic growth. However, most subsequent studies have found a negative impact of natural disasters on long-run growth (e.g. Noy and Nualsri, 2007; Raddatz, 2009; Felbermayr and Gröschl, 2014; Hsiang and Jina, 2014).
Broadly summarized, one might conclude that natural disasters, at least large ones, tend to affect economic growth negatively, both in the short- and in the long-run, although the strength of the effect depends on country characteristics and the type of disaster. Interestingly enough, the existing empirical literature remains relatively vague with respect to the specific channels through which natural disasters might affect long-run economic growth. Only a few papers have engaged in attempts at uncovering these channels.

In this paper we aim to shed additional light on one specific channel through which natural disasters might affect economic growth: saving behavior. The savings rate is well-known to be a decisive factor in determining per-capita income in macroeconomic models of economic growth in closed economies. In open economies, the role of domestic saving for economic growth is less clear, as domestic investments can also be financed by foreign savings. However, there are reasonable theoretical arguments for why domestic saving is also crucial in open economies. Dooley, Folkerts Landau and Garber (2004) argue that poor and instable countries in particular might transfer domestic savings to countries of possible investors, thereby making expropriations of foreign investors capital less likely. Thus, the transfer of domestic savings takes on the role of collateral, which encourages foreign investments and contributes to better economic development. In a similar vein, however, based on a well-defined theoretical model, Aghion et al. (2016) argue that domestic savings play an important role in relatively poor countries that employ production technologies far away from the technological frontier. In these countries, catching up to developed countries requires a joint venture between a foreign investor who is familiar with the frontier technology and a domestic entrepreneur who is familiar with the local conditions. In this scenario, domestic savings are necessary to mitigate the agency problem which would otherwise deter the foreign investor from joining this project. The empirical evidence that Aghion et al. (2016) present supports the relevance of this line of argument. Moreover, studies on the determinants of economic growth suggest that domestic savings have a positive impact on economic growth (e.g. Barro, 1991; Mankiw, Romer and Weil, 1992; Islam, 1995).

In summary, we might conclude that whenever natural disasters have a permanent influence on domestic savings behavior, medium- or even long-term economic growth will also be affected. However, the effect of natural disasters on domestic savings is ex-ante ambiguous. In principle natural disasters can affect individual saving behavior in different ways and through different channels.[7]

Saving is typically seen as a means of consumption smoothing. Naturally, the amount of saving will increase with a corresponding increase in life expectation. Whenever natural disasters make individuals believe that life expectations decrease, this might increase consumption and depress saving. However, the theory of precautionary saving argues that saving does not only serve to spread income over the life cycle, but might also serve as insurance against uncertain events (Lusardi, 1998). In this context, Roson, Calzadilla and Pauli (2006) argue that individuals might react to natural disasters by increasing their savings. Based on a theoretical model of constant absolute risk aversion, Mani, Keen and Freeman (2003) show that the optimal amount of precautionary saving depends positively on expected loss, and thus on both the disaster probability and disaster loss. Natural disasters might increase expected losses and thus increase precautionary saving. This effect should be especially pronounced for more risk-averse individuals (Fuchs-Schundeln and Schundeln, 2005). However, it is also possible that precautionary saving is reduced as a consequence of natural disasters. Often individuals who have suffered from catastrophic losses are supported or even fully compensated by state institutions, private donations, or international aid. All these forms of support decrease the incentives for accumulating ones own precautionary savings. Finally, individuals might be forced to dis-save for a certain period of time in response to natural disasters, due to increases in expenditures (e.g., for repairs or replacements) or negative income shocks.

In order to further investigate the effects of natural disasters on saving behavior, we study whether the occurrence of a large natural disaster (i.e., the flood of August 2002 in central Europe) affects subsequent individual saving behavior in the flooded region. We base our study on micro-level data from the German Socio-Economic-Panel (SOEP), focusing on those panel members who lived in Saxony – the German state that was the most affected by the flood catastrophe.[8] Using geo-referenced maps of the flood, we identify two groups of individuals. The first group of individuals lived in areas of Saxony that were unaffected by the flood; they serve as our control group. The second group lived inside the flooded areas and make up our treatment group. Subsequently, we apply a differences-in-differences approach to analyze the impact of the 2002 flood on individual saving behavior.[9] We find that the flood caused a significant reduc-

[7]We summarize the related literature in section 2.2.

[8]The data used comes from the Socio-Economic Panel (SOEP), data for years 1984-2012, version 29, SOEP, 2013, doi:10.5684/soep.v29.

[9]Bechtel and Hainmueller (2011) use the August 2002 flood to analyze the impact of national aid flows on voter gratitude. Using data on electoral districts, they apply a differences-in-differences analysis to estimate the effect of disaster aid on national election outcomes. In contrast to our paper, they did not use geo-referenced information on floods, but aggregated data on flooding at the level of electoral districts. Regarding flood assistance, they

tion in individual savings. We also show that this finding cannot be explained by income effects alone, and discuss the potential driving forces behind our results.
The remainder of the paper is organized as follows. In the next section, we briefly summarize the related empirical literature. Section 2.3 gives a brief overview on the August 2002 flood catastrophe in central Europe, with a special emphasis on Saxony, introduces the data set, and explains our estimation strategy. In section 2.4, we study the effect of the flood catastrophe on individual saving volume, distinguishing between the extensive and intensive margin of the saving decision. Section 2.5 examines the potential income effects and analyzes the individual savings rates while section 2.6 delivers additional robustness checks. Finally, in section 2.7 we summarize our main results and offer concluding remarks.

2.2 Related Literature

In the standard neoclassical growth model, a natural disaster destroying parts of an economy's capital stock has a negative short-term impact on per-capita Gross Domestic Product (GDP). Thus, in the very short-term perspective, natural disasters should negatively affect the growth rate. As the economy returns to its long-term steady state, the intermediate growth rate must exceed the long-term trend. In the long-run, the growth rate should remain unaffected by the disaster as the economy has returned to its steady state.

Natural disasters might have an influence on long-run economic growth whenever one of the key variables (i.e., those assumed to be exogenous in the standard neoclassical growth model) changes as a result of a natural disaster. The most important factors determining steady state per-capita GDP are the savings rate (i.e., investments), population growth, human capital accumulation, and the rate of technical progress. However, to date the empirical literature has rarely studied whether and how the occurrence of natural disasters influences these factors. To the best of our knowledge, the only study that is explicitly concerned with the effects of natural disasters on these growth factors is the early study by Skidmore and Toya (2002). The authors detected no significant effect of disaster risk (i.e., measured by the average rate of disasters which occurred throughout the sample period of 1960-1990) on the growth of physical capital (and thus saving). The effect of natural disasters on human capital growth and total factor productivity turns out to depend on the type of natural disaster. While the effect of geologic disasters is negative and insignificant, the effect of climatic disasters on both human capital accumulation and total factor productivity is positive and significant. The authors therefore conclude that climatic disasters have a positive impact on long-run economic growth, as climatic disasters provide the opportunity to update capital stock and adapt new technologies. However, subsequent literature has found little support for this hypothesis that climatic natural disasters have a positive long-

assume that every district that was affected by the flood received disaster aid. They conclude that flood aid had a positive impact on the voter share of the incumbent party in the preceding election.

term growth effect.

Based on a life cycle expected utility model, Skidmore (2001) shows that saving should generally increase as a result of rising expected future losses from natural disasters. While this result does not hold when perfect insurance is available, Skidmore (2001) argues that even in highly developed countries, disaster insurance is often unavailable due to the combination of the low likelihood of disaster occurrence and the enormous damages to be covered in the case of disaster events. Based on a very small data set consisting of 15 highly developed countries, Skidmore (2001) found that the more a country is prone to natural disasters, the higher the aggregate saving rate.

In order to investigate how natural disasters might influence long-run growth, it is useful to study behavioral responses to disasters at the microeconomic level. Although the existing empirical evidence is relatively scarce, recently a number of papers have studied this issue, although rarely in a growth context. Sawada and Shimizutani (2008) find that post-disaster consumption behavior patterns after the Kobe earthquake depend strongly on individual borrowing constraints. In an attempt to quantify the welfare costs of floods in European societies, Luechinger and Raschky (2009) find that individual happiness is negatively affected by flood disasters. Berlemann (2016) finds that global hurricanes only depress happiness in the short-run. However, hurricane risk turns out to have a strong negative impact on life satisfaction. Page, Savage and Torgler (2014) find that the Brisbane flood of 2011 had a significant effect on individual risk seeking behavior whenever an individual suffered a substantial loss in wealth. Finally, Cameron and Shah (2015) conducted several field experiments with individuals affected by disasters in rural Indonesia and find them to be more risk adverse. Taken together, this evidence suggests that natural disasters can induce behavioral responses.

2.3 Background, Data and Methodology

In our analysis of behavioral responses to natural disasters, we study whether and how individuals adjusted their saving behavior in response to a severe flood that occurred in central Europe in summer 2002. Before we turn to the empirical analysis, we first summarize the main facts about the flood catastrophe. We then turn to the description of the dataset and explain the basic strategy used to identify individuals who were affected by the flood. The section ends with an introduction of the employed empirical methodology.

2.3.1 The August 2002 Flood in Saxony

In July and the beginning of August 2002, central Europe experienced multiple waves of heavy rainfall and thunderstorms. Several watercourses exhibited increased gauge stages and the soil was saturated with water in many parts of Saxony, Bavaria, the Czech Republic, and Austria (Loepmeier, 2003). The first floods in these areas occurred between August 7 and 11, as water houses were only able to drain off above ground (GWS, 2007). In the early hours of August 12, the storm-front Ilse crossed the Czech Republic and moved towards Saxony. The overall meteorological situation in Europe during that time and the orographic conditions in Saxony caused extreme rain as the storm-front completely unloaded its waters above eastern Germany. In the Ore Mountains, which are close to the Czech boarder, official measures reported 312 liters of water per square meter within 24 hours (Rudolf and Rapp, 2003). This all-time German record exceeded historical precipitation levels by a factor of four. In other central European regions, Ilse dropped between 80 and 167 liters of water per square meter in a 24 hour period. In many affected regions, the water masses caused massive direct damage.

In Saxony, the water masses caused destruction through various channels. First, small watercourses in the Ore Mountains flooded and caused destruction on their way down to the Elbe River. Much of the reported damage was caused by these tributaries that are normally rather small. Second, many of the water reservoirs located in the Ore Mountains already exhibited increased gauge stages. Traditionally the reservoirs had two functions, drinking water storage and flood prevention for the Elbe valley. In late July, many of the reservoirs had gauge stages close to maximum in order to provide ample fresh and drinking water for the summer season. When the somewhat unexpected heavy rain period started, emergency drainages became necessary in various reservoirs to prevent bursting dams. As a consequence of one of these emergency drainages, the Weieritz stream, which is normally a small watercourse in the Ore Mountains, became a torrential river within a matter of minutes and caused massive destruction in several villages including the medium-sized city of Freital and Saxony's capital Dresden, where the water flooded the main station and substantial parts of the city center. Third, the specific orographic constellation of the region from Prague to Dresden makes the Elbe River the only significant drainage for increased water houses. Thus, the heavy rainfall at the beginning of August steadily increased the gauge level of the Elbe River. Finally, several flood waves from the Czech Republic made their way down the Elbe River and reached eastern Germany after this heavy rain period of August 11 and 12. The already high gauge stages of the Elbe thus increased even further, thereby causing severe damage to many settlement areas close to the Elbe River.

Even though the Elbe River is by far the largest watercourse that was affected by the heavy rain, the flood catastrophe was not restricted to its surrounding areas. As mentioned earlier, other areas such as those close to the Mulde River were also affected, and severe damage was often caused by small tributaries such as the Weieritz. Consequently, the flood affected many distinct parts of Saxony. For Germany as a whole, the Center for Research on the Epidemiology of Disasters (CRED) reports that 330,108 people were affected and the damage totaled $11.6 billion. In these two dimensions, the 2002 flood is Germanys most severe natural disaster recorded in the CRED database.

2.3.2 Household Data

For our empirical study, we used the German Socio-Economic Panel (SOEP), a panel data set on German households.[10] The SOEP is a representative annual panel survey which started in 1984 in West Germany, and has included the areas that formerly comprised East Germany since German Reunification in 1990. The survey contains roughly 150 questions that allow researchers to extract information on the socio-economic infrastructure of the included households. Among other variables, the survey includes data such as individual wealth, income, employment, and health status. All household members above the age of 17 are personally interviewed. In addition to a personal interview, the head of the household answers an additional household questionnaire.[11]

All household variables that are used in the empirical estimations are taken from the SOEP. To identify those households that lived in flooded areas, we make use of anonymized regional information on the residences of SOEP respondents.[12] This data is considered as highly sensitive and is subject to particular data protection regulations. We refrain from describing all variables here, but a complete description of the employed variables can be found in table A1. Instead, we focus on describing those variables that serve as dependent variables in our empirical analyses. Our analysis of individual saving behavior is based on the answers to the question:

> "*Do you usually have an amount of money left over at the end of the month that you can save for larger purchases, emergency expenses or to acquire wealth? If yes, how much?*"

The head of the household answers this question by reporting aggregate monthly household savings. In the subsequent empirical analysis we study individual saving behavior.[13] Thus, whenever households consist of more than one person we have to make an appropriate assumption how aggregate monthly saving can be attributed to individual household members. It seems to be reasonable to assume that individual saving is proportional to individual income, which is measured as the sum of revenues from all recorded sources, including wages, social benefits, rents and any other source of income received regularly. We then attribute the household saving volume to the household members based on their individual share in total household income.[14] As we are interested in real rather than in nominal savings, we deflate savings by the German

[10] The SOEP data can be obtained from the SOEP Research Center located at the German Institute for Economic Research (DIW) in Berlin.

[11] For a more detailed description of the SOEP survey, see Wagner, Frick and Schupp (2007).

[12] Section 2.3.3 contains a detailed description of the regional data used to identify flooded regions and how it is matched to SOEP households.

[13] The results of our empirical analysis remain qualitatively unchanged when conducting the analysis on the household level as we will show in the robustness section 2.6.

[14] All subsequently shown empirical results remain qualitatively unchanged when attributing the same share of savings to each household member as a more conservative variant of the applied procedure.

consumer price index and code the result as the variable S.[15]

For the analysis of saving behavior at the extensive margin, we additionally construct the dummy variable SE, which takes the value of one whenever a respondent saves and zero otherwise:

$$S_E = \begin{cases} 1 \mid S > €0 \\ 0 \mid S = €0 \end{cases}$$

Finally, in order to study the saving decision at the intensive margin, we construct the variable S_I. This variable is only defined for individuals in households which declare they save a positive amount of money, i.e.:

$$S_I = S \mid S > €0$$

2.3.3 Definition of Treatment and Control Group using Flood Data

In 2002, the SOEP contained 23,892 people living in 12,605 households. Of these, 1,678 people (or 860 households) lived in Saxony. As Saxony was the German state most heavily affected by the August 2002 flood, we concentrate our analysis on SOEP members living in Saxony when the August 2002 flood occurred. Since the financial freedom of children and adolescents is rather limited, we exclude all respondents younger than 18 from our analysis. In our empirical analysis, we are interested in comparing savings behavior before and after the flood occurred. We therefore exclude all respondents who were interviewed in 2002 (i.e., after the flood occurred in early August).[16] Hence, the remaining 1285 respondents interviewed in 2002 compose our pre-disaster observations.

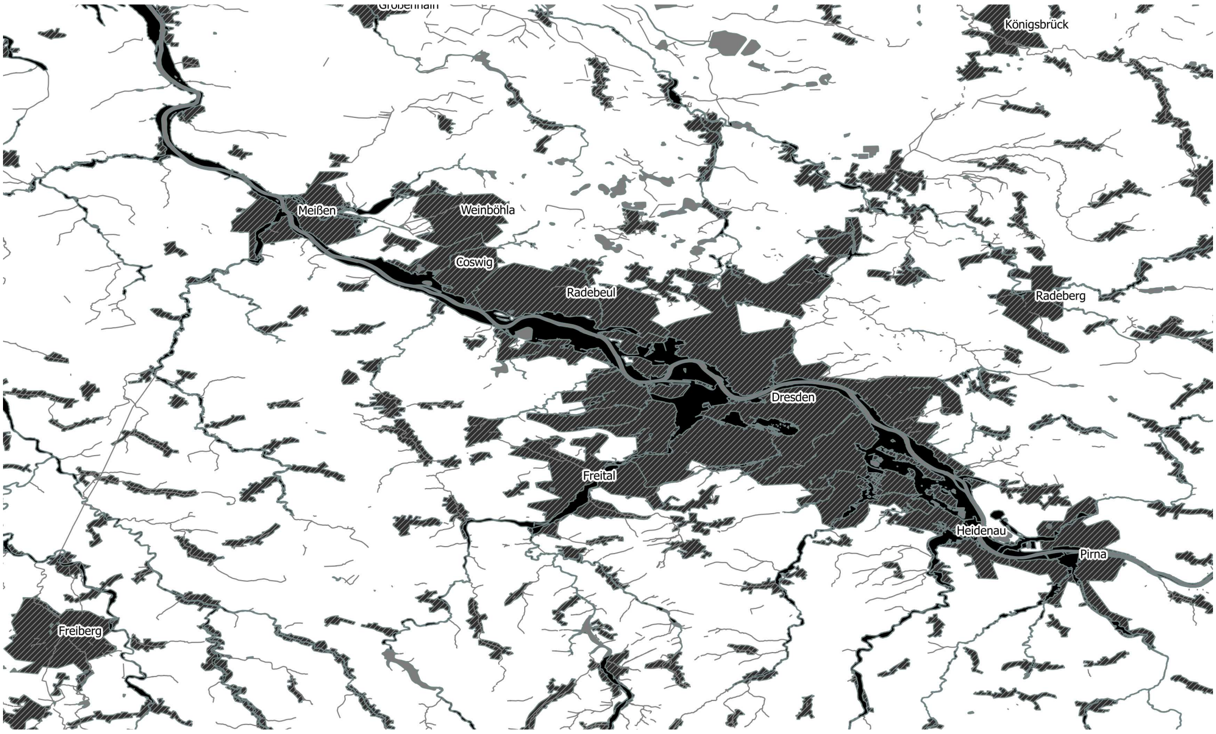
A crucial issue in our empirical analysis is the identification of those SOEP respondents who lived in the flooded area and were therefore strongly affected by the 2002 flood. Not surprisingly, the SOEP dataset does not contain a variable or question that pertains to this issue. However, by applying a three-step procedure which we will describe in detail, we are nevertheless able to identify Saxon SOEP respondents who lived in the area flooded in August 2002. In the first step, we collect detailed geographic data on the flood impact in Saxony. For this purpose, we employ a combination of two flood maps. The first map was constructed by the Saxon State Office for Environment, Agriculture and Geology on the basis of aerial photography and hydraulic computations. The map was refined and updated various times; we use the version dated November 2007. As this first map excludes the city of Dresden (i.e., one of the most heavily affected regions in Saxony), we combined this map with a flood map provided by

[15] We make use of the consumer price index (code 61111-0001) published by the German Statistical Office. Savings are expressed in values as of year 2000.

[16] As the SOEP questionnaire is primarily carried out in the first half of the year, very few observations were excluded for this reason.

the City of Dresden's Department for Environmental Protection. After merging the two maps, we attained one coherent flood map covering the whole state of Saxony.[17] The combined map contains about 220 watercourses. Some 2,800 kilometers, or 11.2 percent of Saxonys watercourses, were affected.[18] The total flooded area amounts to about 40,000 hectares.[19] Roughly 20 percent of this area is classified as settlement area or infrastructure. Figure 1 shows a graphical representation of the employed combined flood map. Dark areas were flooded throughout August 2002 while areas marked in light gray indicate watercourses. The shaded areas depict settlements.

Figure 1: Map of flooded areas in Saxony throughout the August 2002 flood



Source: Constructed from data from the Federal Office of Cartography and Geodesy, the Department for Environmental Protection of the City of Dresden, and the Saxon State Office for Environment, Agriculture and Geology. Notes: The flood map shows areas which were flooded throughout the August 2002 flood (marked in black). Shaded areas (dark gray) indicate settlements while regular watercourses are a solid light gray.

In the second step, we localize the SOEP households within Saxony. Although the standard SOEP dataset provides only information on the state level in order to protect respondents' privacy, more detailed information is available at the SOEP Research Data Center in Berlin.[20] The available geographical units comprise inter alia, official municipality keys, postal codes, and Microm neighborhood data. All of these geographical identifiers have been available since 2000, at the latest. As the Microm neighborhood data contains the most detailed location infor-

[17] The creation of the flood map is based on maps with a scale of 1:10,000 (in cm) and the official topographic map TK 10. For validity checks, more highly scaled maps (e.g., 1:5,000) were used in densely populated areas.

[18] In total there are about 25,000 km of watercourses in Saxony.

[19] About 2.2% of the total surface area of Saxony.

[20] Geo-coordinates of included households can only be obtained at the research center. Less detailed geo-referenced data can be obtained and used outside the research center.

mation, we make use of this location identifier in our analysis. The Microm identifier localizes households by the geo-coordinates of their living places.[21]

In the third and final step, we match the geo-coordinates of Saxon SOEP households with the combined flood map.[22] Doing so allows us to identify which adult Saxon SOEP respondents lived inside the flooded area, and which respondents lived outside of it, when the August 2002 flood occurred.

As our treatment group, we define those respondents who lived inside the flooded area when the flood occurred in 2002. As our control group, we make use of those respondents identified as living outside the flooded areas.[23] In order to ensure that the control group contains exclusively unaffected SOEP respondents, we include only those individuals living at least 500 meters away from flooded areas when the flood event occurred.[24] While in the growth context it is interesting to study the long-run flood impact, our time perspective is somewhat limited as parts of Saxony experienced another, yet less severe flood, in spring 2006. As this would ultimately threaten our identification strategy, we restrict our post-disaster analysis to the years between 2003 and 2005. However, this perspective nevertheless goes well beyond the short-term growth effect of natural disasters.[25] Finally, we drop those respondents who have been interviewed in 2002 before the flood, but have left the treatment or control region before the flood occurred. This leaves us with a treatment group of 50 persons, and a much larger control group of 1225 persons. Table 1 shows the summary statistics for all variables in the pre-disaster year (2002), conditional on being a member of the treatment or control group.

2.3.4 Estimation Strategy

The aim of our empirical analysis is to study whether and how individuals, who lived in the flooded area when the catastrophic flood occurred, adjusted their subsequent saving behavior. In order to study the causal effect of the flood on saving behavior, we apply a differences-in-differences (DD) approach as described below. We hereby follow the basic framework outlined by Angrist and Pischke (2009).

In our setting, we have two regions $(1,2)$; region 2 was hit by the flood in August 2002. Moreover, we have two periods (before August 2002, after August 2002) for which we can observe

[21] For additional information on geographically referenced data and the SOEP, see Hintze and Lakes (2009).

[22] The flood map and the geo-coordinates of Saxon SOEP respondents were matched using the open source software Quantum GIS Dufour.

[23] In our study, being unaffected implies that these individuals should not have suffered directly from the flood catastrophe and therefore did not receive any financial disaster aid by private insurance companies or the state.

[24] Respondents living outside the flooded areas but closer than 500 meters to such an area were excluded from the analysis.

[25] Many empirical studies on the growth effects of natural disasters solely focus on the growth effects in the subsequent year.

Table 1: Summary statistics for treatment and control group (Person-Level)

Year 2002	**Treatment Group**					**Control Group**				
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**N**	**Mean**	**Std. Dev.**	**Min**	**Max**	**N**
Saving (S)	118.97	249.13	0	1450.90	50	130.95	239.35	0	3088.835	1180
Saves (S_E)$^\diamond$	0.66	0.48	0	1	50	0.66	0.47	0	1	1180
Saving (S_I)	180.26	289.21	11.73	1450.90	33	198.62	271.06	2.11	3088.84	778
Saving Rate (SR)	0.07	0.11	0	0.56	50	0.08	0.11	0	0.95	1178
Controls:										
Sex$^\diamond$ (1=male)	0.47	0.50	0	1	50	0.49	0.50	0	1	1225
Age	44.26	16.07	19	77	50	47.45	16.66	18	91	1225
Homeowner$^\diamond$	0.59	0.50	0	1	50	0.43	0.50	0	1	1225
Primary Educ.$^\diamond$	0.14	0.35	0	1	49	0.11	0.32	0	1	1190
Secondary Educ.$^\diamond$	0.59	0.50	0	1	49	0.57	0.50	0	1	1190
Tertiary Educ$^\diamond$	0.27	0.45	0	1	49	0.32	0.47	0	1	1190
Employed$^\diamond$	0.66	0.48	0	1	50	0.57	0.50	0	1	1225
Unemployed$^\diamond$	0.10	0.30	0	1	50	0.09	0.28	0	1	1225
Non-working$^\diamond$	0.24	0.43	0	1	50	0.34	0.48	0	1	1225
Single$^\diamond$	0.26	0.44	0	1	50	0.24	0.43	0	1	1225
Married$^\diamond$	0.65	0.48	0	1	50	0.62	0.49	0	1	1225
Other$^\diamond$	0.10	0.30	0	1	50	0.14	0.35	0	1	1225
No child$^\diamond$	0.57	0.50	0	1	50	0.68	0.47	0	1	1225
1 child$^\diamond$	0.28	0.45	0	1	50	0.20	0.40	0	1	1225
2 children$^\diamond$	0.04	0.20	0	1	50	0.09	0.29	0	1	1225
3+ children$^\diamond$	0.04	0.20	0	1	50	0.09	0.29	0	1	1225
Rural Area$^\diamond$	0	0	0	0	50	0.21	0.41	0	1	1225

Notes: The table presents summary statistics by group category. $^\diamond$ indicates that variable is binary, with Yes=1. Due to missing answers observations can differ across variables. The statistics are based on SOEP answers in 2002 before the flood occurred in August. A detailed description of listed variables can be found in table AI in the appendix.

individual saving behavior.[26] Given this situation, we have two potential outcomes. S_{1irt} is the saving of individual i in region r $(1,2)$ at time t (before August 2002, post 2002) if a flood happened, and S_{0irt} is the saving of individual i in region r at time t if no flood happened. However, in reality, we only observe one or the other event. For example, we can see S_{1irt} in region 2 in 2003 but we cannot observe the counterfactual S_{0irt} in region 2 in 2003, since region 2 was affected by the flood in August 2002. The DD setup is based on an additive structure for potential outcomes in the no-treatment scenario:

$$E[S_{0irt}|r,t] = \gamma + \lambda_t.$$

We therefore assume that saving without a flood is determined by the sum of a time-invariant regional fixed effect (γ_r) and a time effect (λ_t) that is common across regions. Let D_{rt} be a dummy variable for flooded regions and periods. Assuming that $E[S_{1irt} - S_{0irt}|r,t]$ is the constant , the observed saving S_{irt} can be written as:

$$S_{irt} = \gamma_r + \lambda_t + \delta D_{rt} + \varepsilon_{irt}, \tag{1}$$

[26] De facto we have more than one post-treatment period. For simplicity, we explain the DD approach with two periods only.

where $E(\varepsilon_{irt}|r,t) = 0$. The expected differences for the two regions are thus:

$$E[S_{irt}|r=1, t=post\,02] \quad - \quad E[S_{irt}|r=1, t=prior\,Aug\,02] = \lambda_{post\,02} - \lambda_{prior\,Aug\,02}$$

and

$$E[S_{irt}|r=2, t=post\,02] \quad - \quad E[S_{irt}|r=2, t=prior\,Aug\,02] = \lambda_{post\,02} - \lambda_{prior\,Aug\,02} + \delta.$$

The difference between the two expected differences, the difference-in-differences estimator, is thus δ. One way to estimate equations like (1) with additional individual level covariates, $\mathbf{x'}_{rit}$, is:

$$S_{rit} = \alpha + \gamma treat_r + \lambda year_t + \delta(treat_r \times year_t) + \beta \mathbf{x'}_{rit} + \varepsilon_{rit}, \tag{2}$$

where *year* is a dummy variable that switched to 1 in the years after the flood event happened. The dummy *treat* takes the value 1 for region 2 (where the flood occurred in August 2002) and 0 otherwise. When studying saving behavior, we start out with an analysis of the overall saving volume S. As our saving measure cannot be negative, we use the tobit approach in the first step of our analysis. We then turn to separate analyses of the two dimensions of the savings decision: the decision to save at the extensive and intensive margin. As the decision to save or not to save is a binary one, we employ probit regressions for the analysis of the extensive margin of the saving decision (S_E). The decision to save at the intensive margin (S_I) is analyzed based on a linear model using standard OLS techniques.

We conduct all our empirical analyses with our sample of SOEP respondents that were attributed to either the treatment or the control group. As we analyze the effect of the flood on saving behavior in three post-disaster years, we report three different estimates (2002/03, 2002/04, and 2002/05). In order to study the stability of the derived results and to further investigate potential factors driving our results, we conduct a number of additional estimations in sections 2.5 and 2.6. As it is easier to understand these estimates after learning about the main estimation results in section 2.4, we explain those approaches in later sections.

2.4 Empirical Analysis of Individual Saving Behavior

Our empirical analysis of the flood's impact on saving behavior covers three dimensions: the effect on overall saving S, the effect on the extensive margin S_E, and the effect of the intensive margin of the saving decision S_I. Thus, we estimate the differences-in-differences regression outlined in equation (2) using three different dependent variables. As outlined earlier, we make use of different estimation techniques to adequately take the different characteristics of the referring dependent variables into account. In table 2 we report the estimation results.[27] To

[27]The number of observations across specifications varies slightly due to missing values of explanatory variables and the choice of the dependent variable.

ease interpretation, we only report the results for the dummy variables for year and treatment, as well as the interaction between these two dummy variables which captures the treatment effect.[28]

The upper part of table 2 reports the results of tobit models where we estimate the flood's impact on the latent variable S.[29] We report the effect on the uncensored latent variable. Thus, the estimated coefficients can be interpreted as the predicted change in predicted saving levels. The estimates suggest that the flood depressed saving in all three years succeeding the disaster. However, in 2003 (i.e., the first year after the flooding), the effect is not significant. In 2004 and 2005, the effect becomes significant and also increases in magnitude. We also report the conditional marginal effect of the flooding on factual saving, which turns out to EUR -59 in 2004 (-25 percent) and EUR -69 in 2005 (-21 percent).This suggests that the flood had a very strong and lasting effect on individual saving behavior. The center part of table 2 shows the results of the probit models that analyze the decision to save at the extensive margin. Experiencing the flood had a negative impact on the decision to save in all post-disaster years analyzed. As for the overall saving decision, the effect of the flood is significant in the years 2004 and 2005. In order to deliver a meaningful interpretation of the estimated coefficients, we compute marginal effects for an individual with median characteristics (i.e., the year and the interaction term were set to 0).[30] For 2004, we find the median individual, impacted by the flood, had 30.5 percentage points lower probability of saving any money, as compared to 2002. Even in 2005 (i.e., three years after the disaster), flood-affected individuals are 23.9 percentage points less likely to save. These findings are in line with the results from our tobit model and suggest a rather strong behavioral reaction to the flood. Finally, the estimates of the linear model, as reported in the bottom part of table 2, show the floods impact at the intensive margin of the saving decision. Note that for the analysis of S_I, only those respondents that save a positive amount before the flood and in the respective post-flood year are included in the analyses. We find no systematic effect on saving, here; however, the number of observations is also very small in this specification because several respondents reduced their savings to zero in 2004 and 2005, as our estimates at the extensive margin have shown.

To sum up, the results of our estimates show that the flood significantly reduced the savings of affected individuals. While we detected no response to the flood at the intensive margin of the saving decision, the effect is significant at the extensive margin of the saving decision two and three years after the disaster.

[28]Full estimation results are provided in the appendix, tables A.2 to A.4.

[29]Our dependent variable is the logarithm of total household saving S. In cases of households with zero saving, S was manually set to one and hence the logarithm of S was set to zero.

[30]Results are also similar when a linear probability model is used. Estimation results are available from the authors on request.

Table 2: Estimation Results Individual Saving Behavior

Model	**Variable**	**2002/03**	**2002/04**	**2000/05**
Tobit	Dept. Var.: S	(I)	(II)	(III)
	Year	-8.108	-11.637	0.404
		(0.433)	(0.288)	(0.979)
	Treat	-23.774	-16.321	-7.607
		(0.658)	(0.759)	(0.904)
	Year × Treat	-71.623	-175.168**	-178.232***
		(0.232)	(0.015)	(0.005)
	$ME[E(S\|S>0)]$	-26.917	-59.011**	-69.631***
		(0.253)	(0.021)	(0.006)
	Change (in percent)[1]	-10.58	-25.41	-21.40
	Log pseudolikelihood	-10.599.234	-9.888.180	-9.818.446
	Observations	2188	2068	1974
	Left censored Obs.	764	722	683
Probit	Dept. Var.: S_E			
	Year	-0.068	-0.072	-0.034
		(0.224)	(0.210)	(0.604)
	Treated	-0.012	0.018	0.032
		(0.965)	(0.950)	0.913)
	Year × Treated	-0.163	-0.789**	-0.628**
		(0.593)	(0.026)	(0.021)
	ME	-0.058	-0.305**	-0.239**
		(0.581)	(0.018)	(0.016)
	Change (in ppt.)[2]	-5.845	-30.535	-23.91
	Log pseudolikelihood	-1.285.720	-1.218.389	-1.155.581
	Observations	2188	2068	1974
OLS	Dept. Var.: S_I			
	Year	-13.543	5.776	4.700
		(0.494)	(0.788)	(0.877)
	Treated	-110.168	-75.464	-95.487
		(0.174)	(0.523)	(0.368)
	Year × Treated	-108.24	-133.760	-87.587
		(0.133)	(0.333)	(0.300)
	Adjusted R^2	0.122	0.128	0.123
	Observations	1,276	1,196	1,096

Notes: ME stands for marginal effect. ME are computed for a person with characteristics according to the median of all treated individuals in the respective regression sample. Observations can vary between regressions as regression samples are only balanced for the specific year-to-year combination analyzed. All regressions include the control variables depicted in table 1. The interaction terms capture the change in the differences on the outcome between before and after the flood (i.e. the effect of the flood).[1] Refers to the percentage change in predicted S between the two groups using characteristics of an average treated person in 2002. [2] Refers to the change in the likelihood to save any amount of money due to the flood. The variable S is censored at 0. P-values are reported in parenthesis and standard errors are clustered on the household level. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

2.5 Why Do Affected Individuals Reduce their Savings?

In the previous section, we presented empirical results indicating that the flood in Saxony of August 2002 had an economically important and statistically significant negative effect on saving behavior on those individuals who decided to stay in the disaster-affected area. As stated in the introduction, there are several potential explanations for this finding. In this section, we try to identify which of these explanations is the most likely driving force behind our results.

One possible explanation could be that flood-affected individuals updated their life expectancy after the disaster and consequently adjusted their time preferences (cf. Callen, 2015; Cassar, Healy and von Kessler, 2017). The observed reduction in savings would imply that affected individuals expect to die earlier. Such a shift in time preferences in the context of the Elbe flooding is unlikely. Germany is a highly developed country with relatively high protective measures (e.g., strict building codes) that should prevent high death tolls in the event of a natural disaster. Indeed, while more than 330.000 Germans were affected by the 2002 Elbe flood, the death toll was considerably small and amounted to 27 people (CRED/EM-DAT).[31] We therefore consider it very unlikely that an adjustment of life expectation is the driving force behind the decision to decrease saving. A second explanation might be that decreased saving is the consequence of increased expenditures. Individuals severely affected by the flood might require all of their available income to cope with the consequences of the disaster. However, one might have serious doubts about this dis-saving explanation. Significant financial aid flows were allotted to affected individuals in the aftermath of the flood event. Affected households received governmental help, payments from charity organizations, or insurance payments shortly after the flood. In addition, aid in kind and neighborhood support were substantial. For Germany as a whole, more financial aid was available than was needed to deal with the estimated damages of the flood (Mechler and Weichselgartner, 2003). While governmental programs could rely on a national fund of about EUR 7.1 billion, insurance payments and charity payouts for households in Saxony alone amounted to EUR 240 and 362 million, respectively. Governmental aid programs can be divided into two programs: emergency relief and reconstruction relief. As the name emergency relief implies, most of these payments were quickly allotted. In Saxony, nearly all requested emergency relief funds were paid out to affected households by the end of January 2003. Reconstruction relief, aimed at the long term support of affected homeowners, was paid out over the whole period of the reconstruction process. Reconstruction expenses of up to 80% were compensated by the program. By mid-2003, nearly all approved disaster relief was paid out (Striefler, 2003*b*,*a*). In light of these facts, it is somewhat doubtful that post-disaster expenses forced the referring individuals to dis-save. Moreover, the time-pattern we uncover in our estimation results does not support a dis-saving argument. A large share of disaster-related expenses likely occurred soon after the disaster. If in fact disaster-related expenses would have

[31] Access to the database is free but requires registration. Please visit: http://www.emdat.be/database.

enforced dis-saving, we should observe the savings effect to occur quickly after the flood event. However, none of our savings measures decreased significantly before 2004. A third possible explanation is that the flood could have induced the observed reduction in savings through its impact on the local labor market by reducing individual income (Vigdor, 2007; Groen and Polivka, 2010; Deryugina, Kawano and Levitt, 2017). A reduction in individual income would lead to less disposable income and thus fewer savings. However, the flood's short-term impact on the economy in Saxony was rather moderate (Hoffmann, Matticzk and Speich, 2004; Mueller and Thieken, 2005; Berlemann and Vogt, 2008). Moreover, the duration of the estimated effect on individual saving behavior makes it unlikely that the reduction is caused through reduced income alone. Household income might have temporarily declined but this effect would not have persisted over a period of three years. For instance, Mueller and Thieken (2005) report that businesses interrupted production for two to four days after the flood.

In order to formally check whether our results are primarily driven by income effects, we compute individual saving-rates, SR,[32] and run a number of additional regressions. Again we estimate regressions that follow equation (2). However, we now use the saving-rate as our dependent variable. The results of the tobit regressions (table 3) are consistent with the results reported in table 2. As before, we report the conditional marginal effects on the factual saving-rate. While we do not find any significant effect in the first year after the disaster, the flood induced a significant drop in the saving rate in 2004 and 2005. Again, we calculate the percentage change in the saving rate in order to quantify the magnitude. In 2004 and 2005, the flood exerted a reduction in the factual saving-rate of slightly more than 3 percentage points. Given these results and our estimates from the tobit model on total saving, we conclude that the reduction in savings were not primarily driven by a decline in income. Finally, the observed saving pattern could stem from a change in precautionary savings. As outlined earlier, precautionary savings should be reduced whenever the perceived probability of a disaster event and/or a disaster loss decreases. It seems counterintuitive that the occurrence of a disaster should decrease the perceived probability of disasters occurring in the future, as one might expect the opposite to happen instead (see, e.g. Eckel, El-Gamal and Wilson, 2009; Cameron and Shah, 2015). However, the reduction of precautionary savings might stem from decreased perceived disaster loss. At first glance, again there is little reason to believe that perceived disaster loss decreases as a consequence of the occurrence of a disaster. However, it is well possible that perceived loss is decreased by unexpected financial compensation in the aftermath of a disaster. Given that precautionary savings are intended as insurance against unexpected expenditures (Lusardi, 1998), receipt of disaster relief can induce disaster-affected individuals to reduce their savings.[33] In particular, unprecedentedly high compensation rates might induce a reduc-

[32]The individual savings rate is calculated by dividing individual saving S by individual income.

[33]In a similar vein, Raschky and Weck-Hannemann (2007) argue that individuals anticipate governmental and private aid in the case of natural disasters and therefore often refrain from purchasing private disaster insurance. Antwi-Boasiako (2014) provides a more detailed discussion.

Table 3: Estimation Results Individual Saving Rate

Model	**Variable**	**2002/03**	**2002/04**	**2000/05**
Tobit	Dept. Var.: *SR*	(I)	(II)	(III)
	Year	-0.006	-0.006	-0.006
		(0.319)	(0.272)	(0.354)
	Treated	-0.015	-0.012	-0.007
		(0.583)	(0.656)	(0.813)
	Year $\times$ Treated	-0.017	-0.098**	-0.080***
		(0.646)	(0.010)	(0.003)
	$ME[E(S\|S>0)]$	-0.007	-0.035**	-0.031***
		(0.454)	(0.015)	(0.007)
	Change (in ppt.)[1]	-0.71	-3.49	-3.09
	Log pseudolikelihood	214.722	270.495	199.660
	Observations	2180	2064	1972
	Left censored Obs.	764	720	682

Notes: Marginal effects (ME) are computed for a person with characteristics according to the median of all treated individuals in the respective regression sample. Observations can vary between regressions as regression samples are only balanced for the specific year-to-year combination analyzed. All regressions include the control variables depicted in table 1. The interaction terms capture the change in the differences on the outcome between before and after the flood (i.e. the effect of the flood). [1] Refers to the change in the saving rate measured in percentage points. The variable *SR* is computed dividing the individualized monthly amount saved by individual monthly income and is censored at 0. P-values are reported in parenthesis and standard errors are clustered on the household level.Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

tion in precautionary savings through moral hazard effects. This phenomenon, also known as the Samaritan's Dilemma (Buchanan, 1975; Coate, 1995) or the Charity Hazard (Raschky and Weck-Hannemann, 2007; Dobes, Jotzo and Stern, 2014), is theoretically convincing, yet, little empirical evidence exists so far.[34] As discussed earlier, the flood victims of the August 2002 event indeed received an immense amount of financial aid (i.e., in addition to the in-kind aid and neighborhood support already mentioned).[35] Compared to disaster aid in other developed countries, compensation was exceptionally high. Linnerooth-Bayer et al. (2001) reported that compensation rates after disasters in several developed countries average around 40% of occurred losses, whereas the Elbe flooding compensation rates provided almost total compensation.[36] Although we have little information on pre-event expectations on disaster compensation, one might nevertheless suspect that the extraordinary disaster aid of the Elbe flood was at least somewhat unexpected. While the generous disaster aid surely helped to quickly overcome the

[34] There is a small amount of empirical literature on the existence of the Samaritan's Dilemma in natural hazard insurance. While the studies by Kunreuther (1978) and Browne and Hoyt (2000) failed to find that government aid crowded out purchasing of private disaster insurance, the studies by van Asseldonk, Meuwissen and Huirne (2002), Botzen, Aerts and van den Bergh (2009), Brunette et al. (2013), Kousky, Michel-Kerjan and Raschky (2013), and Deryugina and Kirwan (2016) report evidence that supports this argument.

[35] In their analysis of the political consequences of the provided aid in the aftermath of the Elbe flood of 2002, Bechtel and Hainmueller (2011) implicitly assume high compensation rates, which is in line with our line of argument.

[36] In line with this finding, Horwich (2000) reports that the governments of disaster-prone Japan traditionally provide only minimal disaster compensation in order to prevent negative incentive effects.

direct consequences of the flood disaster, it is also highly possible that the enormous level of aid indeed caused a moral hazard effect that led to a reduction in self-insurance via precautionary saving. The explanation of decreased saving by the existence of a Samaritan‘s Dilemma is further supported by the observed time-pattern in our estimation results. The strong reaction in saving behavior happened in 2004 and 2005, and hence after most financial aid was already been paid out.

2.6 Placebo and Robustness Tests

In order to study the appropriateness of our identification strategy and the robustness of our estimation results, we present and discuss several additional estimation results in this section. First and most important, we study whether the assumption of parallel trends in the treatment and control group holds true in the absence of the treatment. As inference in the differences-in-differences approach is based on this assumption, we shed some light on this issue by conducting a falsification test. This strategy estimates placebo differences-in-differences regressions that use the same basic specification that was explained in section 2.3.4 and employed in section 2.4; the only difference is that we assume the flood occurred at some arbitrary point in time before the actual occurrence in August 2002. Whenever the identified differences between the treatment and the control group indeed result from the treatment, we should find that the interaction effect between the treatment and year dummy variables is insignificant in the placebo treatment.

For our placebo treatment, we assume that the flood had already occurred in August 2000 and thus two years before it actually took place.[37] All respondents questioned before August 2000 comprise the pre-treatment sample, and all respondents questioned in 2001 and before August 2002 comprise the post-treatment observations. We included only respondents which took part in the SOEP in between 2000 and 2002. In order to construct the control and the treatment group of the placebo treatment, we use the same procedure as described in section 2.3.4.

The estimation results for individual saving are displayed in table 4. In contrast to the results reported in Section 2.4, the relevant interaction effect between the year and treatment dummy variables is insignificant for both estimations (i.e., the difference between 2000 and 2001 and the difference between 2000 and 2002). Thus, we find no evidence for differing trends between the treatment and the control group in our placebo treatment. In table 5 we show the corresponding estimation results for individual saving rates. Again, we identify no difference in the trends of the treatment and control group.

Throughout the previous empirical analysis of the extensive decision to save, we defined the

[37] The geographic data that was used to divide the SOEP participants into the treatment and the control group was unavailable before 2000. We therefore cannot conduct placebo treatments for earlier points in time.

Table 4: Placebo Estimation Results Individual Saving Behavior

Model	**Variable**	**2000/01**	**2000/02**
Tobit	Dept. Var.: S	(I)	(II)
	Year	-0.393	-15.477
		(0.959)	(0.107)
	Treated	-50.814	-43.741
		(0.116)	(0.166)
	Year $\times$ Treated	-9.201	24.455
		(0.782)	(0.376)
	$ME[E(S\|S>0)]$	-3.638	25.125
		(0.784)	(0.240)
	Change (in percent)[1]	-2.05	5.932
	Log pseudolikelihood	-10.653.729 -10.009.988	
	Observations	2208	2062
	Left censored Obs.	703	670
Probit	Dept. Var.: S_E		
	Year	-0.014 -0.117*	
		(0.815)	(0.076)
	Treated	0.080	0.129
		(0.770)	(0.639)
	Year $\times$ Treated	-0.225	-0.034
		(0.301)	(0.862)
	ME	-0.066	-0.060
		(0.319)	(0.911)
	Change (in ppt.)[2]	-6.581	-0.5991
	Log pseudolikelihood	-1.230.240	-1.192.633
	Observations	2208	2062
OLS	Dept. Var.: S_I		
	Year	12.878*	-2.250
		(0.059)	(0.826)
	Treated	-55.790*	-79.411**
		(0.075)	(0.023)
	Year $\times$ Treated	10.425	30.761
		(0.758)	(0.256)
	Adjusted R^2	0.171	0.122
	Observations	1382	1242

Notes: August 2000 has been chosen as date for the hypothetical flood. ME stands for marginal effect. ME are computed for a person with characteristics according to the median of all treated individuals in the respective regression sample. Observations can vary between regressions as regression samples are only balanced for the specific year-to-year combination analyzed. All regressions include the control variables depicted in table 1. The interaction terms capture the change in the differences on the outcome between before and after the hypothetical flood (i.e. the effect of the flood).[1] Refers to the percentage change in predicted S between the two groups using characteristics of an average treated person in 2002. [2] Refers to the change in the likelihood to save any amount of money due to the flood. The variable S is censored at 0. P-values are reported in parenthesis and standard errors are clustered on the household level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 5: Estimation Results Individual Saving Rate

Model	**Variable**	**2000/01**	**2000/02**
Tobit	Dept. Var.: *SR*	(I)	(II)
	Year	0.003	-0.013**
		(0.618)	(0.022)
	Treated	-0.037*	-0.034*
		(0.050)	(0.064)
	Year $\times$ Treated	-0.006	0.024
		(0.800)	(0.180)
	ME $[E(S\|S>0)]$	-0.003	0.011
		(0.784)	(0.170)
	Change (in ppt.)[1]	-0.258	1.136
	Log pseudolikelihood	404.673	351.105
	Observations	2204	2058
	Left censored Obs.	703	670

Notes: August 2000 has been chosen as date for the hypothetical flood.ME stands for marginal effect. MEs are computed for a person with characteristics according to the median of all treated individuals in the respective regression sample. Observations can vary between regressions as regression samples are only balanced for the specific year-to-year combination analyzed. All regressions include the control variables depicted in table 1. The interaction terms capture the change in the differences on the outcome between before and after the flood (i.e. the effect of the flood). [1] Refers to the change in the saving rate measured in percentage points. The variable *SR* is computed dividing the individualized monthly amount saved by individual monthly income and is censored at 0. P-values are reported in parenthesis and standard errors are clustered on the household level.Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

referring saving variable as:

$$S_E = \begin{cases} 1 \mid S > €0 \\ 0 \mid S = €0 \end{cases}$$

Thus, even individuals with very small but positive reported savings were considered savers. However, one might argue that in the context of disasters, a very low saving volume of only a few Euros should not be considered as precautionary savings. As a second stability test, we defined savers as individuals who declared that they saved more than EUR 50 a month, i.e.:

$$S'_E = \begin{cases} 1 \mid S > €50 \\ 0 \mid S \leq €50 \end{cases}$$

Table 6 presents the corresponding estimates at the extensive margin. The results remain qualitatively unaffected, while the calculated marginal effects slightly increase. Moreover, the results are now significant on the 99% confidence level.
Finally, we study whether our estimation results still hold true when we only look at household heads instead of individual household members. In other words, instead of including all household members we only focus on the household heads without imputing any intra-household

Table 6: Stability of Estimation Results at the Extensive Margin

Model	**Variable**	**2002/03**	**2002/04**	**2000/05**
Tobit	Dept. Var.: S'_E	(I)	(II)	(III)
	Year	0.029	-0.013	0.009
		(0.545)	(0.794)	(0.874)
	Treated	0.019	-0.022	0.031
		(0.940)	(0.931)	(0.903)
	Year $\times$ Treated	-0.270	-1.054***	-0.735***
		(0.389)	(0.001)	(0.001)
	ME	-0.107	-0.360***	-0.286***
		(0.387)	(0.001)	(0.001)
	Change (in ppt.)[1]	-10.742	-35.975	-28.605
	Log pseudolikelihood	-1.279.031	-1.221.087	-1.185.816
	Observations	2188	2068	1974
	Left censored Obs.	764	720	682

Notes: Marginal effects (ME) are computed for a person with characteristics according to the median of all treated individuals in the respective regression sample. Observations can vary between regressions as regression samples are only balanced for the specific year-to-year combination analyzed. All regressions include the control variables depicted in table 1. The interaction terms capture the change in the differences on the outcome between before and after the flood (i.e. the effect of the flood). [1] Refers to the change in the likelihood to save any amount of money due to the flood. The variable S_E is 0 for all persons saving EUR 50 or less. P-values are reported in parenthesis and standard errors are clustered on the household level. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

allocation of household savings. Doing so generally decreases the number of observations but also reduces measurement error in our dependent variable. Due to the small number of observations, we decided to pool all observations from 2002-2005 and include year dummies, our treatment dummy and year dummies interacted with our treatment dummy for all years except for our baseline year (2002). The resulting estimation results are displayed in table 7.

The estimates for total savings and for the extensive decision to save remain qualitatively unchanged. The estimation results for the decision to save at the intensive margin are also similar; however, the effect for the period between 2002 and 2003 is now significant at the 10% level. The pattern of the estimates for the individual saving rate is very similar to the one in our benchmark regression as reported in table 2. The results differ from the earlier reported results only in so far as the effect in 2004 is larger in size. To sum up, the estimates at the level of household heads demonstrate that our benchmark results are not biased by the construction of our individual saving measure and deliver strong support that the flood has affected saving behavior in the described way.

2.7 Summary and Conclusion

In this paper, we presented empirical evidence illuminating the ways in which severe natural disasters might influence individual saving behavior. Using the example of the August 2002 flood catastrophe in Europe, we showed that individual saving decisions were strongly depressed by

Table 7: Estimation Results Household Saving Behavior

Model:	**Tobit**	**Probit**	**OLS**	**Tobit**
Dependent Var.:	(S)	(S_E)	(S_I)	(SR)
	(I)	(II)	(III)	(IV)
Year3 (2003 = 1)	-3.241	-0.036	9.183	-0.003
	(0.864)	(0.527)	(0.686)	(0.615)
Year4 (2004 = 1)	-11.473	-0.046	22.805	-0.005
	(0.570)	(0.435)	(0.330)	(0.462)
Year5 (2005 = 1)	-12.503	-0.068	4.166	-0.010
	(0.616)	(0.315)	(0.868)	(0.226)
Treated	-36.209	0.012	100.166	-0.011
	(0.757)	(0.968)	(0.643)	(0.802)
Year3 × Treated	-134.242	-0.125	-341.152*	-0.036
	(0.188)	(0.689)	(0.094)	(0.406)
Marginal effect	-554.751	-0.047	–	-0.015
	(0.217)	(0.687)		(0.426)
Change	-12.71%	-4.71 ppt.	–	-1.54 ppt.
Year4 × Treated	-341.631***	-0.828**	-308.233	-0.126***
	(0.006)	(0.023)	(0.238)	(0.003)
Marginal effect	-122.858**	-0.321**	–	-0.046**
	(0.013)	(0.015)		(0.012)
Change	-28.15%	-32.10 ppt.	–	-4.57 ppt.
Year5 × Treated	-198.432**	-0.506*	-118.224	-0.072**
	(0.025)	(0.095)	(0.399)	(0.026)
Marginal effect	-77.704**	-0.198*	–	-0.028*
	(0.046)	(0.082)		(0.060)
Change	-17.81%	-19.83 ppt.	–	-2.81 ppt.
Log pseudolikelihood	-10.478.447	-1.174.316	NA	-5.407
Observations	2024	2024	948	1976
Left censored	702	NA	NA	693
Adjusted R^2	NA	NA	0.147	NA

Notes:Reference year for all year dummies is 2002. Personal characteristics refer to the head of the household. Household heads are classified as such by the SOEP. Marginal effects are computed for a household head with characteristics according to the median of all treated household heads in the respective regression sample. The marginal effect for the tobit model is on the expected saving volume conditioned on saving any amount of money ($E[S|S>0]$). The interaction terms capture the change in the differences on the outcome between before and after the flood (i.e. the effect of the flood). Change in case of the tobit model is the percentage change in saving volume due to the flood to average saving volume of a treated individual in 2002. The change in percentage points in case of the probit model is the change in the likelihood to save any amount of money due to the flood. The change in case of the saving rate is the change in the saving rate due to the flood expressed in percentage points. The variables S and SR are censored at 0. SR is computed dividing monthly household net saving by monthly household net income. P-values are reported in parenthesis and standard errors are clustered on household level. Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

the flood event. While we cannot provide a formal test for this line of reasoning, the available empirical evidence supports the hypothesis that the reduction of savings was the consequence of the generous financial support that German policymakers provided to flood victims in the aftermath of the catastrophe. While this policy helped to quickly overcome the direct consequences of the disaster quite, it likely caused the Samaritan's Dilemma by decreasing the incentive for precautionary saving among affected individuals. Thus, our findings highlight the tradeoff between short-term disaster relief and long-run moral hazard effects. Our results must be carefully interpreted, as we derived them from a natural experiment in a highly developed country. Provided our explanation that the presence of the Samaritan's Dilemma is correct, such a dilemma can only occur as a consequence of high compensation rates. Thus, our findings can only be transferred to situations with comparably generous disaster aid. Our finding that natural disasters can affect individual saving behavior might also help to explain why natural disasters tend to have long-run growth effects. Whenever saving rates permanently decrease as a consequence of a natural disaster, this translates into lower per-capita growth. However, in order to find out whether natural disasters do indeed have long-run consequences for individual saving behavior, it would be necessary to track individual saving behavior for an even longer period of time than the three year follow-up period of our empirical analysis. While this perspective is longer than the one taken in most of the existing empirical literature, it would be intriguing to study individual saving decisions for additional years. However, due to the fact that Saxony experienced an additional (although much less severe) flood in 2006, we have to refrain from extending our econometric analysis beyond 2005.

3 Foreign Education and Domestic Productivity[38]

"[...] while the individual man is an insoluble enigma, in the aggregate he becomes a mathematical certainty."

(Sherlock Holmes in: Sign of the Four)

3.1 Introduction

Over the course of the last decades studying abroad has become increasingly popular and is officially promoted in many countries. National governments provide significant amounts of public resources to promote programs for cultural, economic and educational exchange.[39] In 2013 the Fulbright program alone allocated financial funds totaling USD 418.46 million (Fulbright, 2013). Moreover, it appears that globalization encompasses not only the consolidation of economic, but also educational ties. In 2010 some 4.1 million students were enrolled in tertiary educational institutions outside their country of origin, more than tripling from about 1.3 million in 1990 (OECD, 2014). This development outpaced growth of global enrollment in tertiary education by about 1 percentage point per year.[40]

Notwithstanding the growing popularity of foreign education, little is known about the impact internationally educated students have on their home country. In a recent report Engberg et al. (2014) analyze outward mobility scholarships in several countries and conclude that national governments promote international tertiary education inter alia as means of economic development. This suggests that policy makers believe foreign education is a tool to improve domestic economic prosperity.[41] Similarly, Baumol, Litan and Schramm (2007) argue that it might be cost efficient for some (developing) countries to rely on international mobile students as means of improving the quality of their respective domestic educational system. The authors further remark that such improvements in educational quality will actively foster economic prosperity. However, little empirical evidence exists so far (Engberg et al., 2014).

[38]Acknowledgements: My thanks go to Emma Aisbett, Michael Berlemann, Julain Donaubauer, Barry Eichengreen, Thibault Fally, Johannes Jarke, Aprajit Mahajan, Grischa Perino, and Max Steinhardt. Their support and comments are greatly appreciated.

[39]Reasons for public financial support are manifold. The *German Academic Exchange Service* (DAAD) supports over 100.000 students and researchers annually in order to promote scientific, political, and social progress around the globe (see mission statement DAAD, last visit: 4.12.2015). Similarly, the *Bureau for Education and Cultural Affairs* of the U.S. government aims to foster cultural understanding. Programs such as *Fulbright* and *EducationUSA* provide financial and educational resources for people interested in learning about American culture (see mission statement, last visit: 24.01.2016).

[40]Global enrollment in tertiary education grew from 66.9 million in 1990 to 177.684 million in 2010 (UNESCO Institute for Statistics, 2009, 2012). Hence, growth in foreign education (constant annual growth of about 6%) outpaced enrollment in tertiary education (constant annual growth of about 5%) by about one percentage point.

[41]Prominent examples for such a strategy are South Korea, Singapore, Taiwan, and Hong Kong. Also some cities/regions in China and India followed such strategies.

The present paper investigates whether foreign education is indeed beneficial to sending economies and thereby provides empirical evidence to evaluate the returns public support of foreign exchange programs might yield. In order to analyze possible returns, I focus on total factor productivity (TFP) growth. TFP has been identified as the main driver of worldwide income differences (Klenow et al., 1997; Hall and Jones, 1999) and hence is an important determinant of economic development.[42] Moreover, TFP is positively affected by human capital accumulation (cf. Benhabib and Spiegel, 1994; Bils and Klenow, 2000) to which foreign education undoubtedly contributes.
I find that foreign education in the U.S. has a positive impact on domestic TFP growth, especially if the sending economy is classified as developing. The result is based on an unbalanced panel-data-set comprising data on 103 countries spanning at maximum from 1991 to 2011. Data include TFP growth rates, the amount of foreign students studying in the U.S., and some basic control variables. The choice of the U.S. as host country is motivated by the facts that it attracts most of the world's mobile students, has the world's highest-ranked universities, and is generally considered the world's technological leader.[43]

The structure of the paper is as follows. The next section presents the theoretical foundation which leads to my hypotheses. In section 3.3, I discuss the related empirical literature and how it connects to this study. Subsequently, section 3.4 presents the data employed and introduces the empirical strategy. Section 3.5 reports results of my empirical analysis and section 3.6 discusses issues of causality. Finally, section 3.7 concludes.

3.2 Total Factor Productivity Growth and Foreign Education

Total factor productivity growth is decisive for economic prosperity, as in the long run technical progress is the major force driving income per capita growth (Easterly, 2001). Moreover, total factor productivity explains a large proportion of the cross-country variation in income per capita (Hsieh and Klenow, 2010). Hence, increasing total factor productivity nurtures economic development. Several factors that directly or indirectly affect total factor productivity have been identified. These include education, imports, institutions, and geographical predicaments. Yet, for some countries the perhaps most important determinant of TFP growth is innovation.[44]

Innovation usually materializes in new technologies (cf. Isaksson, 2007*b*). New technologies can be understood as result of purposeful and profit orientated investment into research and development (cf. Grossman and Helpman, 1991; Aghion and Howitt, 1992). Such investments,

[42]To the contrary, (Mankiw, Romer and Weil, 1992) argued that the better part worldwide income differences can be explained by differences in human capital.
[43]Please refer to subsection 3.4.1 for further information.
[44] Isaksson (2007*b*) provides a thorough overview on the literature and determinants of TFP growth.

however, require resources and institutions that are readily available only in a small group of economically developed countries. Indeed, data from the UNESCO reveal that in 1996 about 90 percent of the world's investments into research and development (R&D) were carried out by high-income countries. This share decreased to about 70 percent in 2013, as R&D expenditures soared in some Asian countries - mainly China and South Korea.[45] Nonetheless, high-income countries remain the major contributor to worldwide R&D expenditures OECD (2015). It is therefore not surprising that these countries are on or just below the 'World Technology Frontier' and that the development of new technologies (i.e. innovation) is their primary source of TFP growth (Isaksson, 2007*a*).

In contrast, TFP growth in developing economies bases on adoption of existing advanced technologies.[46] Resources to develop new technologies are extremely scarce in these countries making it costly to focus on TFP growth driven by innovation. Fortunately, technologies usually diffuse quickly around the world, e.g. through trade and foreign direct investment (see Keller, 2004, for an overview on the literature of knowledge diffusion). Therefore, a cost efficient strategy could be to adopt and implement existing technologies.

However, developing countries encounter difficulties implementing advanced technologies effectively (Evenson and Westphal, 1995). This means that it is challenging for developing countries to reap the full productive potential of an advanced technology.[47] Several reasons for such difficulties are discussed in the related literature (see Acemoglu, 2015, for a recent discussion). One proposed explanation is that technological change is skill-biased.[48] This implies that a certain technology requires the skill-set prevailing in the inventing country to work effectively. Acemoglu and Zilibotti (2001) show that even if all countries employ the same technology, differences in cross-country skill-sets will lead to cross-country productivity differences. Hence, to effectively make use of an advanced technology, a developing country has to find ways of aligning its skill-set with the one prevailing in the innovating economy.

In order to infer how cross country skill-sets can be aligned, it is critical to determine why skill-sets differ originally. Autor, Katz and Krueger (1998) provide empirical evidence for a link between skills, wages and technology concluding that education plays a decisive role in the creation of skills. Similarly, Goldin and Katz (2008) conclude in their book '*The Race between Education and Technology*' that technological change and education worked hand in

[45] Cf. UNESCO Institute for Statistics (UIS), GERD expenditures in 2005 USD at PPP. URL: http://data.uis.unesco.org/Index.aspx

[46] The term advanced technologies refers to technologies which are invented in industrialized countries and are more productive than existing ones. I refer to a country as industrialized if it is classified as *High Income* and as developing if it is classified as *Low*, *Lower Middle* or *Upper Middle Income* by the *World Bank* in 1991 or the first year data is available. Please refer to B.7 for a complete list of countries and their classification.

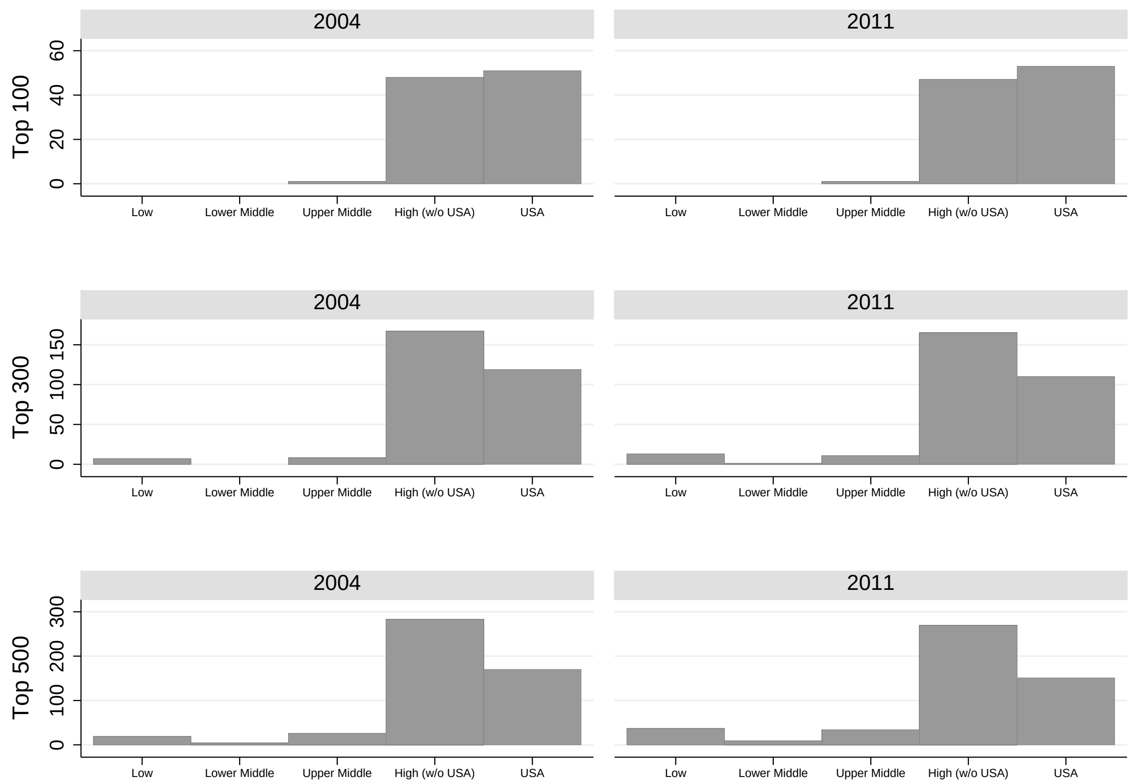
[47] Effective use of an technology means that a technology can be used at is full productive potential.

[48] Alternative explanations are brought forward by the literature on localized and appropriate technology (Atkinson and Stiglitz, 1969; Basu and Weil, 1998) as well as by the literature on induced innovation (Kennedy, 1964; Drandakis and Phelps, 1966)

hand to make the 20th century the 'American Century'; and Nelson and Phelps (1966) claimed that "educated people make good innovators"(Nelson and Phelps, 1966, p.66). This suggests that education is an important determinant in the formation process of country specific skill-sets. Now, if education is so important for the development of skills, the answer to how align skill-sets is rather obvious: developing countries simply have to devote more resources to the educational sector.

At least two caveats remain, however. First, even though education will eventually alter the average skill-set of a developing economy, it might take some time. Goldin and Katz (2008) argue that the 'American Century' had its roots in the high-school-movement, starting around 1915. Subsequently, after World War II, it was consolidated by establishing a multifaceted higher education system. Thus, the current skill-set in the U.S. was generated over decades and it will take developing economies some time to achieve a similar standard. Second, the quality of education might also play a decisive role (cf. Baumol, Litan and Schramm, 2007; Hanushek and Woessmann, 2012). For instance, in a developing economy improvements in the quality of the educational system and the skill-set of its educators, might considerably accelerate the process of skill-set alteration.At present, there are vast cross-country differences in the quality of universities as can be inferred from figure 2. Figure 2 shows the distribution of top-ranked

Figure 2: Distribution of Top Ranked Universities by Income Group



Notes: The ranking is based on ShanghaiRanking (2015). Income groups refer to the classification by the World Bank in the year 1991 or the first year available

universities by country income group. In 2004 nearly all universities ranked in the world's top 100 were located in high-income countries (*Moscow State University* being the exception). Moving down the ranking, the picture remains relatively stable. Only 14 universities from non high-income countries are being ranked in the top 300 and 45 in the top 500. The figure shows that by 2011 the distribution shifted slightly indicating some progress by developing countries in improving the quality of their higher education. Yet, the bulk of the world's top ranked universities remains to be located in industrialized economies, especially in the U.S.
Combining the two caveats it can be concluded that the process of skill-set alteration is time consuming but might be accelerated if certain skill-sets can readily be imported.

The import of skills can increase productivity by either enlarging the share of potential innovators in the domestic population and/or by enabling a larger share of the domestic population to effectively implement imported technologies. One mechanism to import skills, which are ultimately bound to people, is by foreign education. Students educated abroad obtain the skills prevailing in the host country transferring them back to their home country upon their return. Therefore, my first hypothesis states that a country's level of foreign educated students positively affects domestic TFP growth through the import of new skills.

> **Hypothesis 1:** On average, the more students a country sends for education abroad, the higher its subsequent TFP growth rate will be.

Note that sending one's students to a country close to the technological frontier, i.e. a country regularly inventing advanced technologies, is preferable compared to sending them to a country with low innovative potential. Additionally, there is more to be gained by foreign education for a country with a low level of productivity than for a country that is already close to the technological frontier. The reason is that countries on the technological frontier, i.e. industrialized countries, have similar skill-sets, use advanced technologies effectively, and possess high levels of education. Due to diminishing returns there should not be as much to gain as for a country with lower levels of education and/or being further away from the technological frontier, i.e. a developing country. Hence, my second hypothesis states that it will be especially beneficial for developing countries sending students abroad.

> **Hypothesis 2:** The effect on TFP growth by sending students abroad, will, on average, be stronger for developing than for high-income countries.

As this section focused on the theoretical foundation, the next section will discuss empirical evidence concerning the link between skills and productivity as well as the link between foreign education and economic prosperity.

3.3 Survey of the Related Literature

Following Acemoglu and Zilibotti (2001) I argue that advanced technologies are relatively skill-complementary. This implies that cross-country differences in productivity do not necessarily arise from technology but from a technology-skill mismatch. To alleviate this mismatch I propose that foreign education serves as a mechanism aligning skill-sets. So far the related empirical literature has not focused on this particular argument, but some studies focus on the link between skills and productivity as well as on foreign education and economic prosperity.

Showing that certain skills are necessary for the effective implementation of a technology is empirically challenging, but some suggestive evidence exists. In a rather early work, Kalirajan (1991) studies why especially developing countries show limited success in improving agricultural productivity. Using the "green revolution" in rice technology as a productivity treatment, he shows that Indian farmers' education poses a major constraint on the effective use of the new technology. Mohnen and Röller (2005) provide empirical evidence for various countries and industries, indicating that the lack of skills is the most important obstacle for successful innovation; Leiponen (2005) reports that firms benefit less from innovations if their employees lack sufficient skills; and Bloom et al. (2013) show that management skills have a positive causal impact on firm productivity. Finally, Giorcelli (2016), using a unique firm-level panel-data-set, shows that not only technology, but also skills for its effective use are an important determinant of firms' productivity. Exploiting a quasi-natural experiment, Giorcelli finds that training trips of Italian managers to the U.S., as part of the Marshall Plan to rebuild Europe, had a substantial impact on firms' productivity. Moreover, while the positive effect on firms' employment, profit, and productivity was long-lasting if the managers were trained in the U.S., the impact of solely new technology (i.e. new machinery) without the equivalent training was also positive but only short-lived. All these findings suggest that skills and technology are complementarities.

Most of the economic literature dealing with foreign students focuses on their impact on the host country (e.g. Borjas, 2005; Chellaraj, Maskus and Mattoo, 2008; Borjas, 2009; Dreher and Poutvaara, 2011; Borjas and Doran, 2015). In contrast, only a few studies focus on the sending country. However, these studies report evidence that there is a positive link between foreign education and domestic economic development. Kim (1998) finds that foreign education has a positive effect on income per capita growth. He argues that foreign education brings advanced knowledge to the sending economy. Hence, his proposed transmission channel is similar to the one in this study. Differing from this study Kim (1998) focuses on income growth using cross-section data for empirical analysis. Contrary to Kim (1998), Park (2004) investigates whether student flows function as a channel for the diffusion of R&D. Using panel cointegration techniques Park (2004) shows that there is a positive correlation between TFP and student flows weighted by R&D capital-stocks of the respective host countries. While his research question is seemingly similar to mine, Park's study is embedded in the literature on R&D spillovers started by Coe and Helpman (1995). The aim of this literature differs, as it investigates how technologies (measured by R&D expenditures) diffuse. Therefore, Park (2004) only focuses on OECD countries, as this is the group of countries musters the bulk of worldwide annual R&D investments. Le (2010) extends the analysis by focusing on R&D weighted student flows from developing countries to developed ones. Their studies indicate that student outflows positively influence the level of domestic TFP. However, while both authors apply panel cointegration techniques, only the study by Park uses a satisfyingly long time period (1971 to 1990).[49] Finally, Spilimbergo (2009) considers the effect of foreign education on democracy in the sending country. Using dynamic panel estimation techniques he finds that studying in a foreign democratic country increases the level of democracy in the sending country. While seemingly unrelated, this finding does support the hypothesis that foreign education, conditional on the host country, is beneficial for economic development, as democracy seems to be beneficial as well (see e.g. Acemoglu et al. (2014) for the positive link between democracy and economic growth). To conclude, the small group of related literature consistently points to a positive relationship between foreign education and domestic economic development.

3.4 Empirical Strategy and Data

3.4.1 Data and descriptive statistics

The data-set used in this study comprises unbalanced panel-data of 102 countries, excluding the U.S., ranging from 1991 to 2011. It includes data on the number of foreign students in the U.S., imports of goods and services from the U.S., levels of education, population aged 20 to 29, and total factor productivity. The following describes the computation of TFP and provides a detailed description of the data.

In order to study the effect of foreign education on domestic productivity, I follow the literature on knowledge diffusion and construct a basic measure of economy-wide TFP using growth accounting (e.g Coe and Helpman, 1995; Keller, 2004; Park, 2004; Le, 2010). Several reasons can be brought forward justifying the procedure in the present context.

First, one has to recognize that there is no uniformly accepted definition of TFP and therefore it is difficult to interpret. Technological change is notoriously difficult to measure and one has to be careful to interpret changes in TFP as such (see e.g. Baier, Dwyer and Tamura, 2006; Caselli, 2010; Gollin, 2010, for a critical perspective). However, there is a long tradition in the economic literature to interpret changes in TFP as changes in technological progress and/or productivity

[49] Le (2010) considers a time period of at maximum eight years ranging from 1998 to 2005.

(e.g. Tinbergen, 1942; Abramovitz, 1956; Solow, 1957). At present, this interpretation can still be justified if TFP captures externalities; e.g. from R&D or education (skills) (cf. Caselli, 2010). Since I investigate the externality arising from a change in skills-sets – leading to more effective (i.e. productive) technology use – TFP is an appropriate measure and its progression over time can to some extent be interpreted as a change in productivity.
Second, I want to study the positive externality arising from education (foreign or domestic) and therefore use capital and raw labor, instead of human capital, as inputs to compute TFP.[50]
Third, the use of economy-wide TFP is appropriate since skills acquired through domestic and foreign education affect productivity of capital and labor in all sectors of the economy.[51]

Total factor productivity is computed as follows. Assuming that output follows a Cobb-Douglas production function with constant returns to scale, I compute TFP according to:

$$TFP_{it} = \frac{GDP_{it}}{L_{it}^{\alpha} \cdot K_{it}^{1-\alpha}}, \tag{3}$$

where α represent the labor's share of output. In line with the literature I assume that the labor share is constant over time and across countries (e.g. Gollin, 2002; Le, 2010). Furthermore, I assume that the labor share amounts to about two-thirds which is in accordance with available data.[52] Data on countries' gross domestic product (GDP_{it}) and employment (L_{it}) come from the World Bank's *Development Indicators* while data on country specific capital stocks (K_{it}) is taken from Berlemann and Wesselhöft (2014).[53] The resulting TFP levels are then used to compute TFP growth rates, $TFPgr_{it}$.

Table 8 reports on the progress of TFP over time and across countries for a selected sample. Column (1) shows country i's mean annual growth rate for the maximum time period available while column (2) shows its level of TFP relative to the level of TFP in the U.S. for the year

[50] Using only the variation in capital and raw labor should not capture any positive externality arising from changes in skills-sets. Accounting for human capital instead of raw labor as input in TFP computation would blur this effect. Thus, changes in human capital are part of computed total factor productivity growth and can then be analyzed in subsequent TFP growth regressions.

[51] Relying on basic economy-wide TFP is common in related studies (e.g. Coe and Helpman, 1995; Park, 2004; Le, 2010)

[52] Using data for OECD countries shows that the labor share remained very stable over time and across member countries amounting to about two-thirds. Relevant data is not available for many developing countries but Gollin (2002) shows that the labor share is quite similar across countries (developed and developing) once accounted for differences in employment and self-employment. Note that even if the assumption is flawed, any country-specific and time-invariant measurement error will be absorbed by the fixed effects (see section 3.4.2)

[53] Capital stock data computed by Berlemann and Wesselhöft (2014) is based on investment data from the World Bank's *Development Indicators*. Hence, all data for computation of TFP comes basically from the same source. At the time this study was first stared, there was no similar data source available. However, I am aware that similar data is now provided by the latest version of the popular *Penn World Tables* (PWT). Still, the used capital stock data has the advantage that its estimation is more precise. Results using PWT data can be obtained from the author on request.

Table 8: Selection of TFP data across time and countries

	mean annual TFP$_i$ growth rate	$\frac{TFP_{i,1996}}{TFP_{USA,1996}}$
	(1)	(2)
Venezuela	-0.003 (98)	0.344 (36)
Kenia	-0.000 (93)	0.074 (85)
Japan	0.001 (91)	0.924 (4)
Romania	0.005 (76)	0.164 (63)
Canada	0.006 (71)	0.731 (17)
Germany	0.007 (64)	0.693 (19)
U.S.A.	0.009 (58)	1.000 (2)
Hong Kong	0.017 (31)	0.832 (7)
Mozambique	0.024 (13)	0.037 (100)
Poland	0.024 (12)	0.300 (39)
Argentina	0.027 (7)	0.447 (32)

Rank in parenthesis. Mean TFP growth rate is refers to the maximum time period available. 1996 is chosen in column (2) as it is the first year the sample is balanced. There are 103 countries in the sample including the U.S.

1996.[54] The numbers in parenthesis represent the country's rank. Obviously, there are great differences in total factor productivity across countries and over time.

The present study focuses on one host country only: the United States of America. Several reasons justify such constraint. First, the U.S. is widely considered the world's technological leader and invests by far the highest amount in R&D (Acemoglu and Zilibotti, 2001; Caselli and Coleman, 2006). Second, over half of the top 100 universities in the world are located in the United States (cf. figure 2). Third, the U.S. is the major host country of foreign students in the world. Over the time period 1955 to 2005 it hosted on average about 30 percent of the world's mobile students (cf. Spilimbergo, 2009). Even though the U.S. recently lost some ground to other industrialized countries, it remains the world's top host country in 2014 receiving about 22% of all globally mobile students (Project Atlas, 2015). Fourth, data on foreign students comes from one source guaranteeing consistency.[55] Fifth, all foreign students are confronted with the same institutional, cultural, educational and judicial conditions. Finally, the setting allows to differentiate between foreign students from developing and industrialized countries studying in an industrialized economy.

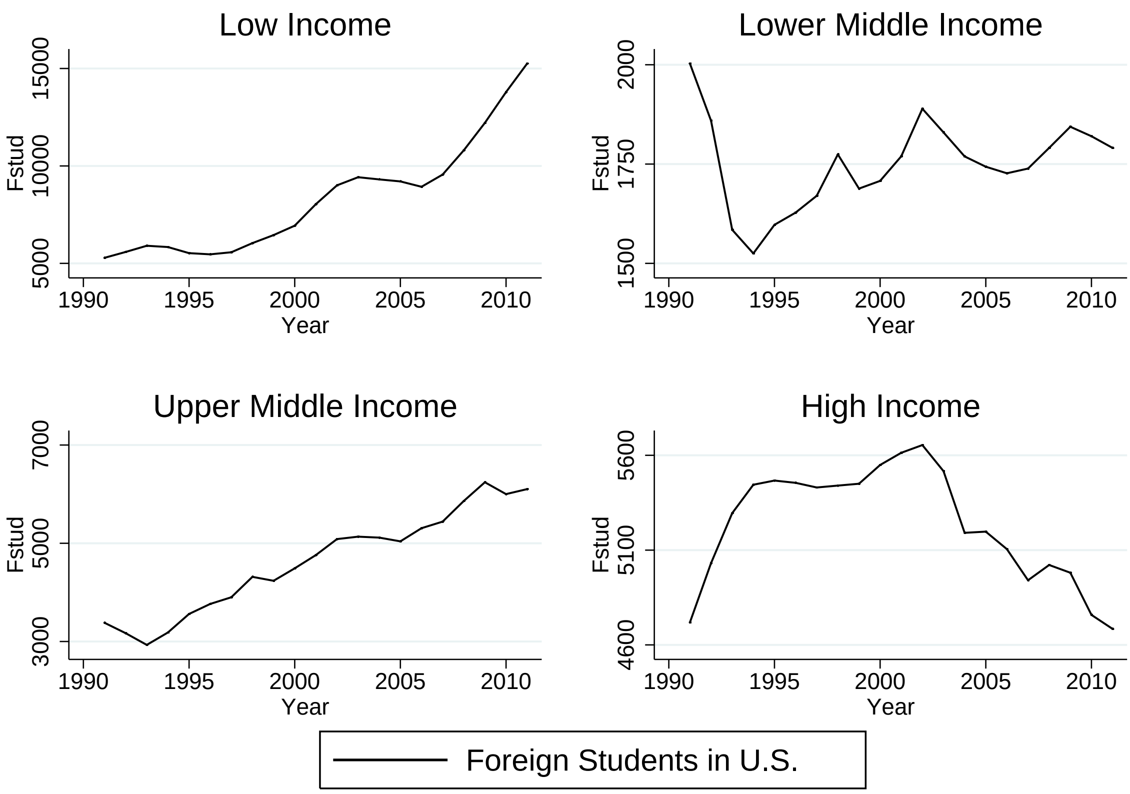
Data on foreign students studying in the U.S. are taken from the *Open Doors Database* which is maintained by the *Institute of International Education* (IIE, 1991-2011). The IIE conducts

[54] 1996 is the first year TFP data is available for all included countries.

[55] Consistency is problematic using other data sources as they all rely on national statistical offices with varying definitions of foreign student. For instance, several European countries differentiate between mobile and international students. Data provided by the UNESCO can differ depending on the direction of students, outbound or inbound. If everything would be correctly reported the amount of outbound students from Egypt to the U.S., reported by authorities in Egypt, should be the same as the amount of inbound students from Egypt into the U.S., reported by the U.S. authorities. Unfortunately this is rarely the case. See also Spilimbergo (2009) on this point.

annual surveys since 1948 and is supported by the U.S. Department of State since the early 1970s. Data on international students are based on an annual Census as well as data from several governmental sources.[56] The data are available online from 2000 onward. Prior to 2000 the data are taken from Open Door's annual reports (hard copy). Information on the country of origin is available for all years included in this study but the group of countries having students studying in the U.S. generally increases over time. Figure 3 shows the evolution of foreign students in the U.S. by the World Bank's income classification. While the amount of foreign students from high income countries declined over the last decade, the U.S. experienced an increase of students from low and upper middle income countries.

Figure 3: Foreign Students Studying in U.S. by Income Group



Notes: Income classifiaction according to World Bank in 1991 or first year available. Data on foreign students in the U.S. comes from IIE (1991−2011)

Ideally, I would like to include some measure of the domestic R&D capital stock. However, data on investments in R&D are rarely available for many developing countries. An alternative is to follow Coe and Helpman (1995) and include bilateral imports of goods and services. The idea is that technologies are incorporated in goods. Hence, technologies from the U.S. become available via exported goods and services. Information on bilateral imports of goods and services was provided by Hühne, Meyer and Nunnenkamp (2014) and comes originally from UN

[56]See on data methodology IIE (1991-2011)

Comtrade. In order to approximate a country's potential level of students, population aged 20 to 29 is included.[57] The data on demographics come from the *United Nations Population Division*. Finally, data to approximate the level of education (average years of schooling) are taken from Barro and Lee (2013). As data are in five year intervals, annual data are computed using linear interpolation. Table 9 shows summary statistics of considered variables. The variables always refer to the sending country (i) and are coded as follows: *TFPrel.US* refers to TFP relative to the U.S., *Fstud* to the level of foreign students in U.S., *Age2029* to the population aged 20 to 29 (in thousand), *HumanCap* to average years of schooling (primary, secondary, and tertiary), and *Imports* to the value of goods and services imported from the U.S. (in million USD).

Table 9: Summary Statistics of Variables Included in Analysis

	Obs.	Mean	Std. Dev.	Min	Max
TFP growth	2031	0.01	0.05	-0.45	0.23
TFP rel. US	2133	0.33	0.28	0.02	1.41
Fstud	2110	4,390.32	11,597.15	10.00	157,558.00
Age2029$^{\diamond}$	2142	8,455.01	28,965.46	41.13	246,798.31
HumanCap	1880	8.04	2.59	0.92	13.02
Imports*	2061	8,409.56	24,158.38	0.36	284,387.31

Notes: Sample is unbalanced. Maximum time dimension ranges from 1991 to 2011.$^{\diamond}$Population is in thousand.*Imports are in million USD. Variables are coded: *TFP rel. US* refers to TFP relative to the U.S., *Fstud* to the level of foreign students in U.S., *Age2029* to the population aged 20 to 29, *HumanCap* to average years of schooling (primary, secondary, and tertiary), and *Imports* to the value of goods and services imported from the U.S.

3.4.2 Empirical model and econometric issues

The hypotheses presented in section 3.2 are tested by estimating the effect of foreign education in a dynamic representation of total factor productivity growth.[58] The specification to be estimated is of the following form:

$$\begin{aligned} TFPgr_{it} = \sigma_0 + \alpha TFPgr_{it-1} &+ \beta \ln(TFPrel.US_{it-1}) + \gamma \ln(Fstud_{it-1}) \\ &+ \delta \ln(Z'_{it-1}) + \lambda_t + \mu_i + \varepsilon_{it}, \end{aligned} \tag{4}$$

where i=1,2,...N indexes countries, t=1,2,...T indexes years, and the vector Z'_{it-1} can include additional explanatory variables. By including the variable $TFPrel.US_{it-1}$ the model explicitly accounts for the potential acceleration in productivity growth a country can achieve by catching up to the U.S. All regressions include year dummies, λ_t, to account for common shocks affecting all countries in a given year. Additionally, all except the pooled OLS regression include

[57]The higher the level of potential students, the higher TFP growth might be. Especially if they are equipped with appropriate skills.

[58]Estimating growth rates is preferable as foreign education is expected to have along lasting effect the productivity, Moreover, TFP growth follows a stationary process which is required to employ the econometric analysis described below.

country dummies, μ_i, to capture country specific effects that are relatively stable over time such as geography, culture and institutions. Finally, ε_{it} represents the regression's error term. In the baseline model *Age2029* is added controlling for the potential level of students in the sending economy. Later, average years of schooling, *HumanCap*, and imports of goods and services from the U.S., *Imports*, are added.

There are several reasons to prefer a dynamic over a static model. First, implementing new skills acquired trough foreign education requires time. Therefore, it is very unlikely that foreign education has an instantaneous effect on TFP growth suggesting that a static model is inappropriate. Second, including lagged TFP growth allows to explicitly model the short- and long-run impact foreign education has on TFP growth. Third, including the lagged dependent variable helps to control for the effect of potentially relevant, yet omitted, variables. Fourth, the dynamic setting helps to control for serial correlation.

Three different estimation techniques will be used: pooled OLS, least square dummy variable (LSDV), and system GMM. First, pooled OLS does not control for country fixed effects and will overestimate the coefficient of the lagged dependent variable. Second, while the LSDV estimation controls for country specific effects, the results will be biased due to the lagged dependent variable and finite T (Nickell, 1981). The bias normally results in smaller estimates of the true coefficient of the lagged dependent variable. Judson and Owen (1999) show that this bias can still be considerable even if T is equal to 30. However, the biased results produced by pooled OLS and LSDV provide a range of the true autoregressive component. Third, the system GMM estimator provides consistent and unbiased estimates. The coefficient of the lagged dependent variable should be inside the range spanned by the former two estimators.
Another econometric problem is that *Fstud* is possibly endogenous. It might be that growth in total factor productivity causes an increase of students studying in the U.S. because with higher TFP growth students are better equipped or simply have better chances to study in the U.S. To control for this endogeneity problem, *Fstud* is teated as potentially endogenous in system GMM and is instrumented by its own second lag. A similar argument can be brought forward for the level of domestic education. Hence, when included, *HumanCap* is also treated as potentially endogenous as will be *TFPrel.US*. Following Feyrer (2007) *Age2029* is treated as predetermined using its own first lag for instrumentation. All other variables enter as specified in equation (4).[59]

[59]Variables are instrumented using its first (second) lag only to avoid having too many instruments.

3.5 Results

Results of the basic model are reported in the first three columns of table 10. As mentioned before, the regression results using OLS and LSDV are biased but span a range including the true estimate for the lagged dependent variable. Indeed, using the consistent and unbiased system GMM estimator, the coefficient of last year's TFP growth is within this range, slightly larger than the downward biased LSDV estimate. However, the unbiasedness of the system GMM estimator might be affected by the instrument count (Windmeijer, 2005; Roodman, 2009*b*).[60] Roodman (2009*a*) suggest to test the robustness of the system GMM estimation by severely reducing the amount of instruments used in the first step. Table B.1 provides results for the system GMM estimations with reduced instrument count. While coefficients keep their signs and remain significant, their magnitude increases somewhat. However, this should be expected as the system GMM estimator becomes less efficient with reduced instrument count.

Returning to table 10 and focusing on the system GMM estimation results, most coefficients show the expected sign. There is a certain persistence in TFP growth as can be inferred from the positive and highly significant coefficient of the lagged dependent variable. $TFPrel.US_{t-1}$ is negatively correlated with future TFP growth indicating that the closer a country is to the productive capacity of the U.S., the slower its domestic TFP growth. That is, the closer a country gets to the productivity level of the U.S., the less the potential for catching up.

The first hypothesis stated in section 3.2 claims that foreign education has a positive effect on domestic TFP growth. Indeed, the coefficient on the level of foreign students in the U.S. is positive and highly significant across specifications. The finding is well in line with the literature cited in section 3.3. It indicates that the more students from a country are studying in the U.S., the higher its future domestic productivity growth will be on average. For instance, increasing the amount of students in the U.S. by one percent increases (short-run) domestic TFP growth by about 0.021 percent (column (3)). Note that the percentage increase in average TFP growth rate is not very informative. To illustrate the magnitude of the effect it is conducive to compute the increase to average annual TFP growth rate in percentage points. The average annual TFP growth rate for the sample is 1.25 percent. Hence, increasing the average amount of foreign students by 1% increases average annual TFP growth by 0.025 percentage points. As the estimated model is dynamic, the long run effect can be calculated by dividing the estimated short-run coefficient by one minus the coefficient on the lagged dependent variable: $\beta/(1-\alpha)$. Thus, in the long-run a 1% increase in the level of students send to the U.S. increases TFP growth on average by about 0.038 percentage points.[61]

[60] Roodman (2009*a*) further notes that instrument proliferation can overfit endogenous variables, fail to expunge their endogenous component, and weakens the power of the Hansen test.

[61] Table B.4 provides estimation results of a static model. Serial correlation is either addressed by clustered (HAC)

Table 10: Estimation Results Dynamic Panel

Dept. Var. TFP growth$_t$	Pooled OLS (1)	LSDV (2)	Sys. GMM (3)	Sys. GMM (4)	Sys. GMM (5)	Sys. GMM (6)
TFP growth$_{t-1}$	0.326***	0.272***	0.278***	0.239***	0.265***	0.216***
	(0.043)	(0.044)	(0.047)	(0.044)	(0.051)	(0.046)
ln TFP rel. US$_{t-1}$	-0.006***	-0.109***	-0.022***	-0.028***	-0.016	-0.016**
	(0.001)	(0.016)	(0.007)	(0.005)	(0.010)	(0.006)
ln Fstud$_{t-1}$	0.002*	0.018***	0.021***	0.021***	0.024***	0.026**
	(0.001)	(0.005)	(0.008)	(0.008)	(0.008)	(0.010)
ln Age2029$_{t-1}$	-0.002	-0.014***	-0.017*	-0.015**	-0.017**	-0.014**
	(0.001)	(0.003)	(0.009)	(0.007)	(0.009)	(0.007)
ln HumanCap$_{t-1}$				0.025*		0.023*
				(0.013)		(0.012)
ln Imports$_{t-1}$					-0.004	-0.009*
					(0.005)	(0.005)
Constant	0.010		0.069	-0.009	0.158**	0.144
	(0.012)		(0.077)	(0.073)	(0.064)	(0.095)
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	No	Yes	Yes	Yes	Yes	Yes
AR(1) Test			0.00	0.00	0.00	0.00
AR(2) Test			0.95	0.95	0.78	0.78
Hansen's J Test			0.15	0.23	0.21	0.33
Number of Instruments			96	98	97	99
Number of Countries	102	102	102	94	101	93
Observations	1911	1911	1911	1765	1853	1707
Adjusted R^2	0.24	0.38				

Notes: Pooled OLS suffers from endogeneity bias as fixed effects are excluded while LSDV suffers from Nickell (1981) bias. System GMM is consistent and unbiased. The lagged dependent variable, *TFPrel.US*, *Fstud* and *HumanCap* are treated as potentially endogenous and are instrumented using their own second lag. *Age2029* is treated as predetermined and instrumented by its own first lag. Lagged imports are not instrumented. For the instrumentation of *Fstud*,*Age2029* and *HumanCap* the collapse option is used to reduce instrument count (see Roodman, 2009*a*). AR(1) and AR(2) are Arellano-Bond test for serial correlation. Hansens J tests the null hypothesis for violation of over-identification restrictions. For all tests p-values are reported. Standard errors are reported in parentheses. For column 1 and 2 HAC standard errors are reported clustered on country-level. For column 3 to 5 robust standard errors are reported using the Windmeijer (2005) procedure. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Somewhat surprisingly, the lagged amount of the domestic population aged 20 to 29 is negatively correlated with current TFP growth. One could expect that the higher the base for potential students, the more students the country actually has and the higher TFP growth should be. However, a larger population aged 20 to 29 could also indicate that parents substitute child quality for child quantity (Galor and Moav, 2002) or that less financial and attentional resources can be devoted to each (potential) student.

The analysis is subsequently extended by adding additional controls. In column (4) the measure for domestic human capital and in column (5) imports from the U.S. are added. In column 6 all variables are included at the same time. All included variables turn out to be statistically significant. While lagged *HumanCap* positively correlates with current TFP growth, the coefficient on lagged imports is negative. One explanation for the negative coefficient could be that the more

standard errors or by using Discroll-Kraay standard errors. The coefficient on foreign education remains positive and statistically significant in all regressions.

goods and services are imported from the U.S. the higher is the need for a skilled workforce to operate the inherent technologies effectively. Note that the amount of countries included in the different estimations varies. Country selection is governed by data availability. Estimation results with a constant sample of 93 countries are reported in table B.2. Results do not change significantly from the ones reported in table 10.

Table 11: Estimation Results for Developing and Industrialized Countries

Dept. Var.	System GMM		LSDV		LSDVc	
TFP growth$_t$	Developing	High	Developing	High	Developing	High
	(1)	(2)	(3)	(4)	(5)	(6)
TFP growth$_{t-1}$	0.203***	0.129	0.266***	0.190***	0.324***	0.251***
	(0.051)	(0.119)	(0.046)	(0.067)	(0.027)	(0.049)
ln TFP rel. US$_{t-1}$	-0.032**	0.096	-0.131***	-0.068	-0.130***	-0.067***
	(0.012)	(0.133)	(0.017)	(0.040)	(0.013)	(0.016)
ln Fstud$_{t-1}$	0.033***	0.014	0.018***	0.001	0.016***	-0.001
	(0.008)	(0.025)	(0.006)	(0.006)	(0.003)	(0.005)
ln Age2029$_{t-1}$	-0.027**	0.004	-0.016***	-0.002	-0.016	-0.017
	(0.011)	(0.007)	(0.003)	(0.003)	(0.013)	(0.012)
Constant	0.120	-0.128				
	(0.124)	(0.205)				
AR(1) Test	0.00	0.00				
AR(2) Test	0.86	0.02				
Hansen's J test	0.82	1.00				
Number of Instruments	35	35			40	40
Number of Countries	77	25	69	24	69	24
Observations	1436	475	1436	475	1436	475
Adjusted R^2			0.39	0.51		

Notes: All regression include time and country fixed effects. System GMM is consistent and unbiased but LSDVc might be preferable if N is moderate (cf. Judson and Owen, 1999). The lagged dependent variable, *TFP rel. US*, and *Fstud* are treated as potentially endogenous and are instrumented using their own second to fourth lag. *Age2029* is treated as pedterimend and is intrumented using its own first to third lag. For the instrumentations the collapse option is used to reduce instrument count (see Roodman (2009*a*)). LSDV suffers from Nickell (1981) bias. LSDVc refers to the bias corrected LSDV estimation using the procedure from Bruno (2005*a*). AR(1) and AR(2) are Arellano-Bond test for serial correlation. Hansens J tests the null hypothesis for violation of over-identification restrictions. For all tests p-values are reported. Standard errors are reported in parentheses. For column (1) and (2) robust standard errors are reported using the Windmeijer (2005) procedure. For column (3) and (4) HAC standard errors are reported clustered on country-level. Significance of estimates in column (5) and (6) is derived by bootstrapping the variance-covariance matrix using 50 repetitions. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

My second hypothesis states that especially developing countries benefit from foreign education. In order to test the hypothesis, I split the sample in developing and industrialized countries and run the baseline model on both samples.[62] As can be inferred from table 11, the coefficient on foreign education remains positive and highly significant for the sample of developing countries but becomes insignificant for the sample of industrialized ones. As the sample for

[62]Results for the fully specified model including all variables are reported in table B.3 in the Appendix. Results do not qualitatively differ from the ones presented here. Note that results using system GMM (esp. column (1) in table 11 and column 6 in table B.3) should be viewed with caution as the panel is no longer one of small T and large N. For an alternative refer to B.5. The table presents results for both samples in a static model. Again, results remain qualitatively comparable.

industrialized countries becomes rather small, the system GMM estimator might not be the preferred estimator anymore (Judson and Owen, 1999).[63] Even though LSDV is biased, it has a relatively small variance compared to the GMM estimator (Arellano and Bond, 1991; Kiviet, 1995). Under certain circumstances, small N, it can thus be preferable to use a bias corrected LSDV estimator (LSDVc). In column (5) and (6) the bias corrected LSDV estimates are reported using the bias correction for unbalanced panels from Bruno (2005*b*), while column (3) and (4) report the original LSDV estimates.[64] The results from the initial estimation using system GMM do not change qualitatively but the magnitude of the coefficient on foreign education is somewhat reduced.

3.6 Issues of Causality

This section elaborates on the causality of the link between foreign education and domestic productivity growth. First, information on the fields of study pursued by foreign students in the U.S. is presented. Second, additional evidence is brought forward supporting the underlying assumption that a sufficient amount of foreign students have to return to their home country after their education in the U.S. is terminated. Finally, the results will be linked to insights from the migration literature which indicate that there are strong economic ties between the country of origin and the diaspora.

Skills obtained by foreign students are likely to be dependent on the field of study they pursue. While some fields of study (and hence skills) can easily be classified as productivity improving others might not. For instance, skills obtained by a student of fine arts will significantly differ from the skills obtained by a student of computer science. It seems reasonable to assume that the latter will obtain and subsequently import skills which are better suited to increase productivity. Therefore, arguing in favor of the positive effect foreign education has on domestic productivity, one would expect that foreign students in the U.S. primarily enroll in fields which are associated with productivity enhancing skills. Table 3.6 shows the amount of U.S. F-1 visas issued by the U.S. State Department between 2008 and 2012 categorized by academic level and field of study.[65] While there are three types of student visas (F-1, J-1 and M-1), F-1 visas are the most relevant – they are issued to foreigners enrolled in a full-time academic program – and most common ones. In fact the share of F-1 visas relative to all types of student visas remained stable over time; 58% in 1997 and 57% in 2011.[66] Contrary to the here presented data on issued F-1 visas, information on the field of study from IIE data, which is used in the empirical analysis, is only available from 2010 onward. However, the distribution of foreign students on

[63] Actually, the estimation becomes unreliable as there seems to be undesired second order serial correlation.
[64] For the LSDVc estimates the bias is approximated up to O(1/T).
[65] Note that only the top four fields of study are listed for each degree type.
[66] See U.S. State Department, Nonimmigrant Visa Statistics, last visit: 1.6.2016.

fields of study are similar. Table B.6 in the Appendix provides further information on IIE data for the top 19 sending countries in 2010.

Table 3.6 shows that foreign students in the U.S. indeed enroll in subjects which are commonly connected with productivity enhancing skills. Especially Business Studies are popular at the Bachelor and Master level while Natural and Engineering Sciences predominate at the Doctorate level. Both fields are relevant for economic development and productivity. On the one hand, advanced management skills seem to be important for economic growth through their impact on inter alia finance, foreign direct investment and entrepreneurship (e.g Nelson and Phelps, 1966; Borensztein, de Gregorio and Lee, 1998; Levine, 2001; Acs, 2006). On the other hand, education in STEM fields (Science, Technology, Engineering, and Mathematics) enhances productivity through technological innovation and adoption (cf. Murphy, Shleifer and Vishny, 1993; Jones, 1995; Lin, 2004; Peri, Shih and Sparber, 2015).[67]

Table 12: Fields of Study by Foreign Students 2008 - 2012

Degree Type		Major Field of Study	Foreign Students	Of Total Foreign Students
Bachelor				
	1	Business, Management, Marketing	173,372	32.4%
	2	Engineering	61,438	11.5%
	3	Liberal Arts and Sciences, General Studies	43,906	8.2%
	4	Social Sciences	37,422	7.0%
Master				
	1	Business, Management, Marketing	146,146	30.4%
	2	Engineering	86,590	18.0%
	3	Computer and Information Sciences	59,000	12.3%
	4	Education	21,377	4.4%
Doctorate				
	1	Engineering	38,201	27.8%
	2	Physical Sciences	16,262	11.8%
	3	Biological and Biomedical Sciences	13,766	10.0%
	4	Health Professions and Related	10,620	7.7%

Source: Ruiz (2014). Notes: The numbers refer to F-1 visas. Original data has been obtained by the Brookings Institute from U.S. Immigration and Customs Enforcement through a Freedom of Information Act Request. Data are aggregates across the 2008 - 2012 time period.

Skills obtained by foreign students have to be transferred back to their country of origin in order to influence productivity growth. That is, foreign students have to return to their country of origin. Somewhat surprisingly there is no official data on students returning from the U.S. to their respective home countries. However, there is evidence suggesting that it is difficult to stay in the U.S. after graduation and/or termination of the F-1 visa status. According to Wadhwa (2012) and Matloff (2013) the U.S. has relatively restrictive visa regulation. For instance, Wadhwa et al. (2009) show that half of all interviewed foreign students would prefer to work in the USA but report it to be difficult, because visa regulations are strict and job perspectives

[67] Walker and Zhu (2013) provide an extensive overview on the benefits of STEM skills.

bleak. Moreover, in order to remain and work in the U.S. after graduation a foreigner needs a H-1B visa. Ruiz (2014) reports that in 2010 only 26% of all H-1B visas approved were issued to foreign students. Given that there were 76.627 H-1B visas approved in 2010, this amounts to about 26.500 H-1B visas granted to former F-1 visa owners. Compared to the 668.513 F-1 visas approved in 2010, this amounts to a rather small group. Beside the strict visa restriction, which seem to hinder foreign students to remain in the U.S., a recent survey of former foreign students in the U.S. provides additional evidence. According to Enders and Kottmann (2013) 90% of surveyed alumni are working in their home country, conditional on having a job, and 75% of all alumni live in their respective home countries.

Even though there is evidence that most foreign students in the U.S. have to return to their respective home countries after expiration of their visa, some might actually stay in the U.S. However, the link between foreign education and domestic productivity might even remain intact for those students that are not returning. Ample evidence from the migration literature indicates that strong networks exist between the diaspora and the home country leading to positive feedbacks. Several studies report that the strong link between diaspora and home country leads to increased founding of businesses, trade and investments (cf. Rauch and Trindade, 2002; Rauch and Casella, 2003; Kugler and Rapoport, 2007; Aga et al., 2013). This contrasts the dreary view of the "brain drain" literature by the much more delectable term "brain circulation" suitably coined by Sexenian (2002).

To sum up, I argued that foreign students in the U.S. are predominantly enrolled in fields which are tied to productivity enhancing skills. Moreover, foreign students likely have to return to their respective home countries after their visas expire, i.e. usually after graduation. Due to the strict visa and immigration restrictions in the U.S. most students are unable to extend or switch their visa status. Finally, evidence from the migration literature suggests that strong links between the diaspora and the respective country of origin remain. These strong links lead to increased trade and investment between the sending and the receiving country.

3.7 Conclusion

National governments provide significant amounts of public resources to finance programs of student exchange. Respective institutions justify such expenses by referring to the positive impact foreign education has on economic growth and thus welfare. However, empirical evidence for the proposed positive effect between foreign education and domestic productivity is scarce (Engberg et al., 2014). This paper presented empirical evidence indicating that foreign education has indeed a positive effect on domestic total factor productivity growth. For instance, estimation results show that a 1 percent increase of the student flow to the U.S. increases TFP growth for an average country by about 0.025 percentage points in the short and 0.038 in the long run. Moreover, the positive effect of foreign education diminishes as countries become more developed indicating that foreign education is especially important for economically less developed countries.

The suggested underlying causal mechanism between foreign education and domestic productivity is the transmission of skills. Skills and technology are often found to be complementarities (Leiponen, 2005; Mohnen and Röller, 2005; Giorcelli, 2016) implying that a certain technology requires specific skills to unfold its full productive potential. Hence, the acquisition of new skills through foreign education can enable a country to employ available technology more productively than before (see also Acemoglu and Zilibotti, 2001). The causal mechanism for this conclusion is based on two reasonable assumptions about technology and skill-sets. First, technologies require specific skills to unfold their full productive potential, an assumption rooted in the literature on skill-biased-technological-change (see Acemoglu, 2015, for a recent discussion). Second, there are cross-country differences in skill-sets which can be explained by differing educational quality (e.g. Hanushek and Woessmann, 2012).

From a policy perspective, the results presented in this paper indicate that support of foreign education can be an effective extension to the existing set of development policies, especially for countries with an educational system of inferior quality. Providing financial means to students pursuing their studies in an industrialized country, preferably one with a superior educational system, will increase welfare for the whole society through higher productivity. However, it seems advisable that policy makers provide incentives for their internationally mobile students to return home after completion of their studies. This would give some guarantee that new skills are indeed transferred. Furthermore, students educated abroad could potentially multiply their impact on the domestic economy by spreading their newly acquired skills through working in the educational sector. Hence, it could also be beneficial if policy makers provide incentives for returning students to work in the domestic educational system. However, little empirical evidence on how foreignly educated educators affect the performance of the domestic educational system exists so far opening a fruitful avenue for further research.

4 The Scrubber Rip-Off. Regulation-Based Price Discrimination: Evidence from the Acid Rain Program[68]

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."

(Sherlock Holmes in: The Adventure of the Speckled Band)

4.1 Introduction

Environmental regulation internalizes externalities and provides incentives to develop and adopt abatement technologies with clear differences in performance across regulatory instruments.[69] With imperfect competition in the market for abatement technologies, e.g. due to patents, regulatory interventions in the polluting industry affect the elasticity of demand for these devises, implying instrument specific mark-ups.[70] Observable differences in the willingness-to-pay for abatement created by different types of environmental regulation have a typically unintended side effect. At least in principle, it allows providers of abatement to price discriminate.

U.S. regulation of sulfur dioxide (SO_2) emissions from coal-fired power plants provides an illustrative empirical example. Starting in the early 1970's different regulatory instruments were introduced, most restricting SO_2 emissions by some kind of command-and-control regulation. Title IV of the 1990 Clean Air Act Amendments (CAAA) introduced a novel program of emissions allowance trading and extended SO_2 regulation to a set of power plants that had previously been exempted from federal regulation.

To comply with SO_2 emission regulation coal-fired power plants have basically two economically feasible abatement strategies. First, they can switch their fuel from high to low sulfur coal. Second, they can install a flue-gas desulfurization [FGD] device, also called "scrubber", to remove SO_2 from the emission stream.[71] Considering the first compliance strategy, Busse and Keohane (2007) provide empirical evidence that railway operators delivering (low sulfur) coal to U.S. coal-fired power plants price discriminate on the basis of geographic location (power plants' distance to the Powder River Basin, the major mining area of low sulfur coal) and the

[68]Acknowledgements: This paper has been co-authored with Grischa Perino. We would like to thank participants at the Young Economist Conference 2016 in Lisbon, the annual conference of the European Association of Environmental and Resource Economists 2017 in Athens, and the Workshop on Environmental Regulation 2017 at ETH Zurich. We are especially thankful to Marc Luik.

[69]See Milliman and Prince (1989); Jung, Krutilla and Boyd (1996) and Requate and Unold (2003).

[70]Denicolò (1999); Fischer, Parry and Pizer (2003); David and Sinclair-Desgagne (2005); Requate (2005); Perino (2010) and David, Nimubona and Sinclair-Desgagne (2011).

[71]There is a small empirical literature investigating the interaction between SO_2 regulation and compliance of coal-fired power plants. Lange and Bellas (2005) provide evidence that both installation and operating costs of scrubbers decreased after the introduction of tradable permits for SO_2 emissions by Title IV of the 1990 CAAA. Keohane (2005) compares the price elasticity of installation decisions and finds that adoption of scrubbers is more sensitive to total scrubbing costs under permit trading than under command-and-control.

regulatory instrument (Title IV of the 1990 CAAA).[72] That is, the imperfect competitive market structure of railway operation allows operators to charge higher prices to those power plants that are regulated by the program. Hughes (2011) conducts a similar analysis for the transportation costs of ethanol used in reformulated gasoline and alternative transportation fuels. His results are based on non-attainment areas instead of SO_2 regulation as an instrument of environmental regulation. To sum up, the admittedly sparse empirical evidence suggests that providers of clean fuel charge mark-ups based on environmental regulation.

In the spirit of but contrary to Busse and Keohane (2007), we analyze price discrimination based on SO_2 regulation but focus on the market for flue gas desulfurization devices, the alternative suitable abatement technology for coal-fired power plants. We exploit two different sources of variation to identify differences in the willingness-to-pay for scrubbers. First, we test whether scrubber vendors charge mark-ups to power plants regulated by Title IV. Second, we test whether SO_2 emission stringency effects scrubber prices. To the best of our knowledge no empirical evidence on the link between price discrimination on abatement technology and environmental policies exists to date.[73] Abatement technologies deserve special attention relative to clean fuel access since mark-ups charged on abatement devices affect R&D incentives and diffusion rates. Consequences can be severe impacting the future set of abatement options and current abatement efficiency. The present paper fills this gap.

Our subsequent analysis reveals that scrubber prices are significantly higher for power plants that are regulated by Title IV of the 1990 CAAA. This finding persists even after controlling for technical and other observable characteristics and is robust to several sensibility checks. Mark-ups can change by millions of U.S. dollars per scrubber and hence are economically relevant. We find no evidence of price discrimination based on stringency of emission rate standards. However, we find that the costs of clean coal are positively related to scrubber prices indicating yet another channel how scrubber vendors identify power plants with higher willingness-to-pay.

The remainder of this paper is organized as follows. The next section gives relevant information on SO_2 regulation in the U.S. (section 4.2.1), elaborates on the market structure for scrubbers (section 4.2.2), and presents our hypotheses based on the interaction between regulation and the willingness-to-pay for an abatement technology (section 4.2.3). Section 4.3 describes our empirical strategy and section 4.4 our data. Subsequently, section 4.5 presents our results on the determinants of scrubber prices and elaborates on the possibility of price discrimination (section 4.5.2). Finally, section 4.6 concludes.

[72] Their results have been contested by Gerking and Hamilton (2008).

[73] Goeschl and Perino (2017) present a theoretical model where a monopolist owning a new abatement technology price-discriminates between signatories and non-signatories to an international environmental agreement due to the difference in price elasticities induced by the abatement commitment associated with signing the agreement.

4.2 Background and Hypotheses

4.2.1 Coal-Fired Power Plants and Regulation of Sulfur Dioxide Emissions in the U.S.

Coal-fired power plants have been identified as the "largest contributor to external costs" (Muller, Mendelsohn and Nordhaus, 2011, p. 1649) in the U.S., with damages ranging up to five times the value added by the industry. External costs of burning coal include the emission of CO_2, NO_x, SO_2, and Mercury. Emitting SO_2 is considered a major hazard, as it not only causes acid rain leading to acidification of soil and water sources but also contributes to the formation of $PM_{2.5}$. Several studies emphasize the negative effects of $PM_{2.5}$ on human health (see e.g. review in EPA 2004 or Chestnut and Mills 2005). Hence, reducing SO_2 emissions provides not only environmental but also considerable health benefits.

Over the last decades there has been a massive transformation in the U.S. electricity generating mix and a corresponding reduction in associated emissions. In the 1980's and 1990's of the last century, about 50 percent of electricity in the U.S. were produced by some 1,400 coal-fired power plants. Annual sulfur dioxide emissions in the 1970's and 1980's averaged around 17 million metric tons (EPA, 2016). Today coal's share in the U.S. electricity generating mix is down to about 30% and SO_2 emissions are reported to amount to roughly one million metric tons per year (EPA, 2016). This development is partly due to a shift in the electricity generating mix from coal to natural gas but also to significant investments in the abatement of hazardous pollutants by coal-fired power plants. For instance, Hewson and Graeter (2016) report that U.S. coal-fired power plants invested in total about USD 14.4 billion in SO_2 emission control systems before 2000 and a total of about USD 45.4 billion by the end of 2015. One major cause for this development is the continuous evolution of SO_2 regulation in the U.S. and a shift from command-and-control to an emission allowance trading system - and back.

With the advent of the Environmental Protection Agency (EPA), the federal government signed the Clean Air Act Amendments (CAAA) into law in December 1970.[74] The legislation was the first serious attempt to install federal emission standards. Before, coal-fired power plants were regulated by state or local emission rate standards that varied widely across and within states. According to Taylor, Rubin and Hounshell (2005) it required the much more stringent and uniform federal emission standards imposed by the 1971 and 1979 New Source Performance Standards (NSPS) to create a market for scrubbers. The 1971 NSPS regulates all new and substantially modified sources of SO_2 to employ best technology and the 1979 NSPS effec-

[74] Before 1970 the federal government undertook three major attempts to address air pollution: the Air Pollution Control Act of 1955, the Clean Air Act of 1963, and the Air Quality Control Act of 1967. In all three cases the Department of Health, Education, and Welfare was provided with funds for research and later very limited enforcement powers to take legal action against interstate polluters.

tively requires installation of a scrubber for all generating units installed after 1978.[75] During the 1980's the Reagan administration regarded environmental regulation predominantly as a concern of states, resulting in a state-level response to the emerging acid rain problem (Perino and Talavera, 2014). Thus, older units (built before 1971) remained mainly to be regulated by local or state authorities. Title IV of the 1990 CAAA (Public Law 101-549) marks the first serious federal attempt to address acid rain establishing an emission allowance trading program that also includes already existing units; at first only the dirtiest (Phase I, 1995-1999) and later nearly all emitting coal-fired power plants (Phase II, starting in 2000). All units that have to participate in Phase I are listed in a table (Table-A) attached to the legislation. It is therefore very easy to identify which unit has to participate and which does not. Similarly, the stringency of SO_2 regulation of a specific coal-fired power plant is in principle observable to outsiders as it depends on its spatial location and/or the time of in-service.[76] To conclude, serious SO_2 emission regulation of coal-fired power plants started with the 1971 NSPS and became comprehensive with Phase II of Title IV in 2000. During this period the evolution and variation of SO_2 regulation is most informative for our research question and hence we focus on scrubber installations between 1970 and 1999.

Price discriminating behavior requires at least two conditions. First, the market for scrubbers has to exhibit some degree of market power so that competition by price is imperfect (section 4.2.2). Second, scrubber vendors have to be able to observe differences in the willingness-to-pay of potential costumers (section 4.2.3). The next two subsections elaborate how the two conditions relate to the market for flue gas desulfurization devices.

4.2.2 The Market for Flue Gas Desulfurization Units

The market for flue gas desulfurization units in the U.S. was historically dominated by a small group of vendors. Between 1970 and 1999, 218 scrubber were installed by 25 manufacturers.[77] Over the course of this period several mergers occurred so that the actual amount of active and

[75] According to the 1971 NSPS, all emitting stationary sources built after 1971 can at most emit 1.2 lbs of SO_2 per million British thermal units of heat input. While this rate is technology neutral, the requirement of the 1979 NSPS to reduce potential SO_2 emissions per heat input by either 70 or 90% is not. Irrespective of the sulfur content of the coal the 1979 NSPS required that the sulfur content in the emission stream had to be reduced by 70 or 90%. This can only be achieved through scrubbing. The 70 or 90% mark depends on the sulfur content per unit of heat input of the burned coal.

[76] While the empirical analysis focuses on scrubbers it is worthwhile to notice that: a) plants usually consist of one or more boiler-generator units and b) one or more boiler-scrubber units. Environmental regulation is usually tied to boilers. This poses a potential problem as a scrubber can be connected to more than one boiler with possibly different levels of environmental regulation. For the data considered here, there are 8 cases in which more than one boiler is connected to a particular scrubber. In all cases respective boilers have the same level of regulation at the date the scrubber was installed.

[77] Due to missing data on installation costs the sample of scrubbers only includes 204 scrubber installations. Five scrubbers were inhouse constructions, i.e. designed and constructed by the power plant. Those installations are not considered in table 13 and in the computation of measures for market concentration.

independent vendors was much smaller at any given point in time (Lange and Bellas, 2005; Taylor, Rubin and Hounshell, 2005). While there was a relatively vibrant market for scrubbers in the late 1970s and early 1980s, likely due to the 1979 NSPS, market concentration increased steadily thereafter. Taylor, Rubin and Hounshell (2005) report that market exits, mergers, and acquisitions occurred in the late 1980s and early 1990s because the expected increase in demand for (retro-fitted) scrubbers did not materialize after the 1979 NSPS and/or the 1990 CAAA were passed.[78] For instance, the authors report that 16 scrubber vendors were active at the end of the 1970s but due to exits, mergers and acquisitions in the 1980s the number of independent scrubber vendors was considerably reduced. Table 13 reports measures for market concentration: the market share of the four leading vendors (CR4) and the Herfindahl-Hirschman-Index (HHI). The concentration measures presented are lower bounds because vendors are identified by name and hence ownership structures between different vendors are not taken into account. While the 1979 NSPS initially led to market entries by newcomers and probably to increased competition in the early 1980s, the 1990s were marked by a relatively concentrated market (cf. table 13). Moreover, Taylor, Rubin and Hounshell (2005) estimate that between 1976 and 1996 some 1,237 to 1,593 patents concerning FGD systems were filed. Of those 76% were attributable to firms. Given the small number of vendors one can expect a high concentration of patents, considerably constricting market entrance for newcomers. Consequently, the scrubber market that started to emerge in the mid 1980s is characterized by a rather oligopolistic market structure with active scrubber vendors exhibiting some degree of market power.

Table 13: Market concentration and average costs

	Concentration measures			Real scrubber prices (in 1,000 USD)		
	CR4	HHI	installations (distinct)	all units	Non-Table-A	Table-A
1970 - 1999	46.79	0.078	218 (25)	76,560.26	71,932.80	116,885.20
1970s	48.45	0.089	50 (16)	77,812.58	77,812.58	NA
1980s	54.00	0.101	97 (17)	84,243.40	84,243.40	NA
1990s	69.01	0.150	71 (12)	63,609.28	35,639.42	116,885.20

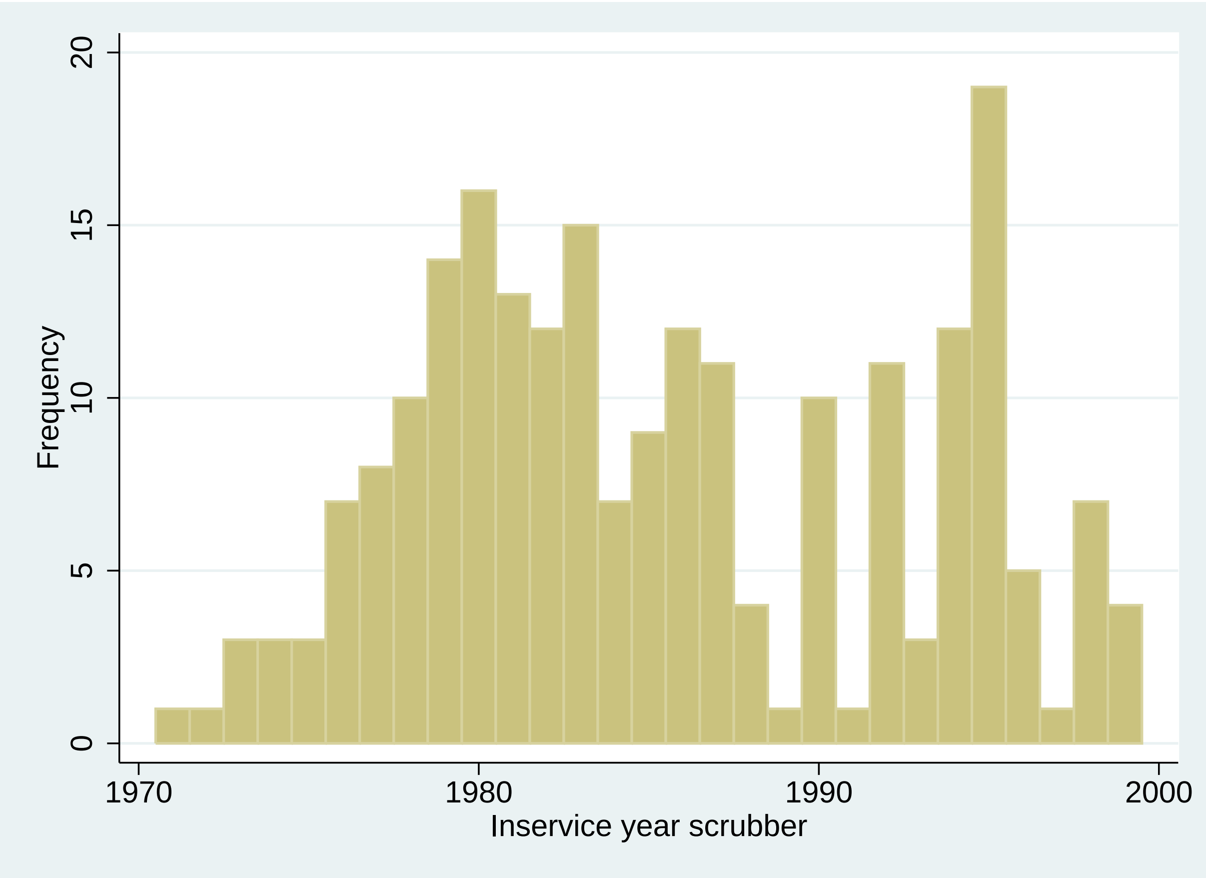
Notes: CR4 refers to the market share of the four biggest companies and HHI refers to the HerfindahlHirschman Index which ranges from 0 to 1. Inhouse installations are not listed here and were not considered in the computation of both measures. Table-A refers to units that had to participate in Phase I of the Acid Rain Program. Distinct installations are scrubber installations carried out by a different vendors (i.e. distinct by name). Installation costs are deflated using the Handy-Whitman index for public utility construction costs (electric). The number of scrubber price observations is smaller because prices are missing for some installations. However, inhouse installations are considered if price information exists.

Scrubbers are durable goods with a lifetime of several decades and installing them is an irreversible investment. Resale is prohibitively expensive. Borenstein (1985), Holmes (1989) and Armstrong and Vickers (2001) analyze third-degree price discrimination in oligopoly but without explicitly considering the durable goods problem. However, the theoretical literature on durable goods and monopolies identifies at least two factors that help to overcome the Coase conjecture (Coase, 1972) and which are relevant in the market for scrubbers. Both discrete

[78]One reason for the absence of a sizable market for scrubber is the continuous usage of generating units exceeding their anticipated retirement age.

(Bagnoli, Salant and Swierzbinski, 1989) and varying demand (Board, 2008) enable a durable goods monopolist to raise prices above marginal costs. Figure 4 shows the annual frequency of scrubber installations over time. With no more than 19 units installed in any given year, demand for scrubbers is clearly discrete and, with on average 7.5 scrubber installations per year and a standard deviation of 5.3, demand is also highly variable.

Figure 4: Histogram of scrubber installations over time



Mean scrubber installation cost, also known as the price of a scrubber (Lange and Bellas, 2005), depicted in table 13, declined steadily over the decades.[79] One likely cause for such decrease in prices are patent innovations, which according to Taylor, Rubin and Hounshell (2005) first resulted in increased efficiency and redundancy, and later in lower installation costs. Similarly, Lange and Bellas (2005) and Popp (2003) attribute reductions in scrubber prices over time to technological change. Finally, Taylor, Rubin and Hounshell (2005) note that the majority of cost reducing innovations occurred before the 1990s. Indeed, mean scrubber prices reported in table

[79]Information on the transaction between a power plant and a scrubber vendor is scarce. It is our understanding, similar to the related literature (e.g. Lange and Bellas, 2005), that scrubber installation costs comprise all necessary expenditures and that all components are provided by the scrubber vendor. Therefore, we interpret reported installation costs as the vendor's price of a scrubber.

13 drop significantly in the 1990s. Surprisingly, such reduction in prices cannot be extended to the sub-sample of Table-A units that appear to be much more expensive. This indicates that not all customers profited from cost reducing innovations.

4.2.3 Downstream Regulation, Technology Adoption, and Hypotheses

This section illustrates how environmental regulation affects a coal-fired power plant's willingness-to-pay [WTP]. Power plants only have an incentive to abate if there are costs associated with pollution either directly through a price on emissions or indirectly through penalties for non-compliance with emission standards. All SO_2 emitting stationary sources are subject to emission standards and the sub-sample listed in Table-A also participate in the cap-and-trade scheme from 1995 onward. Note that Table-A units still have to comply with emission standards despite the additional and overall more stringent tradable permit program. While there was historically little change in emission stringency for Table-A units, Title IV required participating plants to cut their SO_2 emissions by roughly 40% creating a huge incentive for investments into abatement.[80]

In order to reduce SO_2 emissions per unit of heat input, power plants have essentially two options: switching fuel to burn low-sulfur coal or installing a FGD unit. The former is costly (Lange and Bellas, 2007) because the coal with the lowest sulfur content per unit of heat is predominantly mined in the Powder River Basin (PRB) which is located in Wyoming and Montana and hence further away from most coal-fired power plants than the two other main mining areas in the U.S. (Illinois Basin and the Appalachia region) (Busse and Keohane, 2007). For most plants shipping costs are the main component of the costs of coal (Ellerman and Montero, 1998; Gerking and Hamilton, 2008). However, Ellerman and Montero (1998) show that shipping costs from Wyoming declined in the 1980s due to deregulation in railroads. Many coal-fired power plants in the Midwest switched (partially) to burning low sulfur coal in order to meet tightening SO_2 regulation. In contrast, access to low-sulfur coal from PRB remained prohibitively expensive for several (mostly eastern) coal-fired power plants throughout the 1990s. Note that reductions in output are not a suitable compliance strategy, since they do not directly affect the rate of emissions per unit of heat input. Moreover, due to the type of coal found in the U.S. and the very high costs of removing sulfur from the coal before it is burned, there is a natural lower bound for the emission rate that can be achieved. Therefore, for many mainly eastern coal-fired power plants scrubbing is the economically most viable solution.

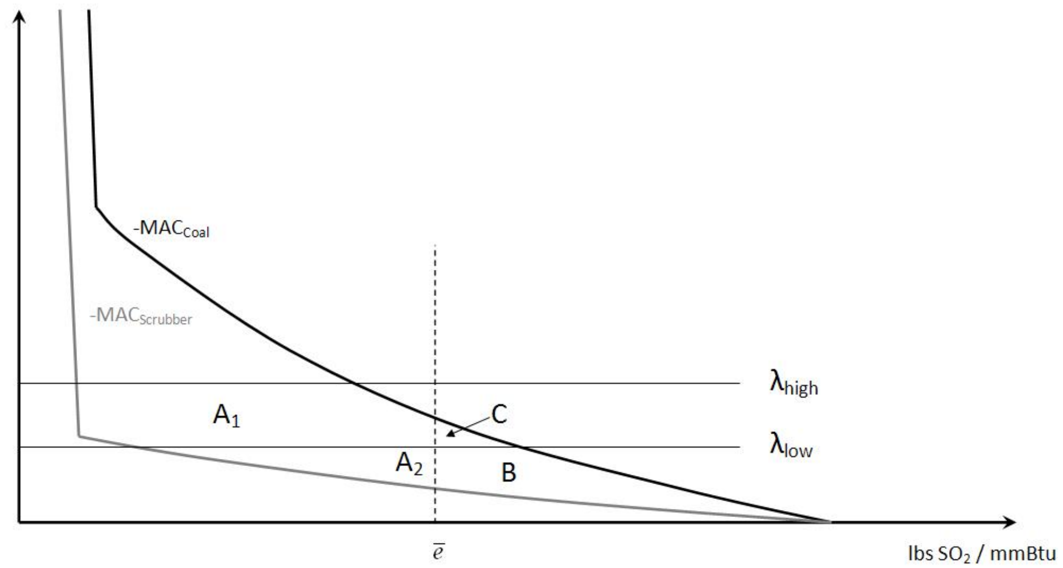
A scrubber removes up to 98 percent of SO_2 from the emission stream and hence allows for

[80] In 1985 Table-A units had an average emission rate of 4.24 lbs/mmBtu (Ellerman and Montero, 1998). In 1995 (1999) average emission standard for Table-A for which we have suitable data amounted to 4.78 (4.56) lbs/mmBtu. Hence, there was little change in emission stringency.

using coal of effectively any sulfur content – even at units facing very strict emission standards (EPA, 2003). Scrubbers can be retro-fitted but doing so is a major and irreversible investment. Scrubber prices can vary over time (likely due to technological progress), their timing of installation (retrofitted to vs. installed together with boiler), its technical specifications (e.g. removal efficiency, reliance, and size), and possibly mark-ups by scrubber vendors. For instance, the average real price for a retro-fitted (non-retro-fitted) scrubber is \$80.10 (74.71) million (in 1996 USD).[81]

It has been shown that abatement technologies reducing the emission-output ratio imply that marginal abatement costs with and without adoption intersect (Amir, Germain and van Steenberghe, 2008; Bauman, Lee and Seeley, 2008) and that this results in investment incentives to be decreasing in the stringency of environmental regulation once abatement moves beyond the point where the two marginal cost schedules intersect (Perino and Requate, 2012). Due to the high removal efficiency of scrubbers and the natural lower bound to what can be achieved by input substitution, this intersection of marginal abatement cost curves is practically irrelevant in this specific case since it would be located much below the most stringent emission standard implemented.[82] However, operating a scrubber is not for free and hence there are still strictly positive marginal costs of abatement. Figure 5 illustrates the savings in variable abatement

Figure 5: Willingness-to-pay for scrubbers per unit of heat input under emission standards ($B+C$) and tradable permits & emission standards for low (A_2+B+C) and high (A_1+A_2+B+C) permit price scenarios.



costs under the two policy regimes. Given the marginal abatement cost schedules $-MAC_{Coal}$

[81]Installation costs are deflated using the Handy-Whitman index for public utility construction costs (electric)

[82]This does not conflict with the empirical evidence in Bauman, Lee and Seeley (2008) since they look at total emission not at emission rate standards and a sample that excludes scrubbers.

and $-MAC_{Scrubber}$, an emission standard of $\bar{e}$ implies cost savings of $B+C$ per unit of heat input if a scrubber is installed.[83] Increasing the emission standard (i.e. moving $\bar{e}$ to the left) also increases the cost saving.

Hypothesis 1 *The willingness-to-pay to install a scrubber is increasing in the stringency of the emission standard.*

This effect will be larger for units not participating in permit trading.[84] For a Table-A unit the cost savings also depend on the equilibrium permit price. Two examples (λ_{low} and λ_{high}) are presented. If the permit price is relatively high, i.e. above $-MAC_{Coal}(\bar{e})$, then cost savings are A_1+A_2+B+C per unit of heat input. For a relatively low permit price (λ_{low}), cost savings reduce to A_2+B+C. Importantly, area C is included since compliance with the emission standard is still required. This implies that the incentives to install a scrubber are always (weakly) higher under tradable permits with emission rate standard than under an emission rate standard alone.[85] Hence, power plants facing tradable permits and an emission rate standard should *ceteris paribus* have a higher WTP for a scrubber than power plants facing emission rate standards alone. This increase in the WTP for scrubbers, induced by the permit trading program, is in line with none of the Table-A units having installed a scrubber before the 1990 CAAA but 23 doing so during Phase I. Note also that there was no perceptible change in average emission standards for Table-A units before and after the passing of the 1990 CAAA.

Hypothesis 2 *Participation in the permit trading program increases the willingness-to-pay for a scrubber.*

Third-degree price discrimination can be an issue in the U.S. market for scrubbers because the specific market structure and concentration of patents allow for market power. Heterogeneity in environmental regulation allows to identify differences in power plants' WTP. Provided that stringency and type of environmental regulation can easily be observed, scrubber vendors can use this information to charge group specific mark-ups.

Thus, there are reasons to conjecture that scrubber vendors could well be in a position to raise prices above marginal costs and to vary the size of the mark-up based on environmental regulation. However, in the end this is an empirical question and the key contribution of this paper is to test whether heterogeneity in regulation can serve to (ex-post) identify third-degree price discrimination.

[83] For a recent example of how to estimate marginal abatement costs (for NO_x) at coal-fired power stations see Fowlie (2010).

[84] It is zero if the permit price is above $-MAC_{Coal}(\bar{e})$.

[85] Adoption incentives are the same if the permit price is lower than $-MAC_{Scrubber}(\bar{e})$. The stricter the permit scheme (which, *ceteris paribus*, determines the equilibrium permit price) is compared to the emission standard, the larger is an increase in the willingness-to-pay for scrubbers when permits are introduced.

4.3 Empirical Strategy

In line with the related literature, we assume that scrubber installation costs are mainly driven by technical and technological factors (Srivastava and Jozewicz, 2001; EPA, 2003; Lange and Bellas, 2005; Taylor, Rubin and Hounshell, 2005). However, observed scrubber prices will not only include installation costs but any mark-up charged by scrubber vendors. We assume that type and stringency of environmental regulation create differences in WTP. Our empirical strategy exploits observable differences in regulation to identify the variation in scrubber prices that is due to third-degree price discrimination.

In the following we refer to technical factors T_k as determinants of installation costs (C_k). Such factors are scrubber capacity, efficiency, redundancy, technology, and the overall size of the power plant. Ideally, we would be able to control for all factors that determine installation costs and could estimate an equation equal to:

$$C_k = \alpha + \rho \cdot T_k + \lambda \cdot D_k + \varepsilon_k. \tag{5}$$

While we are able to control for the most important factors, we are unable to control for the actual engineering and construction effort. Most of this variation is likely to be correlated to the technical factors. If the relative importance of an unobserved parameter (including a general mark-up) does not change across units its influence on average scrubber prices will be captured by the constant α. However, in order to account for regional differences, e.g. costs of labor, we also include region dummies D_k following the classification of the Bureau of Economic Analysis.

In a world absent of any price discriminating behavior, signals identifying customers' WTP should not have any significant effect on prices (P_k) and thus $C_k = P_k$. As discussed in section 4.2 there are reasons to doubt that prices only reflect installation costs. Regulation affects a plant's WTP and if serving as a signal to scrubber vendors, helps us to identify mark-ups. Therefore, the extended version of equation 5 includes measures for environmental regulation:

$$P_k = \alpha + \rho \cdot T_k + \gamma \cdot TableA_k + \delta \cdot R_k + \tau \cdot O_k + \lambda \cdot D_k + \varepsilon_k, \tag{6}$$

where $TableA_k$ is a group identifier equal to one if a scrubber is installed at a Table-A unit and R_k is a vector containing measures of regulatory stringency. A statistically significant coefficient on the Table-A variable ($TableA_k$) would indicate that i) scrubbers installed at Table-A units are fundamentally different than those installed at non-Table-A units and/or ii) scrubber vendors price discriminate based on the higher WTP of Table-A units. The latter conclusion relies on the absence of unobservable drivers with heterogeneous effects on scrubber prices.[86] If there are unobservable characteristics that systematically influence scrubber prices our estimations suffers from omitted variable bias. This is a general problem of estimations aiming to discover price discrimination (Fortin, Lemieux and Firpo, 2011). Yet, based on the discussion in section 4.2, we are confident that our estimations indeed capture mark-ups due to downstream environmental regulation and not just spurious correlations. To alleviate concerns about omitted confounders, we test coefficient stability of the treatment effects at the end of section 4.5.2. Note that a similar argument applies to the interpretation of the coefficient of R_k.[87] Finally, we include a proxy for the costs of clean coal from the PRB O_k, the best alternative strategy to comply with emission standards (i.e. the power plant's opportunity costs). While those costs should not affect scrubber installation costs, they might affect a power plant's WTP for a scrubber. If scrubber vendors are able to identify plant specific costs of clean coal, they might also be able charge mark-ups for those plants that have high costs of clean coal and hence a higher WTP. Therefore, it is possible that O_k affects scrubber prices.

Price discriminating behavior is notoriously difficult to identify (Lott and Roberts, 1991). However, strategies to identify discriminating behavior can be found in the economic labor market literature. One often used approach is the so called Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973). The procedure divides the differential of group means into an "explained" and a residual part. The first part is explained by group differences in characteristics. That is, some of the difference between Table-A and non-Table-A scrubber prices can be allocated to differences in technical parameters and hence be explained. The remaining "unexplained" part measures the difference in estimated coefficients and is usually referred to as a measure for discrimination. That is, the price effect of e.g. increasing scrubber efficiency by one unit is different for Table-A than for non-Table-A units. This results in the major advantage of the decomposition approach, as group specific regressions allow for differing constants and parameter coefficients.[88] The decomposition can formally be represented by:

$$\begin{aligned} P^B - P^A &= \beta^* \cdot \left(T^B - T^A\right) + \\ &+ T^B \cdot (\beta^B - \beta^*) + T^A \cdot \left(\beta^* - \beta^A\right), \end{aligned} \tag{7}$$

[86]One possible heterogeneous effect could be differences in economic performance across regions. We include BEA region dummies to control for time invariant but region specific effects.

[87]However, as smaller measures of R_k indicate increased regulatory stringency, a significant negative coefficient would indicate price discrimination based on regulatory stringency. A complete description of included variables can be found in section 4.4.

[88]Note that in this strategy the identification of price discrimination also relies on the absence of important unobservables. Testing coefficient stability of the treatment effect with respect to the importance of unobservables is no panacea but can alleviate concerns regarding causality.

where A and B identify Table-A and non-Table-A units, respectively. Results of the decomposition depend on the non-discriminatory coefficient vector β^* which is not readily available. The related literature has introduced several suggestions. For instance, one could assume $\beta^* = \beta^B$ (cf. Oaxaca, 1973) or that β^* is some combination of weighted average of β^A and β^B (cf. Reimers, 1983; Cotton, 1988).[89] Based on theoretical work by Neumark (1988), Oaxaca and Ransom (1994) show that β^* can be estimated from a pooled regression over both groups (β^{pooled}). Jann (2008) notes that using the approach by Oaxaca and Ransom (1994) can inappropriately transfer some of the unexplained part of the differential into the explained component if the pooled regression does not include the group identifer (here *TableA*). As we do not want to assume any add hoc constraints on the structure of β^*, we follow Jann (2008) and estimate it from a pooled regression equal to equation (6).
However, we also present estimation results assuming that $\beta^* = \beta^B$ (i.e. negative discrimination against Table-A units) and $\beta^* = \beta^A$ (i.e. positive discrimination against non-Table-A units) in order to clarify the direction of third-degree price discrimination. While the earlier is sensible, the latter requires some modifications because not all non-Table-A units are potential candidates for positive third-degree price discrimination caused by the permit trading program. Only units built after the 1990 CAAA was passed can potentially be affected. To investigate whether there has been positive discrimination against non-Table-A units after 1990, we construct two new groups R and N, where group N includes all non-Table-A scrubbers going into service after 1990 while group R comprises the rest. Hence, by assuming $\beta^* = \beta^R$ we can check whether scrubber vendors charge unexplainable low prices to group N scrubbers, i.e. (positively) price discriminate against this group.[90]

4.4 Data

4.4.1 The Sample

The sample of 204 scrubbers includes all FGD units installed at U.S. coal-fired power plants, with a total capacity of at least 100MW, still operational in 1985 and installed between 1970 and 1999.[91] A sub-sample used to estimate the effect of regulation stringency includes 192 scrubber installation because not all units report their sulfur emission standard in pounds of sulfur dioxide per million British thermal units (mBtu) in heat input or a translatable measure. Most of the used data comes from the U.S. Energy Information Administration (EIA) form EIA-767 (**?**). It was partially extended by its successor from EIA-860.[92] Additional data sources will be named

[89] Assuming that $\beta^* = \beta^B$ can be understood as assuming that group A is discriminated, i.e. that the estimate for group B corresponds to the true value.
[90] Note that this assumes that the estimates of group R (including Table-A units) are the reference group.
[91] While there are 218 scrubber installations, for 14 of those no installation costs were reported.
[92] We used information from EIA-860 to verify and/or complete information in the EIA-767 survey. In a few cases coding mistakes had to be corrected. In others missing data was filled mostly with the help of data from other waves or the EIA-860. Additional information is provided in appendix C.1

as needed. The EIA-767 was an annual survey (discontinued in 2005) collecting information from all U.S. coal-fired power plants with a generator nameplate rating of 10MW or larger.[93] The survey must be completed partly by each plant with at leat 10MW generating capacity and fully by all plants with 100MW generating capacity or more. The first wave of the survey dates back to 1985 but the information in the survey is back-looking. That is, installations of scrubbers before 1985 are covered if they were still operational in 1985.

Table 14: Descriptive statistics

	Whole Sample			Non-Table-A			Table-A		
	count	mean	sd	count	mean	sd	count	mean	sd
ScrbPrice	204	76,560.26	74,889.76	183	71,932.8	67,676.97	21	116,885.2	115,512.5
InservYear	204	1985.054	6.883	183	1983.913	6.323	21	1995	1.095
Trains	204	2.961	1.805	183	3.120	1.821	21	1.571	0.811
RemovalUnit	204	2.145	0.736	183	2.039	0.691	21	3.066	0.405
GasExit	204	1,443,705	920,165.7	183	1,406,639	891,919.8	21	1,766,710	1,110,268
FGDwet	204	0.794	0.405	183	0.771	0.422	21	1	0
NameplatePlant	204	1150.946	802.998	183	1114.486	778.644	21	1468.667	953.635
Retrofit	204	0.343	0.476	183	0.268	0.444	21	1	0
Inhouse	204	0.015	0.121	183	0.016	0.127	21	0	0
OutlierLow	204	0.015	0.121	183	0.006	0.074	21	0.095	0.301
5yrHHI	204	0.135	0.055	183	0.129	0.053	21	0.182	0.043
$ProxyMAC_{Coal}$	204	39.029	27.608	183	37.529	28.548	21	52.092	10.972
Regulation	192	1.228	1.333	172	0.869	0.575	20	4.315	1.918

Notes: *ScrbPrice* are deflated using the Handy-Whitman index for public utility construction costs (electric) and are expressed in thousands of 1996 USD. *FGDwet*, *Retrofit*, *Inhouse*, and *OutliersLow* are dummy variables with [1=yes]. Reported are values for the whole sample, non-Table-A units only, and Table-A units only.

Scrubber prices (*ScrbPrice*) are deflated using the Handy-Whitman index for public utility construction costs (electric) and are expressed in thousands of 1996 USD. Mean (median) prices in the sample are 75.28 (56.98) million 1996 USD.

Table 14 depicts summary statistics for the whole sample as well as Table-A and Non-Table-A scrubbers, respectively. While Table-A units are on average significantly more expensive, the two groups also vary in many of the included parameters. This is not surprising as scrubbers are hardly a homogeneous good. Therefore, to make a reasonable comparison of scrubber prices, technical characteristics have to be accounted for. We use the listed information on plant size and scrubber characteristics to account for heterogeneity. The selection of characteristics is based on the usual set of variables found in the related literature (e.g Lange and Bellas, 2005; Taylor, Rubin and Hounshell, 2005; Poullikkas, 2015). The following section presents a detailed description of the included variables.

4.4.2 Description of Variables

The following variables contain information about technical characteristics affecting scrubber prices.

[93] Since 2006 similar information is collected through EIA forms 923 and 860.

NamePlate refers to the maximum generating capacity of the power plant. Power plants with higher nameplate ratings usually have several boiler and generating units. This means that the plant has more buildings making integration of scrubbers more costly.

Related to the size of the whole generating plant, *GasExitRate* is included to measure the scrubber's size. The 'flue gas exit rate' (in cubic feet/minute) describes how much flue gas can pass through the FGD unit. The more flue gas a scrubber can handle, the more expensive it should be.

Several studies report that retrofitting a scrubber to an existing generating unit is significantly more expansive than for new installation (EPA, 2003; Lange and Bellas, 2005; Taylor, Rubin and Hounshell, 2005, e.g.). According to EPA (2003) capital costs can increase up to 30%. The binary variable *Retrofit* takes the value [1] if the FGD unit went active more than two years after the boiler went into operation and is [0] otherwise. In our sample 70 units are retrofits with mean (median) costs of 80.1 (68.9) million USD. For the 134 new installations mean (median) costs are with 74.7 (54.4) million USD slightly lower.

InsrvYearFGD is a vector containing the year a scrubber went into operation. Accounting for process innovation is important, as there has been some remarkable progress over the period under consideration (Popp, 2003; Taylor, Rubin and Hounshell, 2005). Taylor, Rubin and Hounshell (2005) analyze patents related to SO_2 scrubbing and report that innovations translated first into improved efficiency and reliability. Later those improvements resulted in lower capital costs through elimination of now unnecessary components. Moreover, the authors note that the majority of those improvements occurred before the 1990 CAAA. Hence, the newer a scrubber is (i.e. the later it went into operation), the cheaper it is expected to be on average.

We include three variables to account for the technical specifications of a scrubber. First, the removal efficiency of the scrubbing unit. Following Lange and Bellas (2005) the removal efficiency of the scrubber is specified in terms of standard removal units. A standard removal unit removes approximately 63.2 percent of incoming sulfur. The variable *RemovalUnit* is calculated as follows

$$RemovalUnit = \ln \frac{1}{1-x},$$

where x is the removal efficiency of the scrubber. Note that high values of *RemovalUnit* translate into more efficient scrubber units and likely into higher scrubber prices. Second, the number of included scrubbing compartments. Scrubbing compartments, denoted *Trains*, are a measure of redundancy. Early scrubbers had usually more compartments as reliability was problematic. More trains should translate into higher costs. Finally, *FGDwet* is a binary variable indicating whether the scrubber unit operates under a wet [1] or dry [0] system design. While wet systems are more efficient and the most common ones, they are also more expensive (Taylor, Rubin and Hounshell, 2005; EPA, 2003).

To account for regional differences such as economic activity or labor costs, we include region dummies. Regional classification refers to the eight BEA regions: Far West, Rocky Mountain, Southwest, Plains, Great Lakes, Southeast, Mideast, and New England. Rocky Mountain has been chosen as reference category inter alia for the location of PRB.

The market for flue gas desulfurization units became increasingly concentrated over the period under consideration (cf. table 13). To capture this variation we constructed 5 year average Herfindahl-Hirschman-Indices dubbed *5yrHHI*. We expect scrubber prices to increase with market concentration.

Three variables are included that potentially influence a power plant's WTP for a scrubber. Two of them are measures of environmental regulation (*Regulation* and *TableA*). The third proxies the power plant's opportunity costs at the time of scrubber installation ($ProxyMAC_{Coal}$).

All units in the sample are subject to an emission rate standard denoted *Regulation*. The variable refers to the emission rate standard a unit faces when first entering into service or the first year data is available.[94] Emission regulation can vary across and within states and hence are usually plant specific. Included emission rate standards are measured in SO_2 lbs/mBtu of heat input. An increase in stringency should ceteris paribus increase the WTP for a scrubber. However, this effect should be especially pronounced for units not regulated by Title IV of the 1990 CAAA. The reported average emission standard for Table-A units is 4.78 (4.56) lbs/mBtu of heat input in 1995 (1999) and hence much higher then the de facto average emission standard of 2.5 lbs/mBtu implied by the emission trading program. It is therefore unlikely that emission standards for Table-A units are binding. For this reason we construct an interaction variable *RegNonTableA* that is 0 for all Table-A and equal to *Regulation* for all non-Table-A units. [95]

The dummy variable *TableA* tells whether or not a unit participates in Phase I of the permit trading program introduced by Title IV of the 1990 CAAA (value of 1 if it does). A unit has to participate if it is listed in the table (Table A) attached to the legislation. [96] In total 263 units (generators) connected to 247 boilers are listed. The list of regulated units is easily accessible

[94]For scrubbers installed before 1985 the first year data on regulation stringency is available is usually 1985, the first year of the survey.
[95]We also run regressions only including non-Table-A units to test the effect of *Regulation*.
[96]Some Non-Table A units opted into the permit scheme but none of them installed a scrubber in the relevant period.

and we assume that it is common knowledge to all scrubber vendors.[97]

The variable $ProxyMAC_{Coal}$ is a proxy measure for the cost of clean coal. The intention behind this variable is to have a measure for a power plant's (best) alternative to scrubbing. It is constructed by multiplying a power plant's distance to the PRB with the average costs of coal transportation by rail of the year the scrubber was installed. Hence, the measure tells us how much a power plant would have to pay for a ton of clean coal from the Powder River Basin in the year it installed a scrubber.[98] Distance is measured "as the crow flies" using google maps and data of average annual transportation costs of coal by rail comes from the EIA.[99] Starting point for all measured distances is Casper, Wyoming.

Finally, two binary variables are included to account for features in the data. *Inhouse* marks all scrubber installations that were not bought but home-made. According to the survey an inhouse installation means that the engineering personal of the power plant constructed the scrubbing unit. *Inhouse* installations in our sample are significantly cheaper, have low removal efficiency, and comparably small scrubbing compartments. While *Inhouse* installations are very cheap – between USD 1.5 and 2.2 million – some additional reported scrubber prices are unconvincingly low. To account for this possible measurement error, we mark all reported scrubber prices below USD 1 million by *OutliersLow*.

4.5 Results

This section presents results on the determinants of scrubber prices (section 4.5.1) and the investigation of price discrimination (section 4.5.2).

4.5.1 Scrubber Prices

The first step in our empirical investigation focuses on the determinants of scrubber prices and permit trading. Table 15 presents results from OLS regressions of reported scrubber prices using 204 observations. As several scrubbers can be installed at the same power plant, all regressions are clustered at the plant level.
Column (1) of table 15 presents results of participation in permit trading on scrubber prices and column (2) presents an extension including the dummy variables *Inhouse* and *OutliersLow*. In both specifications the coefficient of *TableA* is positive and significant providing evidence that TableA units are more expensive than their counterparts. Scrubbers can hardly be characterised as homogenous goods. To account for scrubber specific technical heterogeneity, the

[97]For a extensive review of the program the interested reader is referred to Ellerman et al. (2000).
[98]We do not include coal prices as they should be similar for each plant in a given year.
[99]The EIA Coal Transportation Rates to the Electric Power Sector can be downloaded here.

Table 15: Scrubber installation prices and permit tarding

	OLS I	OLS II	OLS III	OLS IV	OLS V	OLS VI	OLS VII
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TableA	44952.4*	53589.1**	33039.4*	37117.3**	35027.4*	35877.8**	
	(25328.3)	(24761.3)	(17023.6)	(18192.7)	(18629.2)	(18017.8)	
Inhouse		-71878.0***	12677.7	-11430.7	-10693.8	-16518.0	-18308.2
		(7635.6)	(19914.9)	(22475.0)	(20632.7)	(20077.5)	(19962.9)
OutliersLow		-109331.2***	-97260.0***	-103894.0***	-100069.0***	-98868.1***	-97534.4***
		(22032.0)	(28187.6)	(28070.9)	(28523.8)	(26718.7)	(25800.1)
InsrvYear			-1739.6**	-1884.6***	-2310.3***	-1955.4***	-1916.4***
			(705.9)	(675.2)	(695.0)	(677.4)	(664.3)
Trains			10591.7*	9997.1*	10011.3*	10122.5*	10263.3*
			(5573.7)	(5462.3)	(5570.2)	(5391.2)	(5445.7)
RemovalUnit			30213.9***	23271.6***	23079.9***	21460.4**	19913.6**
			(7248.1)	(8478.4)	(8488.8)	(8206.6)	(8138.7)
GasExitRate			0.0352***	0.0365***	0.0370***	0.0388***	0.0385***
			(0.00810)	(0.00859)	(0.00904)	(0.00880)	(0.00880)
WetFGD			-9414.1	-18067.2*	-14577.5	-15989.3	-17540.7*
			(9925.8)	(10590.1)	(10621.7)	(10350.0)	(10291.0)
NameplatePlant			8.862	9.463	9.685	7.992	7.304
			(6.836)	(6.146)	(6.096)	(6.305)	(6.122)
Retrofit			16197.0**	12148.3	11582.3	6922.5	5721.5
			(7813.3)	(7558.8)	(7790.4)	(8288.6)	(8111.5)
5yrHHI					117724.2	143267.4	113073.4
					(96066.9)	(92442.9)	(90003.6)
ProxyMAC$_{Coal}$						359.8**	348.7**
						(155.6)	(152.4)
TableA $\times$ 5yrHHI							232906.4**
							(99921.9)
Constant	71932.8***	73708.6***	3372333.1**	3692165.7***	4519400.5***	3813824.8***	3745294.1***
	(7493.2)	(7633.9)	(1401249.2)	(1343466.1)	(1373579.9)	(1338736.2)	(1314696.3)
BEA Region Dummies	No	No	No	Yes	Yes	Yes	Yes
Observations	204	204	204	204	204	204	204
R^2	0.033	0.076	0.570	0.595	0.599	0.607	0.611
Adjusted R^2	0.029	0.062	0.547	0.558	0.560	0.566	0.571

Notes: Regressions (1) to (5) include all observations in our data-set. Contrary, regression (6) only includes scrubbers that went into service between 1985 and 1999. Standard errors are corrected for hetroskedasticity, clustered at the plant level, and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

specification presented in column (3) includes a barrage of controls. Four important technical characteristics turn out to be significant in determining scrubber prices. First, the SO_2 removal efficiency has a positive impact. Second, the size of the scrubber, approximated by *GasExitRate*, also significantly increases scrubber prices. Third, redundancy, measured by the amount of scrubbing *Trains*, affects prices. Finally, the coefficient of *Retrofit* is positive and significant indicating that a retrofit is more expensive than a new installation. Importantly, the estimate on *TableA* remains positive and significant. As all Table-A scrubbers are retrofitted, failing to control for higher costs of retrofitting would cast doubt on the estimate of *TableA*. *InsrvYear* has a significant negative coefficient. The timing of scrubber installation can be interpreted as a time trend capturing the effect of technological innovation on scrubber prices. The size of the whole plant, approximated by *NameplatePlant*, has the expected positive sign but its coefficient is not statistically significant at conventional levels. Accounting for a retrofit reveals that retrofitting a FGD unit is significantly more expensive than a new installation. Withstanding all controls for scrubber characteristics, Table-A units still paid significantly higher prices of about USD 33m. Including regional dummies (BEA reagions) slightly increases the estimated effect of *TableA* and renders the effect of retrofitting a scrubber statistically insignificant. Assuming that scrubber vendors exploit power plants' WTP, we would expect that market concentration has a positive impact on scrubber prices. In column (5) we include *5yrHHI*. While the estimated coefficient is positive, it is not statistically significant. Thus, market concentration alone does not allow for higher prices. However, $ProxyMAC_{Coal}$ significantly raises scrubber prices (column 6). The higher the delivery cost of clean coal, i.e. the higher the opportunity costs, the higher the scrubber price. There is no ad-hoc reason why the opportunity costs to a scrubber should affect scrubber installation costs. However, they do affect a power plant's WTP for a scrubber. Due to the high cost of transporting coal the effect is mainly driven be a power plant's distance to PRB. Hence, finding a positive effect of $ProxyMAC_{Coal}$ suggests that scrubber vendors use the geographic location of a coal-fired power plant to approximate its WTP and charge mark-ups accordingly. Finally, it is likely that market power only affects Table-A units as their higher willingness to pay is easily observable. The last specification therefore interacts *TableA* with *5yrHHI*. For Table-A units increasing market power has a significant positive impact on scrubber prices. This indicates that scrubber vendors are aware of Table-A units' higher WTP and when possible exploit this information.

Table 16 presents results on a sub-sample of scrubber installations for which we have suitable data on SO_2 emission limits. The estimation strategy is similar to the one in table 15. However, this time we substitute *TableA* by *Regulation* to estimate the effect of regulation stringency on scrubber prices. Though, in theory, the WTP for a scrubber depends on the level of environmental regulation *Regulation* does not affect observed scrubber prices at a statistically significant level with and without controls (columns (1) and (2)). However, once we estimate within re-

Table 16: Scrubber installation prices and sulfur regulation

	OLS I	OLS II	OLS III	OLS IV	OLS V	OLS IV	OLS VII
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Regulation	8580.1	3435.3	6807.1*	1858.1	-731.2		-2344.8
	(6293.9)	(2990.2)	(3532.1)	(3804.5)	(3518.8)		(7851.1)
Inhouse		3603.4	-19478.0	-16767.3	-17215.4	-14984.1	-15187.5
		(17495.3)	(19244.6)	(18681.8)	(18572.3)	(18928.3)	(18874.7)
OutliersLow		-81070.9***	-82041.1***	-93362.1***	-96024.5***	-96627.0***	-35375.0**
		(26491.1)	(27128.8)	(32233.2)	(31509.5)	(30930.8)	(16400.8)
InsrvYear		-1667.6**	-1873.9***	-2186.8***	-2204.6***	-2261.8***	-2035.1***
		(806.5)	(683.9)	(704.0)	(691.8)	(736.2)	(761.0)
Trains		6169.1*	5450.3*	5916.5*	5985.3*	5869.1*	6511.2*
		(3570.0)	(3122.5)	(3429.8)	(3563.9)	(3498.7)	(3885.6)
RemovalUnit		21902.7***	16885.0**	14518.4**	12191.7*	13875.7*	10421.7
		(6281.0)	(6588.2)	(6635.4)	(6155.1)	(7215.5)	(7015.3)
GasExitRate		0.0412***	0.0412***	0.0404***	0.0401***	0.0405***	0.0365***
		(0.00761)	(0.00769)	(0.00773)	(0.00785)	(0.00801)	(0.00854)
WetFGD		-8062.3	-17162.0	-17351.0	-17966.2*	-16034.1	-14969.9
		(9484.0)	(10533.6)	(10485.2)	(10553.6)	(11285.5)	(11501.5)
NameplatePlant		8.437	9.732	9.001	8.044	8.729	8.574
		(7.598)	(6.750)	(6.648)	(6.433)	(6.788)	(6.826)
Retrofit		13633.9	7256.8	5551.3	4696.3	5922.0	4478.9
		(8953.3)	(8502.2)	(8706.7)	(8549.3)	(8783.7)	(8674.0)
ProxyMAC_{Coal}		368.1**	347.0**	348.3**	331.8**	347.8**	270.7*
		(162.9)	(163.1)	(165.0)	(162.9)	(167.6)	(162.9)
5yrHHI		112468.2	135684.7	137391.9	105457.6	142223.9	70919.8
		(97302.9)	(97141.0)	(92948.1)	(89407.8)	(97107.5)	(94628.9)
TableA				32742.9*		39188.9**	
				(18621.9)		(18664.0)	
TableA $\times$ 5yrHHI					272483.3***		
					(102526.2)		
RegNonTableA						-939.7	
						(7612.3)	
Constant	64331.7***	3220058.2**	3655568.9***	4284852.8***	4332013.8***	4435059.3***	4003400.3***
	(10174.5)	(1587917.4)	(1346352.2)	(1390054.4)	(1366847.9)	(1454317.4)	(1504987.2)
BEA Region Dummies	No	No	Yes	Yes	Yes	Yes	Yes
Observations	192	192	192	192	192	192	172
Adjusted R^2	0.022	0.513	0.527	0.529	0.537	0.529	0.573

Notes: Regressions (1) to (6) include all observations a suitable measure of emission standard is reported. Contrary, regression (7) further restricts the sample to all non-Table-A units. *RegNonTableA* is zero for all *TableA* units and takes the value of Regulation otherwise. *FGDwet*, *Retrofit*, *Inhouse*, and *OutliersLow* are dummy variables with [1=yes]. Standard errors are clustered at the plant level and reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

gion effects (column (3)) the coefficient of *Regulation* becomes significant but with a positive sign. Table-A units have significantly more lenient but not binding emission standards. Average scrubber prices are also significantly higher so that the coefficient might pick up this rather spurious correlation. Including *TableA* in column (4) documents that the significant positive coefficient of *Regulation* is indeed driven by this pattern. Once controlled for the effect of the emission trading program *Regulation* becomes insignificant again while the coefficient of *TableA* is significant and positive, also for this sub-sample. Similar to the large sample, market concentration does not affect scrubber prices but our proxy for marginal abatement costs of clean coal does. Column (5) shows that higher market concentration does significantly affect scrubber prices for Table-A units. To test whether stringency in emission standards does affect scrubber prices when binding, we substitute *RegNonTableA* for Regulation in column (6). We continue to control for *TableA*. While not statistically significant, the coefficient of *RegNonTableA* becomes negative. This goes at least into the direction of what our theory predicts. To make a more stringent test of our hypothesis, we exclude all Table-A units from the regression in column (7). Though increasing in magnitude and significance, the coefficient of *Regulation* remains statistically insignificant. Therefore, while scrubber vendors seemingly exploit Table-A units' higher WTP, they do not do so for units with tighter emission standards. One reason for this behavior could be that accessing a unit's emission standard involves some effort. In principle emission standards are available for the public but a variety of measures and regulatory authorities exit making it more difficult to extract the right standard. Moreover, the variable is continuous rather than binary and hence it might be impractical for the use by scrubber vendors for identification of a unit's WTP.

4.5.2 Price Discrimination

Observed installation costs can reasonable be modeled by the proposed specification. However, a significant price difference between Table-A and non-Table-A scrubber installations remains. In this section we pursue its investigation. The Oaxaca-Blinder decomposition approach allows to explain the difference in means of an outcome variable between two groups. The gap in prices between the two groups is decomposed into the part that is due to differences in characteristics and the part that is due differences in coefficients.

Table 17 depicts the results of our decomposition. The first finding is that the explained part of the decomposed differential is statistically insignificant (column (1)). That is, the difference between the two groups is not due to differences in characteristics. Second, the unexplained part of the differential is statistically significant. That is, the remaining gap of about USD 36 million is due to differences in coefficients or unobserved covariates.

Table 17: Decomposition of scrubber prices

	(1) Pooled ($\beta^* = \beta^{pooled}$)	(2) Negative ($\beta^* = \beta^B$)	(3) Positive ($\beta^* = \beta^R$)
Price difference	-44,952.42*	-44,952.41*	54,081.12***
Characteristics	-9,074.58	-7,789.56	43,048.97**
Coefficients	-35,877.85**	-37,162.86*	11,032.15

Notes: The table reports differences between the groups Table-A and Non-Table-A in columns (1) and (2). Column (3) is different, as now differences between non-Table-A installed after 1990 and all other scrubber installations are reported. The estimated differences are decomposed using the Oaxaca-Blinder procedure and have been carried out using Jann's (2008) Stata command. Regressions include the covariates included in specification VI of tables 15 or 16. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Columns (2) and (3) are dedicated to determine the direction of price discrimination. While column (2) reports on the difference between non-Table-A and Table-A installations, column (3) reports on the difference between non-Table-A scrubbers installed after 1990 and all other scrubber installations. In column (2) we assume that there is only (negative) discrimination against Table-A units. Indeed, the results are very similar to the specification assuming no ad-hoc direction of discriminating behavior (cf. column (1)). Contrary, in column (3) we assume that there is only (positive) discrimination against non-Table-A scrubbers going into service after 1990. The decomposition shows that non-Table-A units that went into service after 1990 are significantly cheaper than their counterparts (cf. also table 13). However, the difference can be explained by differences in characteristics indicating that there is no discriminating behavior against this group.

Finally, we address the impact of possibly omitted variables on our findings. As mentioned above the estimated effects might be due to third-degree price discrimination or to some omitted variable(s). In economics it is nearly impossible to be sure that a treatment effect is not driven by an unobserved confounder (cf. Leamer, 2010). To test how sensible the reported degree of price discrimination is to unobserved confounders, we employ the method proposed in Altonji, Elder and Taber (2005) and further developed by Emily Oster (2017). The idea of assessing the importance of unobservables on coefficient stability (usually the coefficient on the treatment variable) bases on the assumption that the selection of observables and unobservables is proportional. In order to test coefficient stability in this setting one has to make two assumptions. First, one has to choose the degree of relative selection (δ). Second, one has to decide how much of the variance of the dependent variable can be explained by the variance of the fully specified model. That is, one has to choose a maximal R^2. There is a clear suggestion in Altonji, Elder and Taber (2005) and Oster (2017) that δ should be bounded between [0,1]. For instance, choosing δ to be bounded by 1 is similar to assuming that observables are at least as important as unobservables. Oster (2017) provides empirical evidence in favor of the bounding condition. Choosing an appropriate R_{max} is less obvious. Based on her empirical exercise Oster (2017) suggests that researchers should choose R_{max} as a function of the R^2 derived in the researcher's model with controls ($\tilde{R}$). The suggested cutoff is $R_{max} = 1.3\tilde{R}$ chosen because most non-causal effects from randomized data do not survive this cutoff. The true value of the treatment coefficient should then lie within the range of δ [0,1]. Assuming $\delta = 1$ ($\delta = 0$) and $R_{max} = 1.3\tilde{R}$ yields a coefficient on *TableA* of 28,971.38 (35,877.85).[100] Hence, given the assumed upper bounds on δ and R_{max}, the true coefficient of *TableA* should lie within the range [28,971.38, 35,877.85].[101] Though the treatment effect could arguably be nearly 20% below the OLS estimate, the coefficient on TableA is still positive and the range does not include 0. This alone does not provide proof of a causal relationship but provides additional evidence. In appendix C.2 we also test coefficient sensibility of $ProxyMAC_{Coal}$ (specification (6) in table 15) and *Regulation* (specification (7) in table 16). Coefficients of both variables are robust in sign (table C.1). Given the evidence collected, we conclude that scrubber vendors identify differences in power plants' WTP through plant specific environmental regulation (participation in Title IV of 1990 CAAA) and geographic location. Moreover, they use this information to charge group specific mark-ups.

4.6 Conclusion

Using regulation of sulfur dioxide emissions from U.S. coal-fired power plants as an example, we show that market power in the upstream eco-industry interacts with plant-specific environmental regulation. The imperfectly competitive upstream eco-industry charges different mark-ups on the primary abatement technology (scrubbing) depending on the type of regulatory instrument a unit is subject to (tradable permits vs. emission rate standards). In line with our theoretical predictions scrubber vendors are able to price discriminate by identifying differences in plant's willingness-to-pay through differences in regulation. Two cases are proposed. First, under command-and-control regulation coal-fired power plants willingness-to-pay for a scrubber increases with regulatory stringency. We find sparse empirical evidence supporting third-degree price discrimination based on regulatory stringency. Second, comparing command-and-control regulation with permit trading power plants regulated under the latter policy regime are willing to pay more for a scrubber than the former ones. Indeed, power plants participating in the SO_2 permit trading program face substantially higher scrubber prices than otherwise equivalent plants only subject to emission standards. We show that the price difference is not due to differences in technological and technical parameters that are typically used in the related literature. This constitutes an additional channel by which instrument choice in environmental regulation affects abatement costs. Mark-ups on abatement technologies can influence diffusion and research incentives of those technologies resulting in repercussions on dynamic efficiency of environmental regulation. Moreover, we find that a coal fired power plant's geographic loca-

[100] Note that OLS implicitly assumes $\delta = 0$ as there are, in theory, no omitted variables.
[101] Further tests on coefficient stability and information about the applied method are provided in appendix C.2.

tion affects scrubber prices. The effect depends on the distance to PRB, the major mining area for low sulfur coal. Low sulfur coal poses the best alternative to scrubbing and hence higher transportation costs of clean coal from PRB increase a power plant's WTP for a scrubber.

So far price discrimination by an imperfectly competitive eco-industry has been ignored both by the theoretical and empirical literature studying the link between downstream regulation and diffusion of abatement technologies. The evidence presented suggests that price discrimination is empirically relevant and requires further theoretical and empirical research to better understand how market power and price discrimination in an upstream industry should be reflected in the design of environmental regulation.

5 Ad Ultimum

"It has been a case for intellectual deduction, but when this original intellectual deduction is confirmed point by point by quite a number of independent incidents, then the subjective becomes objective and we can say confidently that we have reached our goal."

(Sherlock Holmes in: The Adventure of the Sussex Vampire)

Applied empirical research plays an important role in the overall process of scientific progress. Scientific progress starts with the curious mind thinking about individual perceptions of the world in new and unfamiliar ways. However, the resulting new theory is initially nothing more than a subjective interpretation of individual observations. To successfully transfer the subjective to the set of the objective (i.e."truth"), initial interpretations have to be confirmed by further independent observations. Generating those independent studies confirming or rejecting individual ideas is an important task of empirical research.[102] It is this peculiar role in the formation of "truth" from which the empirical researcher receives a position of power and influence.

With power there comes responsibility and writing this dissertation I learned two important lessons.

Firstly, when testing a proposed theory (or hypothesis), the empirical researcher, especially in economics, should keep in mind that a theory only holds until there is sufficient conflicting evidence. Thus, theory is provisional. However, one should also be very careful before dismissing a whole theory. A theory usually does not depend only on one single statement and therefore cannot be easily rejected as a whole (e.g. Quine's holism in Quine, 1951). Quine (1951) cleverly stated: "[...] total science is like a field of force whose boundary conditions are experience [...] A conflict with experience on the periphery occasions readjustments in the interior of the field." (found in: Heckman and Vytlacil, 2007, p. 4782). While experience (empirical evidence) cannot contradict whole theories, it can, according to the Duhem-Quine thesis, contradict parts inherent to statements of the theory. The result of how to subsequently adjust the theory is, however, not clear. A priori it is not obvious which statement should be rejected and which should not. Quine argues that this depends on convention and/or opinion.

Secondly, even though the empirical researcher works in the business of forming objectivity, she herself can never be objective. In his seminal paper Leamer (1983) gives a great example: "Economists have inherited from the physical sciences the myth that scientific inference is objective, and free of personal prejudice. This is utter nonsense. All knowledge is human belief; more accurately, human opinion" (p. 36). Later he continues: "[...] To emphasize this hierarchy of statements, I display them in order: truths; facts; opinions; conventions. Note that I have added to the top of the order, the category truths. This will appeal to those of you who feel

[102]This does not imply that empirical researchers cannot or should not participate in the business of curious thinking.

compelled to believe in such things. At the bottom are conventions" (p. 37). Consequently, no single study can be enough to claim objectivity.

The three pieces of empirical research included in this dissertation all address different hypotheses related to different subfields in economics. It has been my personal ambition not only to establish connections between events and outcomes, but also to provide evidence for causal relationships. However, recognizing my responsibility as an empirical researcher, I want to put my results into perspective. In order to do so I provide a judgement on the internal as well as external validity of the results in each study. It is my intention that this discussion helps the reader to better understand the significance and limitation of each contribution.

The degree of internal as well as external validity naturally varies between each study. In section 2 my coauthors and I used a flooding event as a quasi-experiment to identify the causal relationship between personal experience of a diaster and precautionary behavior afterwards. While we are able to attribute the observed change in personal saving behavior to the flooding event by clearly identifying affected and non-affected individuals, the finding that affected individuals reduce precautionary saving seems implausible. Moreover, as we analyze the effect of experiencing a flooding in Germany, the result might not be transferable to a rarely developed country such as Bangladesh. This might sound as if our study is weak on internal and external validity. Yet it is quite the contrary. A profound analysis of the flooding event reveals that the behavioral change is not directly attributable to the flooding event, but to an inflow of unexpectedly high financial aid. This suggests that we actually identify the causal relationship between individual precautionary behavior and the implicit insurance through a third party. As moral hazard behavior is a well established and rather universal human trait, this explanation increases both internal and external validity. Our study therefore shows how important careful studying of causal mechanisms can be (see also Heckman, 1992; Deaton, 2009).
Compared to the study in section 2, my paper on foreign education and domestic productivity (section 3) shows a lower degree of internal validity. Though I am exploiting the time dimension of the data to establish that the cause comes before the effect, the limitations of the data make rigorous causal inference challenging. For instance, the data do not reveal whether foreign students actually return to and work in their country after studying in the U.S. For my argument, it is important that they indeed return. Only a positive coefficient might not be enough as its does not exclude alternative causal mechanisms. However, through a combination of regrouping country data and by including additional information on staying-rates of foreign students in the U.S., I am able to increase the degree of internal validity. Contrary to internal validity, the degree of external validity is rather high. The analysis uses aggregate data at the national level including most of the world's nations. However, whether the results are also generalizable to different periods in time and countries, not included in the study, remains to be seen.

Last but not least, in "The Scrubber Rip-Off" my coauthor and I combine the advantages of a structural model with those of a quasi-experiment. Although this strengthens the internal and external validity of our study, it still leaves us with a complex environment of intertwined causal mechanisms that makes the dismissal of alternatives (i.e. confounders) challenging. Nonetheless, we are able to control for many important confounders, greatly reducing the possibility of ignoring other probable causal mechanisms. However, doing so considerably strains the possibilities of our data. It is in this study where it becomes most obvious how qualitative and quantitative conditions of data can prevent a high degree of internal validity.

To conclude, the importance of empirical research in the overall process of scientific progress requires the reader of such work to adopt a critical perspective when judging a study's degree of internal as well as external validity. In the three included papers of empirical research, such a critical perspective is adopted and strategies to strengthen internal validity are developed and applied for each respective study. Independent of a study's degree of internal validity or a reader's judgement of it, the reader should recall that it requires more than one piece of evidence to create objectivity. However, this does not mean that individual research should be disregarded, as each result of genuine empirical research resembles a subtle push towards objectivity and therefore has its own right to exist. Ultimately, I hope that some insight from each study in this dissertation enters into the reader's subjective truth.

“Education never ends, Watson. It is a series of lessons with the greatest for the last.”

(Sherlock Holmes in: The Adventure of the Red Circle)

References

Abramovitz, Moses. 1956. "Resource and output trends in the United States since 1870." *The American Economic Review*, 46(2): 5–23.

Acemoglu, Daron. 2015. "Localised and Biased Technologies: Atkinson and Stiglitz's New View, Induced Innovations, and Directed Technological Change." *Economic Journal*, 125(583): 443–463.

Acemoglu, Daron, and Fabrizio Zilibotti. 2001. "Productivity Differences." *Quarterly Journal of Economics*, 116(2): 563–606.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2014. "Democracy does cause growth." *NBER Working Paper Series*, 20004.

Acs, Zoltan. 2006. "How Is Entrepreneurship Good for Economic Growth?" *Innovations: Technology, Governance, Globalization*, 1(1): 97–107.

Aga, Gemechu Ayana, Christian Eigen-Zucchi, Sonia Plaza, and Ani Rudra Silwal. 2013. "Migration and Development Brief."

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Aghion, Philippe, Diego Comin, Peter Howitt, and Isabel Tecu. 2016. "When Does Domestic Savings Matter for Economic Growth?" *IMF Economic Review*, 64(3): 381–407.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources*, XL(4): 791–821.

Amir, Rabah, Marc Germain, and Vincent van Steenberghe. 2008. "On the impact of innovation on the marginal abatement cost curve." *Journal of Public Economic Theory*, 10(6): 985–1010.

Anbarci, Nejat, Monica Escaleras, and Charles A. Register. 2005. "Earthquake fatalities: The interaction of nature and political economy." *Journal of Public Economics*, 89(9): 1907–1933.

Anderson, Terence, David Schum, and William Twining. 2005. *Analysis of Evidence.* Cambridge:Cambridge University Press.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3–30.

Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ:Princeton Univ. Press.

Antwi-Boasiako, Benjamin Addai. 2014. "Why Do Few Homeowners Insure Against Natural Catastrophe Losses?" *Review of Economics*, 65(3): 217–240.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277.

Armstrong, Mark, and John Vickers. 2001. "Competitive Price Discrimination." *RAND Journal of Economics*, 32: 579–605.

Atkinson, Anthony B., and Joseph E. Stiglitz. 1969. "A new view of technological change." *Economic Journal*, 79(315): 573–578.

Autor, David H., Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Econmics*, 113(4): 1169–1213.

Bagnoli, Mark, Stephen W. Salant, and Joseph E. Swierzbinski. 1989. "Durable-Goods Monopoly with Discrete Demand." *Journal of Political Economy*, 97(6): 1459–1478.

Baier, Scott L., Gerald P. Dwyer, and Robert Tamura. 2006. "How Important are Capital and Total Factor Productivity for Economic Growth?" *Economic Inquiry*, 44(1): 23–49.

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, 106(2): 407–443.

Barro, Robert J., and Jong Wha Lee. 2013. "A new data set of educational attainment in the world, 1950–2010." *Journal of Development Economics*, 104: 184–198.

Basu, Susanto, and David N. Weil. 1998. "Appropriate Technology and Growth." *Quarterly Journal of Economics*, 113(4): 1025–1054.

Bauman, Yoram, Myunghun Lee, and Karl Seeley. 2008. "Does technological innovation really reduce marginal abatement costs? Some theory, algebraic evidence, and policy implications." *Environmental and Resource Economics*, 40(4): 507–527.

Baumol, William J., Robert E. Litan, and Carl J. Schramm. 2007. *Good capitalism, bad capitalism, and the economics of growth and prosperity.* New Haven:Yale University Press.

Bechtel, Michael M., and Jens Hainmueller. 2011. “How Lasting Is Voter Gratitude? An Analysis of the Short- and Long-Term Electoral Returns to Beneficial Policy.” *American Journal of Political Science*, 55(4): 852–868.

Benhabib, Jess, and Mark M. Spiegel. 1994. “The role of human capital in economic development evidence from aggregate cross-country data.” *Journal of Monetary Economics*, 34(2): 143–173.

Berlemann, Michael. 2016. “Does hurricane risk affect individual well-being? Empirical evidence on the indirect effects of natural disasters.” *Ecological Economics*, 124: 99–113.

Berlemann, Michael, and Gerit Vogt. 2008. “Kurzfristige Wachstumseffekte von Naturkatastrophen: Eine empirische Analyse der Flutkatastrophe vom August 2002 in Sachsen.” *Zeitschrift für Umweltpolitik und Umweltrecht*, 31(2): 209–232.

Berlemann, Michael, and Jan-Erik Wesselhöft. 2014. “Estimating Aggregate Capital Stocks Using the Perpetual Inventory Method: A survey of Previous Implementations and New Empirical Evidence for 103 Countries.” *Review of Economics*, 65(1): 1–34.

Bils, Mark, and Peter J. Klenow. 2000. “Does Schooling Cause Growth?” *The American Economic Review*, 90(5): 1160–1183.

Blinder, Alan S. 1973. “Wage Discrimination: Reduced Form and Structural Estimates.” *The Journal of Human Resources*, 8(4): 436.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. “Does Management Matter? Evidence from India.” *Quarterly Journal of Economics*, 128(1): 1–51.

Bluedorn, John C. 2005. “Hurricanes: Intertemporal Trade and Capital Shocks.”

Board, Simon. 2008. “Durable-Goods Monopoly with Varying Demand.” *Review of Economic Studies*, 75: 391–413.

Borenstein, Severin. 1985. “Price discrimination in free-entry markets.” *RAND Journal of Economics*, 16(3): 380–397.

Borensztein, Eduardo, Jose de Gregorio, and Jong-Wha Lee. 1998. “How does foreign direct investment affect economic growth?” *Journal of International Economics*, 45(1): 115–135.

Borjas, George. 2005. “The Labor-Market Impact of High-Skill Immigration.” *The American Economic Review*, 95(2): 56–60.

Borjas, George. 2009. “Immigration in High-Skill Labor Markets: The Impact of Foreign Students on the Earnings of Doctorates.” In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment.* , ed. Richard B. Freeman and Daniel Goroff, 131–161. University of Chicago Press.

Borjas, George, and Kirk B. Doran. 2015. “How High-Skill Immigration Affects Science: Evidence from the Collapse of the USSR.” *Innovation Policy and the Economy*, 15(1): 1–25.

Botzen, Wouter, Jeroen Aerts, and Jeroen van den Bergh. 2009. “Willingness of homeowners to mitigate climate risk through insurance.” *Ecological Economics*, 68(8): 2265–2277.

Browne, Mark J., and Robert E. Hoyt. 2000. “The Demand for Flood Insurance: Empirical Evidence.” *Journal of Risk and Uncertainty*, 20(3): 291–306.

Brunette, Marielle, Laure Cabantous, Stéphane Couture, and Anne Stenger. 2013. “The impact of governmental assistance on insurance demand under ambiguity: A theoretical model and an experimental test.” *Theory and Decision*, 75(2): 153–174.

Bruno, Giovani S. F. 2005*a*. “Estimation and inference in dynamic unbalanced panel-data models with a small number of individuals.” *Stata Journal*, 5(4): 473–500.

Bruno, Giovanni S. F. 2005*b*. “Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models.” *Economics Letters*, 87(3): 361–366.

Buchanan, James M. 1975. “The Samaritan’s Dilemma.” In *Altruism, Morality, and Economic Theory*. 71–86. Russell Sage Foundation.

Busse, Meghan R., and Nathaniel O. Keohane. 2007. “Market effects of environmental regulation: Coal, railroads, and the 1990 Clean Air Act.” *RAND Journal of Economics*, 38(4): 1159–1179.

Callen, Michael. 2015. “Catastrophes and time preference: Evidence from the Indian Ocean Earthquake.” *Journal of Economic Behavior & Organization*, 118: 199–214.

Cameron, Lisa, and Manisha Shah. 2015. “Risk-Taking Behavior in the Wake of Natural Disasters.” *Journal of Human Resources*, 50(2): 484–515.

Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and quasi-experimental designs for research.* Boston Mass. u.a.:Houghton Mifflin.

Carson, David. 2009. “The Abduction of Sherlock Holmes.” *International Journal of Police Science & Management*, 11(2): 193–202.

Caselli, Francesco. 2010. “Growth Accounting.” In *Economic Growth.* , ed. Steven N. Durlauf and Lawrence E. Blume, 91–96. London:Palgrave Macmillan UK.

Caselli, Francesco, and Wilbur John Coleman. 2006. "The World Technology Frontier." *The American Economic Review*, 96(3): 499–522.

Cassar, Alessandra, Andrew Healy, and Carl von Kessler. 2017. "Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand." *World Development*, 94: 90–105.

Chellaraj, Gnanaraj, Keith E. Maskus, and Aaditya Mattoo. 2008. "The Contribution of International Graduate Students to US Innovation." *Review of International Economics*, 16(3): 444–462.

Coase, Ronald H. 1972. "Durability and Monopoly." *Journal of Law and Economics*, 15: 143–149.

Coate, Stephen. 1995. "Altruism, the Samaritan's Dilemma, and Government Transfer Policy." *The American Economic Review*, 85(1): 46–57.

Coe, David T., and Elhanan Helpman. 1995. "International R&D spillovers." *European Economic Review*, 39(5): 859–887.

Cotton, Jeremiah. 1988. "On the Decomposition of Wage Differentials." *Review of Economics and Statistics*, 70(2): 236.

David, Maia, Alain-Desire Nimubona, and Bernard Sinclair-Desgagne. 2011. "Emission taxes and the market for abatement goods and services." *Resource and Energy Economics*, 33(1): 179–191.

David, Maia, and Bernard Sinclair-Desgagne. 2005. "Environmental Regulation and the Eco-Industry." *Journal of Regulatory Economics*, 28(2): 141–155.

Deaton, Angus. 2009. "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development." *NBER Working Paper*, 14690.

Denicolò, Vincenzo. 1999. "Pollution-reducing innovations under taxes and permits." *Oxford Economic Papers*, 51: 184–199.

Deryugina, Tatyana, and Barrett Kirwan. 2016. "Does The Samaritan's Dilemma Matter? Evidence From U.S. Agriculture." *NBER Working Paper*, , (22845).

Deryugina, Tatyana, Laura Kawano, and Steven Levitt. 2017. "The Economic Impact of Hurricane Katrina on its Victims: Evidence from Individual Tax Returns." *American Economic Journal: Applied Economics (Forthcoming)*.

Dobes, Leo, Frank Jotzo, and David I. Stern. 2014. "The Economics of Global Climate Change: A Historical Literature Review: 65 / 3." *Review of Economics*, 281–320.

Dooley, Michael P., David Folkerts Landau, and Peter M. Garber. 2004. "The US Current Account Deficit and Economic Development: Collateral for a Total Return Swap." *NBER Working Paper*, 10727.

Drandakis, E. M., and E. S. Phelps. 1966. "A Model of Induced Invention, Growth and Distribution." *Economic Journal*, 76(304): 823–840.

Dreher, Axel, and Panu Poutvaara. 2011. "Foreign Students and Migration to the United States." *World Development*, 39(8): 1294–1307.

Easterly, W. 2001. "What Have we Learned From a Decade of Empirical Research on Growth? It's Not Factor Accumulation: Stylized Facts and Growth Models." *The World Bank Economic Review*, 15(2): 177–219.

Eckel, Catherine C., Mahmoud A. El-Gamal, and Rick K. Wilson. 2009. "Risk loving after the storm: A Bayesian-Network study of Hurricane Katrina evacuees." *Journal of Economic Behavior & Organization*, 69(2): 110–124.

Ellerman, A. Dennis, Paul L. Joskow, Richard Schmalensee, Juan-Pablo Montero, and Elizabeth M. Bailey. 2000. *Markets for Clean Air: The U.S. Acid Rain Program.* Cambridge:Cambridge University Press.

Ellerman, A. Denny, and Juan-Pablo Montero. 1998. "The Declining Trend in Sulfur Dioxide Emissions: Implications for Allowance Prices." *Journal of Environmental Economics and Management*, 36: 26–45.

Enders, Jürgen, and Andrea Kottmann. 2013. "The international fellowships program: Experiences and outcomes: Final report of the formative evaluation."

Engberg, David, Greggr Glove, Laura Rumbley, and Philip Altbach. 2014. "The Rationale for Sponsoring Students to Undertake International Study: An Assessment of National Student Mobility Scholarship Programmes." *GoingGlobal*.

EPA. 2003. "Air Pollution Control Technology Fact Sheet."

EPA. 2016. "Air Pollutant Emissions Trends Data: Average Annual Emissions: Criteria pollutants National Tier 1 for 1970 - 2016."

Evenson, Robert E., and Larry E. Westphal. 1995. "Chapter 37 Technological change and technology strategy." In *Handbook of Development Economics*. Vol. Volume 3, Part A, , ed. Jere Behrman and T. N. Srinivasan, 2209–2299. Elsevier.

Felbermayr, Gabriel, and Jasmin Gröschl. 2014. "Naturally negative: The growth effects of natural disasters." *Journal of Development Economics*, 111: 92–106.

Feyrer, James. 2007. "Demographics and Productivity." *Review of Economics and Statistics*, 89(1): 100–109.

Fischer, Carolyn, Ian W. H. Parry, and William A. Pizer. 2003. "Instrument choice for environmental protection when environmental technological innovation is endogenous." *Journal of Environmental Economics and Management*, 45: 523–545.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics." In *Handbook of labor economics*. Vol. 4 of *Handbook in economics*, , ed. Orley Ashenfelter, 1–102. Amsterdam [u.a.]:North-Holland.

Fowlie, Meredith. 2010. "Emissions Trading, Electricity Industry Restructuring, and Investment in Pollution Abatement." *American Economic Review*, 100(3): 837–869.

Fuchs-Schundeln, N., and M. Schundeln. 2005. "Precautionary Savings and Self-Selection: Evidence from the German Reunification Experiment." *Quarterly Journal of Economics*, 120(3): 1085–1120.

Fulbright. 2013. "2013 Annual Report."

Galor, Oded, and Omer Moav. 2002. "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of Economics*, 117(4): 1133–1191.

Gerking, Shelby, and Stephen F. Hamilton. 2008. "What explains the increased utilization of Powder River Basin coal in electric power generation?" *American Journal of Agricultural Economics*, 90(4): 933–950.

Giorcelli, Michela. 2016. "The Long-Term Effects of Management and Technology Transfers." *SIEPR Discussion Paper*, , (16-010).

Goeschl, Timo, and Grischa Perino. 2017. "The Climate Policy Hold-Up: Green Technologies, Intellectual Property Rights, and the Abatement Incentives of International Agreements." *Scandinavian Journal of Economics*, 119(3): 709–732.

Goldin, Claudia Dale, and Lawrence F. Katz. 2008. *The race between education and technology.* Cambridge, Mass.:Belknap Press of Harvard University Press.

Gollin, Douglas. 2002. "Getting Income Shares Right." *Journal of Political Economy*, 110(2): 458–474.

Gollin, Douglas. 2010. "Agricultural Productivity and Economic Growth." In *Handbook of Agricultural Economics*. Vol. 4, , ed. P. Pingali and R. Evenson, 3825–3866. Elsevier.

Groen, Jeffrey A., and Anne E. Polivka. 2010. "Going home after Hurricane Katrina: Determinants of return migration and changes in affected areas." *Demography*, 47(4): 821–844.

Grossman, Gene M., and Elhanan Helpman. 1991. *Innovation and growth in the global economy.* Cambridge, Mass.:MIT Press.

GWS. 2007. "Die Elbeflut im August 2002 - Fünf Jahre danach."

Hall, R. E., and C. I. Jones. 1999. "Why do Some Countries Produce So Much More Output Per Worker than Others?" *Quarterly Journal of Economics*, 114(1): 83–116.

Hanushek, Eric A., and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth*, 17(4): 267–321.

Heckman, James J. 1992. "Randomization and Social Policy Program Evaluation." In *Evaluating welfare and training programs.* , ed. Charles F. Manski, 201–230. Cambridge Mass. u.a.:Harvard Univ. Press.

Heckman, James J., and Edward J. Vytlacil. 2007. "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation." In *Handbook of Econometrics*. Vol. Volume 6, Part B, , ed. James J. Heckman Leamer and Edward E., 4779–4874. Elsevier.

Hewson, Thomas, and Philip Graeter. 2016. "Capital Investments in Emission Control Retrofits in the U.S. Coal-fired Generating Fleet through the Years: 2016 Update." *American Coalition for Clean Coal Electricity*.

Hintze, Peter, and Tobias Lakes. 2009. "Geographically Referenced Data in Social Science: A Service Paper for SOEP Data Users."

Hochrainer, Stefan. 2009. "Assessing The Macroeconomic Impacts Of Natural Disasters: Are There Any?" *World Bank Policy Research Working Paper*, , (4968).

Hoffmann, Carola, Heike Matticzk, and Wolf-Dietmar Speich. 2004. "Wirtschaftsentwicklung 2003 in Sachsen." *Statistik in Sachsen*, , (4): 1–21.

Holmes, Thomas J. 1989. "The effect of third-degree price discrimination in oligopoly." *American Economic Review*, 79(1): 244–250.

Horwich, George. 2000. "Economic Lessons of the Kobe Earthquake." *Economic Development and Cultural Change*, 48(3): 521–542.

Hsiang, Solomon, and Amir Jina. 2014. "The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence From 6,700 Cyclones." *NBER Working Paper*, , (20352).

Hsieh, Chang-Tai, and Peter J. Klenow. 2010. "Development Accounting." *American Economic Journal: Macroeconomics*, 2(1): 207–223.

Hughes, Jonathan E. 2011. "The higher price of cleaner fuels: Market power in the rail transport of fuel ethanol." *Journal of Environmental Economics and Management*, 62(2): 123–139.

Hühne, Philipp, Birgit Meyer, and Peter Nunnenkamp. 2014. "Who Benefits from Aid for Trade? Comparing the Effects on Recipient versus Donor Exports." *The Journal of Development Studies*, 50(9): 1275–1288.

IIE, ed. 1991-2011. *Open Doors Report on International Educational Exchange: International Students at All Institutions.* New York:Institute of International Education.

Isaksson, Anders. 2007*a*. "Determinants of Total Factor Productivity: A Literature Review." *UNDAT Research and Statistics Branch Staff Working Paper*, , (02/2007).

Isaksson, Anders. 2007*b*. "Productivity and aggregate growth: A global picture." *UNCDAT Research and Statistics Branch Staff Working Paper*, , (05/2007).

Islam, Nazrul. 1995. "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics*, 110(4): 1127–1170.

Jann, B. 2008. "The Blinder-Oaxaca decomposition for linear regression models." *Stata Journal*, 8(4): 453–479.

Jones, Charles Irving. 1995. "Time series tests of endogenous growth models." *Quarterly Journal of Economics*, 110(2): 495–525.

Judson, Ruth A., and Ann L. Owen. 1999. "Estimating dynamic panel data models: A guide for macroeconomists." *Economics Letters*, 65(1): 9–15.

Jung, Chulho, Kerry Krutilla, and Roy Boyd. 1996. "Incentives for Advanced Pollution Abatement Technology at the Industry Level: An Evaluation of Policy Alternatives." *Journal of Environmental Economics and Management*, 30: 95–111.

Kahn, Matthew E. 2005. "The Death Toll from Natural Disasters: The Role of Income, Geography, and Institutions." *Review of Economics and Statistics*, 87(2): 271–284.

Kalirajan, Kaliappa Pillai. 1991. "The importance of efficient use in the adoption of technology: A micro panel data analysis." *Journal of productivity analysis*, , (2): 113–126.

Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature*, 42(3): 752–782.

Kennedy, Charles. 1964. "Induced Bias in Innovation and the Theory of Distribution." *Economic Journal*, 74(295): 541.

Keohane, Nathaniel O. 2005. "Environmental Policy and the Choice of Abatement Technique: Evidence from Coal-Fired Power Plants." *Working Paper*.

Kim, Jinyoung. 1998. "Economic Analysis of Foreign Education and Students Abroad." *Journal of Development Economics*, 56(2): 337–365.

Kiviet, Jan F. 1995. "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models." *Journal of Econometrics*, 68(1): 53–78.

Klenow, Peter Joseph, N. Gregory Mankiw, Andrés Rodríguez-Clare, and Charles Irving Jones. 1997. "The Neoclassical Revival in Growth Economics: Has it Gone too far?" *NBER macroeconomics annual*.

Kousky, Carolyn, Erwann Michel-Kerjan, and Paul A. Raschky. 2013. "Does Federal Disaster Assistance Crowd Out Private Demand for Insurance?" *Risk Management and Decision Process Center Working Paper*, , (2013-10).

Kugler, Maurice, and Hillel Rapoport. 2007. "International labor and capital flows: Complements or substitutes?" *Economics Letters*, 94(2): 155–162.

Kunreuther, Howard. 1978. *Disaster insurance protection: Public policy lessons.* New York NY u.a.:Wiley.

Lange, Ian, and Allen Bellas. 2005. "Technological Change for Sulfur Dioxide Scrubbers under Market-Based Regulation." *Land Economics*, 81(4): 546–556.

Lange, Ian, and Allen S. Bellas. 2007. "The 1990 Clean Air Act and the implicit price of sulfur in coal." *The BE Journal of Economic Analysis & Policy*, 7(1).

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *The American Economic Review*, 73(1): 31–43.

Leamer, Edward E. 2010. "Tantalus on the Road to Asymptopia." *Journal of Economic Perspectives*, 24(2): 31–46.

Leiponen, Aija. 2005. "Skills and innovation." *International Journal of Industrial Organization*, 23: 303–323.

Le, Thanh. 2010. "Are Student Flows a Significant Channel of R&D Spillovers From the North to the South?" *Economics Letters*, 107(3): 315–317.

Levine, Ross. 2001. "International Financial Liberalization and Economic Growth." *Review of International Economics*, 9(4): 688–702.

Linnerooth-Bayer, Joanne., Simon Quijano-Evans, Ragnar Löfstedt, and Shirin Elahi. 2001. "The Uninsured Elements of Natural Catastrophic Losses: Seven Case Studies of Earthquake and Flood Disasters."

Lin, Tin-Chun. 2004. "The role of higher education in economic development: An empirical study of Taiwan case." *Journal of Asian Economics*, 15(2): 355–371.

Loayza, Norman V., Eduardo Olaberría, Jamele Rigolini, and Luc Christiaensen. 2012. "Natural Disasters and Growth: Going Beyond the Averages." *World Development*, 40(7): 1317–1336.

Loepmeier, F.-J. 2003. "Die agrarmeteorologische Situation." In *Klimastatusbericht 2002*. 92–100. Offenbach, Main:DWD.

Lott, John R., and Russell D. Roberts. 1991. "A Guide to the Pitfalls of Identifying Price Discrimination." *Economic Inquiry*, 29(1): 14–23.

Luechinger, Simon, and Paul A. Raschky. 2009. "Valuing flood disasters using the life satisfaction approach." *Journal of Public Economics*, 93(3-4): 620–633.

Lusardi, Annamaria. 1998. "On the Importance of the Precautionary Saving Motive." *The American Economic Review*, 88(2): 449–453.

Mani, Muthukumara, Michael Keen, and Paul K. Freeman. 2003. "Dealing with Increased Risk of Natural Disasters: Challenges and Options." 37 pages.

Mankiw, N. Gregory, David H. Romer, and David Nathan Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics*, 107(2): 407–437.

Matloff, Norman. 2013. "Are Foreign Students the 'Best and Brightest'? Data and implications for immigration policy." *Economic Policy Institute: Briefing Paper*, , (356).

Mechler, Reinhard. 2009. "Disasters And Economic Welfare: Can National Savings Help Explain Post-Disaster Changes In Consumption?" *World Bank Policy Research Working Paper*, , (4988).

Mechler, Reinhard, and Juergen Weichselgartner. 2003. "Disaster Loss Financing in Germany - The Case of the Elbe River Floods 2002."

Milliman, Scott R., and Raymond Prince. 1989. "Firm Incentives to Promote technological change in pollution control." *Journal of Environmental Economics and Management*, 17: 247–265.

Mohnen, Pierre, and Lars-Hendrik Röller. 2005. "Complementarities in innovation policy." *European Economic Review*, 49(6): 1431–1450.

Mueller, Meike, and Annegret Thieken. 2005. "Hochwasserschäden bei Unternehmen in Sachsen: Erfahrungen aus dem Augusthochwasser 2002." *Schadensprisma*, , (2): 22–31.

Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *American Economic Review*, 101(5): 1649–1675.

Murphy, Kevin M., Andrej Shleifer, and Robert W. Vishny. 1993. "Why is rent-seeking so costly to growth?" *The American Economic Review*.

Nelson, Richard R., and Edmund S. Phelps. 1966. "Investment in Humans, Technological Diffusion, and Economic Growth." *The American Economic Review*, 56(1/2): 69–75.

Neumark, David. 1988. "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination." *The Journal of Human Resources*, 23(3): 279.

Nevo, Aviv, and Michael D. Whinston. 2010. "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2): 69–82.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49(6): 1417–1426.

Noy, Ilan. 2009. "The macroeconomic consequences of disasters." *Journal of Development Economics*, 88(2): 221–231.

Noy, Ilan, and Aekkanush Nualsri. 2007. "What do Exogenous Shocks Tell Us about Growth Theories?"

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, 14(3): 693–709.

Oaxaca, Ronald L., and Michael R. Ransom. 1994. "On discrimination and the decomposition of wage differentials." *Journal of Econometrics*, 61(1): 5–21.

OECD. 2014. *Education at a Glance 2014: OECD Indicators.* Paris:OECD Publishing.

OECD. 2015. *Main Science and Technology Indicators, Volume 2015 Issue 1.* Vol. 2015, OECD Publishing.

Oster, Emily. 2017. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 40: 1–18.

Page, Lionel, David A. Savage, and Benno Torgler. 2014. "Variation in risk seeking behaviour following large losses: A natural experiment." *European Economic Review*, 71: 121–131.

Park, Jungsoo. 2004. "International Student Flows and R&D Spillovers." *Economics Letters*, 82(3): 315–320.

Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. "STEM Workers, H-1B Visas, and Productivity in US Cities." *Journal of Labor Economics*, 33(S1): 225–255.

Perino, Grischa. 2010. "Technology Diffusion with Market Power in the Upstream Industry." *Environmental and Resource Economics*, 46(4): 403–428.

Perino, Grischa, and Olena Talavera. 2014. "The Benefits of Spatially Differentiated Regulation: The Response to Acid Rain by US States Prior to the Acid Rain Program." *American Journal of Agricultural Economics*, 96(1): 108–123.

Perino, Grischa, and Till Requate. 2012. "Does more stringent environmental regulation induce or reduce technology adoption? When the rate of technology adoption is inverted U-shaped." *Journal of Environmental Economics and Management*, 64(3): 456–467.

Popp, David. 2003. "Pollution Control Innovations and the Claen Air Act of 1990." *Journal of Policy Analysis and Management*, 22(4): 641–660.

Poullikkas, Andreas. 2015. "Review of Design, Operating, and Financial Considerations in Flue Gas Desulfurization Systems." *Energy Technology & Policy*, 2(1): 92–103.

Project Atlas. 2015. "Trends and Global Data Fact Sheet."

Quine, W. V. 1951. "Main Trends in Recent Philosophy: Two Dogmas of Empiricism." *The Philosophical Review*, 60(1): 20.

Raddatz, Claudio. 2009. "The Wrath Of God: Macroeconomic Costs Of Natural Disasters." *World Bank Policy Research Working Paper*, , (5039).

Raschky, Paul A., and Hannelore Weck-Hannemann. 2007. "Charity hazard—A real hazard to natural disaster insurance?" *Environmental Hazards*, 7(4): 321–329.

Rauch, James E., and Alessandra Casella. 2003. "Overcoming Informational Barriers to International Resource Allocation: Prices and Ties." *Economic Journal*, 113(484): 21–42.

Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics*, 84(1): 116–130.

Reimers, Cordelia W. 1983. "Labor Market Discrimination Against Hispanic and Black Men." *Review of Economics and Statistics*, 65(4): 570.

Requate, Till. 2005. "Dynamic incentives by environmental policy instruments - a survey." *Ecological Economics*, 54: 175–195.

Requate, Till, and Wolfram Unold. 2003. "Environmental policy incentives to adopt advanced abatement technology: Will the true ranking please stand up?" *European Economic Review*, 47: 125–146.

Roodman, D. 2009*a*. "How to do xtabond2: An introduction to difference and system GMM in Stata." *Stata Journal*, 9(1): 86–136.

Roodman, David. 2009*b*. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics*, 71(1): 135–158.

Roson, Roberto, Alvaro Calzadilla, and Francesco Pauli. 2006. "Climate Change and Extreme Events: An Assessment of Economic Implications." *FEEm Working Paper*, , (44).

Rudolf, B., and J. Rapp. 2003. "Das Jahrhunderthochwasser der Elbe: Synoptische Wetterentwicklung und klimatologische Aspekte." In *Klimastatusbericht 2002*. 172–187. Offenbach, Main:DWD.

Ruiz, Neil G. 2014. "The Geography of Foreign Students in U.S. Higher Education: Origin and Destinations."

Sawada, Yasuyuki, and Satoshi Shimizutani. 2008. "How Do People Cope with Natural Disasters? Evidence from the Great Hanshin-Awaji (Kobe) Earthquake in 1995." *Journal of Money, Credit and Banking*, 40(2-3): 463–488.

Sexenian, Anna. 2002. "Brain circulation: How high-skill immigration makes everyone better off." *The Brooking Review*, 20(1): 28–31.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference.* Belmont, CA:Wadsworth Cengage Learning.

Sims, Christopher A. 2010. "But Economics Is Not an Experimental Science." *Journal of Economic Perspectives*, 24(2): 59–68.

Skidmore, Mark. 2001. "Risk, natural disasters, and household savings in a life cycle model." *Japan and the World Economy*, 13(1): 15–34.

Skidmore, Mark, and Hideki Toya. 2002. "Do Natural Disasters Promote Lon-Run Growth?" *Economic Inquiry*, 40(4): 664–687.

Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39(3): 312–320.

Spilimbergo, Antonio. 2009. "Democracy and Foreign Education." *The American Economic Review*, 99(1): 528–543.

Srivastava, R. K., and W. Jozewicz. 2001. "Flue Gas Desulfurization: The State of the Art." *Journal of the Air & Waste Management Association (1995)*, 51(12): 1676–1688.

Striefler, Christian. 2003*a*. *Augusthochwasser 2002: Der Wiederaufbau im Freistaat Sachsen ein Jahr nach der Flut.* . 1. Aufl. ed., Dresden:Zentraler Broschürenversand der Sächsischen Staatsregierung.

Striefler, Christian, ed. 2003*b*. *Schadensausgleich und Wiederaufbau im Freistaat Sachsen: Augusthochwasser 2002.* . 1. Aufl., Stand: 29. Januar 2003 ed., Dresden:Zentraler Broschürenversand der Sächs. Staatsregierung.

Strobl, Eric. 2012. "The economic growth impact of natural disasters in developing countries: Evidence from hurricane strikes in the Central American and Caribbean regions." *Journal of Development Economics*, 97(1): 130–141.

Taylor, Margaret R., Edward R. Rubin, and David A. Hounshell. 2005. "Control of SO_2 emissions from power plants: A case of induced technological innovation in the U.S." *Technological Forecasting & Social Change*, 72: 697–718.

Thomas, Vinod. 2014. "Confronting climate-related disasters in Asia and the Pacific." *Jahrbuch für Wirtschaftswissenschaften/Review of Economics*, 65(2): 121–136.

Tinbergen, Jan. 1942. "Zur Theorie der langfristigen Wirtschaftsentwicklung." *Weltwirtschaftliches Archiv*, 55(3): 511–549.

Uchii, Soshichi. 2010. "Sherlock Holmes on reasoning." *PhilPapers*.

UNESCO Institute for Statistics. 2009. *Global Education Digest 2009: Global trends in tertiary education. Global Education Digest*, Montreal:United Nations Educational Scientific and Cultural Organization (UNESCO).

UNESCO Institute for Statistics. 2012. *Global Education Digest 2012: Opportunities lost: The impact of grade repetition and early school leaving.* Vol. 2012 of *Global Education Digest*, Montreal:United Nations Educational Scientific and Cultural Organization (UNESCO).

UNISDR - United Nations International Strategy for Disaster Reduction Secretariat. 2009. *Global assessment report on disaster risk reduction: Risk and poverty in a changing climate : invest today for a safer tomorrow.* New York:United Nations International Strategy for Disaster Reduction Secretariat (UNISDR).

van Asseldonk, Marcel A. P. M., Miranda P. M. Meuwissen, and Ruud B. M. Huirne. 2002. "Belief in Disaster Relief and the Demand for a Public-Private Insurance Program." *Review of Agricultural Economics*, 24(1): 196–207.

Vigdor, Jacob L. 2007. "The Katrina Effect: Was There a Bright Side to the Evacuation of Greater New Orleans?" *The B.E. Journal of Economic Analysis & Policy*, 7(1).

Wadhwa, Vivek. 2012. *The immigrant exodus: Why America is Losing the Global Race to Capture Entrepreneurial Talent.* Philadelphia, PA:Wharton Digital Press.

Wadhwa, Vivek, AnnaLee Saxenian, Richard Freeman, and Alex Salkever. 2009. *Losing the World's Best and Brightest: America's New Immigrant Entrepreneurs.* Vol. 5.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp. 2007. *The German socio-economic panel study (SOEP): Scope, evolution and enhancements.* Vol. 01 of *SOEP papers on multi-disciplinary panel data research*, Berlin.

Walker, Ian, and Yu Zhu. 2013. "The Benefit of STEM Skills to Individuals, Society, and the Economy: Report to Royal Society's Vision for Science and Mathematics." In .

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126(1): 25–51.

Appendix

A Do Natural Disasters Affect Individual Saving? Evidence from a Natural Experiment in a Highly Developed Country?

Table A.1: Description of Variables

Variable	Description
S	Amount saved per month discounted to 2000 prices using the price index provided by the German federal statistical office. As the variable is only reported on the household level, the amount saved is allocated to all members of the household older than 18 using individual gross income relative to the sum of gross individual income[a] of all household members in a given year.
S_E	Binary variable. Extensive margin of the saving decision. One if S is greater than one and zero otherwise.
S_I	Intensive margin of the saving decision. Only observations for which S is greater than zero.
S'_E	Binary variable. One if S is greater than EUR 50 and zero otherwise.
SR	Saving rate. The individual monthly amount saved (S) divided by the individuals gross income.
Sex	Binary variable. One if individual is male and zero if female.
Age	Age of individual.
$Homeowner$	Binary variable. One if household owns the dwelling they live in and zero otherwise.
$Rural$	Binary variable. One if household lives in a rural area and otherwise.
$PrimaryEducation$	Reference category.
$SecondaryEducation$	Binary (factor) variable. One if the highest education obtained is secondary and zero otherwise.
$TertiaryEducation$	Binary (factor) variable. One if the highest education obtained is tertiary and zero otherwise.
$Single$	Reference category.
$Married$	Binary (factor) variable. One if the person is married, zero otherwise.
$Other$	Binary (factor) variable. One if the person is divorced, widowed or married but living apart and zero otherwise.
$Nochild$	Reverence category.
$OneChild$	Binary (factor) variable. One if one child lives in household and zero otherwise.
$TwoChildren$	Binary (factor) variable. One if two children live in household and zero otherwise.
$Two+Children$	Binary (factor) variable. One if three or more children live in household and zero otherwise.
$RuralArea$	Binary variable. One if household is located in rural area according to the regional classification of the Federal Institute for Research on Building, Urban Affairs and Spatial Development and zero otherwise.

[a] Includes all reported income sources of an individual reported in the person-level survey of the SOEP.

Table A.2: Effect on Individual Saving Volume (balanced regressions)

Tobit Regression Var./ Dep. Var.	(1) S	(2) S	(3) S
Year	-8.108	-11.637	0.404
	(10.341)	(10.955)	(15.355)
Treated	-23.774	-16.321	-7.607
	(53.620)	(53.078)	(62.799)
Year × Treated	-71.623	-175.168**	-178.232***
	(59.872)	(72.204)	(62.958)
Sex	54.680***	59.920***	69.552***
	(12.400)	(10.588)	(13.810)
Age	2.712***	3.088***	2.392**
	(0.985)	(0.855)	(0.983)
Homeowner	35.780	23.098	23.106
	(22.651)	(19.822)	(24.755)
Primary			
1. Secondary	31.927	28.474	-12.134
	(19.497)	(18.427)	(36.217)
2. Tertiary	168.600***	141.275***	127.179***
	(28.538)	(23.495)	(40.217)
Working			
1. Unemployed	-280.895***	-260.394***	-363.850***
	(33.822)	(29.022)	(46.502)
2. Non-Working	-107.319***	-113.326***	-132.818***
	(21.647)	(20.061)	(24.290)
Single			
1. Married	6.039	5.552	35.565
	(35.851)	(29.222)	(35.481)
2. Other	19.299	-7.971	38.422
	(43.054)	(37.585)	(47.547)
No Child			
1. One Child	-34.553	-25.633	-20.089
	(27.573)	(25.665)	(28.820)
2. Two Children	-141.416***	-95.973**	-90.577
	(44.773)	(38.974)	(56.508)
3. Two+ Children	-127.893*	-83.553	-190.916***
	(74.787)	(77.553)	(65.959)
Rural Area	-56.259**	-25.805	-33.961
	(25.303)	(22.827)	(31.527)
Constant	-101.187***	-108.376***	-82.989*
	(39.023)	(35.444)	(44.887)
Sigma	288.269***	258.998***	342.330***
	(27.238)	(17.740)	(40.681)
Log Pseudolikelihood	-10599.234	-9888.180	-9818.446
Observations	2188	2068	1974
Censored	764	722	683

Notes: We report coefficients on the latent variable. See footnotes table 2. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Effect on Individual Extensive Margin (balanced regression)

Probit Regression Var./ Dep. Var.	(1) S_E	(2) S_E	(3) S_E
Year	-0.068	-0.072	-0.034
	(0.056)	(0.058)	(0.066)
Treated	-0.012	0.018	0.032
	(0.286)	(0.289)	(0.295)
Year × Treated	-0.163	-0.789**	-0.628**
	(0.304)	(0.355)	(0.272)
Sex	0.041	0.043	0.046
	(0.042)	(0.045)	(0.045)
Age	0.011**	0.013***	0.012***
	(0.004)	(0.004)	(0.004)
Homeowner	0.034	0.003	-0.057
	(0.105)	(0.104)	(0.101)
Primary			
1. Secondary	0.002	0.013	-0.069
	(0.107)	(0.112)	(0.117)
2. Tertiary	0.316**	0.327**	0.243*
	(0.127)	(0.128)	(0.132)
Working			
1. Unemployed	-1.014***	-0.983***	-1.173***
	(0.115)	(0.113)	(0.120)
2. Non-Working	-0.194**	-0.194**	-0.228**
	(0.095)	(0.092)	(0.091)
Single			
1. Married	0.127	0.046	0.074
	(0.142)	(0.148)	(0.141)
2. Other	-0.136	-0.353**	-0.342**
	(0.164)	(0.173)	(0.162)
No Child			
1. One Child	-0.037	0.028	0.079
	(0.133)	(0.136)	(0.129)
2. Two Children	-0.489***	-0.319*	-0.323*
	(0.189)	(0.188)	(0.180)
3. Two+ Children	-0.655**	-0.682**	-0.523**
	(0.290)	(0.271)	(0.248)
Rural Area	-0.160	0.033	-0.074
	(0.119)	(0.130)	(0.124)
Constant	-0.001	-0.095	0.052
	(0.193)	(0.192)	(0.199)
Log pseudolikelihood	-1285.720	-1218.389	-1155.581
Observations	2188	2068	1974

Notes: See footnotes table 2. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Effect on Individual Intensive Margin (balanced regression)

OLS Regression	(1)	(2)	(3)
Var./ Dep. Var.	S_I	S_I	S_I
Year	-13.543	5.776	4.700
	(19.782)	(21.420)	(30.342)
Treated	-110.168	-75.464	-95.487
	(80.817)	(117.896)	(105.987)
Year × Treated	-108.249	-133.760	-87.587
	(71.871)	(137.898)	(84.398)
Sex	-5.358	-0.696	-7.204
	(17.386)	(11.151)	(12.947)
Age	-1.976	-1.658	-3.013
	(1.831)	(1.638)	(2.089)
Homeowner	178.735***	148.158***	176.024***
	(59.852)	(40.140)	(46.960)
Primary			
1. Secondary	84.674**	60.101*	69.931*
	(37.167)	(34.764)	(39.227)
2. Tertiary	293.710***	185.286***	254.676***
	(65.896)	(47.075)	(61.181)
Working			
1. Unemployed	-177.198***	-154.377***	-206.351***
	(59.604)	(50.069)	(63.410)
2. Non-Working	-104.474**	-107.957***	-149.555***
	(41.115)	(37.178)	(45.781)
Single			
1. Married	-3.838	73.798	144.077**
	(66.881)	(50.017)	(67.697)
2. Other	-46.808	24.119	124.527*
	(73.682)	(60.837)	(74.527)
No Child			
1. One Child	-121.692	-86.350	-90.664*
	(74.535)	(54.689)	(54.910)
2. Two Children	-269.129***	-190.542***	-237.705**
	(91.900)	(73.449)	(94.793)
3. Two+ Children	-125.298	177.137	-250.787***
	(104.758)	(134.341)	(96.387)
Rural Area	-99.170*	-76.005*	-84.561*
	(59.401)	(45.328)	(46.803)
Constant	407.723***	344.274***	373.064***
	(84.230)	(76.932)	(80.665)
Adjusted R^2	0.122	0.128	0.123
Observations	1276	1196	1096

Notes: See footnotes table 2. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Effect on Individual Intensive Margin (balanced regression)

Tobit Regression Var./ Dep. Var.	(1) *SR*	(2) *SR*	(3) *SR*
Year	-0.006	-0.006	-0.006
	(0.006)	(0.005)	(0.006)
Treated	-0.015	-0.012	-0.007
	(0.027)	(0.027)	(0.028)
Year × Treated	-0.017	-0.098**	-0.080***
	(0.037)	(0.038)	(0.027)
Sex	0.005	0.007	0.009*
	(0.005)	(0.005)	(0.005)
Age	0.002***	0.002***	0.001***
	(0.001)	(0.000)	(0.000)
Homeowner	0.021*	0.018	0.019*
	(0.011)	(0.011)	(0.011)
Primary			
1. Secondary	0.011	0.013	0.005
	(0.013)	(0.012)	(0.013)
2. Tertiary	0.046***	0.042***	0.043***
	(0.014)	(0.013)	(0.015)
Working			
1. Unemployed	-0.110***	-0.102***	-0.135***
	(0.015)	(0.014)	(0.017)
2. Non-Working	0.003	-0.002	-0.004
	(0.011)	(0.011)	(0.011)
Single			
1. Married	-0.007	-0.008	-0.002
	(0.016)	(0.015)	(0.016)
2. Other	-0.014	-0.019	-0.020
	(0.020)	(0.020)	(0.020)
No Child			
1. One Child	-0.013	-0.007	-0.005
	(0.014)	(0.014)	(0.013)
2. Two Children	-0.063***	-0.048**	-0.058***
	(0.020)	(0.019)	(0.020)
3. Two+ Children	-0.082**	-0.082***	-0.103***
	(0.034)	(0.031)	(0.026)
Rural Area	-0.030**	-0.015	-0.021
	(0.012)	(0.012)	(0.013)
Constant	-0.033	-0.039*	-0.021
	(0.021)	(0.020)	(0.022)
Sigma	0.139***	0.132***	0.139***
	(0.007)	(0.007)	(0.007)
Log pseudolikelihood	214.722	270.495	199.660
Observations	2180	2064	1972
Censored	764	720	682

Notes:We report coefficients on the latent variable. See footnotes table 3. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Placebo Individual Saving Behavior (balanced regression)

	(1)	(2)	(3)	(4)	(5)	(6)
Model	Tobit	Tobit	Probit	Probit	OLS	OLS
Var. / Dep. Var.	S	S	S_E	S_E	S_I	S_I
Year	-0.393	-15.477	-0.014	-0.117*	12.878*	-2.250
	(7.707)	(9.609)	(0.059)	(0.066)	(6.793)	(10.221)
Treated	-50.814	-43.741	0.080	0.129	-55.790*	-79.411**
	(32.284)	(31.581)	(0.273)	(0.274)	(31.279)	(34.711)
Year $\times$ Treated	-9.201	24.455	-0.225	-0.034	10.425	30.761
	(33.313)	(27.627)	(0.217)	(0.196)	(33.858)	(27.025)
Sex	36.125***	38.011***	-0.002	-0.019	49.201***	50.385***
	(8.654)	(9.554)	(0.045)	(0.044)	(8.978)	(10.471)
Age	4.473***	3.516***	0.018***	0.011***	4.016***	3.305***
	(0.777)	(0.771)	(0.004)	(0.004)	(0.732)	(0.904)
Homeowner	54.474***	45.484***	0.295***	0.163	17.404	23.243
	(15.630)	(17.524)	(0.109)	(0.106)	(14.286)	(17.284)
Primary						
1. Secondary	50.931***	42.419***	0.168	0.134	37.564***	40.042***
	(16.226)	(16.419)	(0.118)	(0.106)	(13.353)	(15.102)
2. Tertiary	137.215***	125.127***	0.452***	0.406***	102.006***	103.133***
	(19.716)	(22.546)	(0.136)	(0.123)	(16.471)	(21.553)
Working						
1. Unemployed	-205.862***	-210.379***	-0.936***	-0.881***	-131.681***	-141.965***
	(20.797)	(24.236)	(0.113)	(0.122)	(13.375)	(15.433)
2. Non-Working	-101.660***	-98.829***	-0.192**	-0.235***	-117.413***	-114.438***
	(16.986)	(17.457)	(0.096)	(0.091)	(16.558)	(19.900)
Single						
1. Married	-34.516	-20.267	0.075	0.133	-68.123***	-50.820
	(24.725)	(27.477)	(0.135)	(0.128)	(23.362)	(32.493)
2. Other	-22.632	-14.964	-0.274	-0.159	0.264	12.251
	(32.245)	(32.102)	(0.167)	(0.154)	(30.905)	(37.625)
No Child						
1. One Child	-15.276	-14.284	-0.063	-0.097	4.635	0.401
	(22.454)	(25.783)	(0.147)	(0.146)	(21.324)	(30.606)
2. Two Children	-34.080	-57.121**	-0.371**	-0.420**	33.692	-2.070
	(28.936)	(28.792)	(0.170)	(0.164)	(27.244)	(28.299)
3. Two+ Children	17.925	-64.528	-0.550**	-0.865***	190.123	141.987
	(80.924)	(78.042)	(0.280)	(0.262)	(147.471)	(141.382)
Rural Area	-39.760**	-32.729	-0.189	-0.102	-22.362	-34.542*
	(18.835)	(20.314)	(0.122)	(0.125)	(18.327)	(19.039)
Constant	-158.122***	-114.959***	-0.404**	-0.042	-18.362	12.494
	(31.457)	(32.602)	(0.186)	(0.179)	(28.068)	(30.021)
Sigma	205.368***	227.751***	NA	NA	NA	NA
	(11.560)	(20.325)				
Log pseudolikelihood	-10653.729	-10009.988	-1230.240	-1192.633	-9037.910	-8330.881
Adjusted R^2	NA	NA	NA	NA	0.171	0.122
Observations	2208	2062	2208	2062	1382	1242
Censored	703	670	NA	NA	NA	NA

Notes: We report coefficients on the latent variable in case of the tobit model. See footnotes table 4. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Placebo Test Individual Saving Rate (balanced regression)

Tobit Regression Var / Dep. Var.	(1) *SR*	(2) *SR*
Year	0.003	-0.013**
	(0.005)	(0.006)
Treated	-0.037*	-0.034*
	(0.019)	(0.018)
Year × Treated	-0.006	0.024
	(0.022)	(0.018)
Sex	0.002	0.001
	(0.005)	(0.005)
Age	0.003***	0.002***
	(0.001)	(0.001)
Homeowner	0.041***	0.032***
	(0.011)	(0.011)
Primary		
1. Secondary	0.021*	0.020*
	(0.012)	(0.011)
2. Tertiary	0.048***	0.043***
	(0.014)	(0.013)
Working		
1. Unemployed	-0.093***	-0.091***
	(0.014)	(0.014)
2. Non-Working	-0.004	-0.002
	(0.010)	(0.010)
Single		
1. Married	-0.022	-0.014
	(0.017)	(0.016)
2. Other	-0.032	-0.021
	(0.021)	(0.020)
No Child		
1. One Child	-0.014	-0.013
	(0.014)	(0.014)
2. Two Children	-0.028	-0.032*
	(0.019)	(0.018)
3. Two+ Children	-0.048*	-0.088***
	(0.028)	(0.027)
Rural Area	-0.032***	-0.024**
	(0.012)	(0.012)
Constant	-0.074***	-0.041*
	(0.023)	(0.021)
Sigma	0.129***	0.129***
	(0.005)	(0.006)
Log pseudolikelihood	404.673	351.105
Observations	2204	2058
Censored	703	670

Notes: We report coefficients on the latent variable. See footnotes table 5. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Stability Test Individual Extensive Margin (balanced regression)

Probit Regression Var. / Dep. Var.	(1) S'_E	(2) S'_E	(3) S'_E
Year	0.029	-0.013	0.009
	(0.048)	(0.052)	(0.058)
Treated	0.019	-0.022	0.031
	(0.246)	(0.253)	(0.256)
Year × Treated	-0.270	-1.054***	-0.735***
	(0.314)	(0.330)	(0.229)
Sex	0.153***	0.198***	0.186***
	(0.050)	(0.051)	(0.050)
Age	0.024***	0.026***	0.024***
	(0.004)	(0.004)	(0.004)
Homeowner	0.034	0.021	-0.014
	(0.098)	(0.096)	(0.095)
Primary			
1. Secondary	0.374***	0.269**	0.192
	(0.116)	(0.119)	(0.120)
2. Tertiary	0.813***	0.668***	0.608***
	(0.132)	(0.133)	(0.132)
Working			
1. Unemployed	-1.380***	-1.451***	-1.395***
	(0.142)	(0.131)	(0.138)
2. Non-Working	-0.489***	-0.516***	-0.533***
	(0.102)	(0.094)	(0.097)
Single			
1. Married	0.035	-0.021	-0.053
	(0.138)	(0.143)	(0.137)
2. Other	-0.086	-0.211	-0.248
	(0.168)	(0.173)	(0.160)
No Child			
1. One Child	-0.044	0.099	0.077
	(0.127)	(0.128)	(0.123)
2. Two Children	-0.413**	-0.237	-0.167
	(0.171)	(0.178)	(0.172)
3. Two+ Children	-0.767***	-0.472*	-0.641**
	(0.260)	(0.279)	(0.273)
Rural Area	-0.139	-0.061	-0.106
	(0.111)	(0.116)	(0.119)
Constant	-1.325***	-1.314***	-1.083***
	(0.204)	(0.198)	(0.191)
Log pseudolikelihood	-1279.031	-1221.087	-1185.816
Observations	2188	2068	1974

Notes: S'_E is binary taking the value [1] if $S >$ €50 and zero otherwise. See footnotes table 6. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Stability Test Household Saving Behavior (balanced regression)

Model	(1) Tobit S	(2) Probit S_E	(3) OLS S_I	(4) Tobit SR
Year3	-3.241	-0.036	9.183	-0.003
	(18.851)	(0.057)	(22.709)	(0.007)
Year4	-11.473	-0.046	22.805	-0.005
	(20.196)	(0.059)	(23.380)	(0.007)
Year5	-12.503	-0.068	4.166	-0.010
	(24.936)	(0.067)	(25.025)	(0.008)
Treated	-36.209	0.012	100.166	-0.011
	(116.914)	(0.305)	(215.624)	(0.044)
Year3 × Treated	-134.242	-0.125	-341.152*	-0.036
	(101.949)	(0.313)	(202.821)	(0.044)
Year4 × Treated	-341.631***	-0.828**	-308.233	-0.126***
	(124.324)	(0.365)	(260.373)	(0.043)
Year5 × Treated	-198.432**	-0.506*	-118.224	-0.072**
	(88.267)	(0.303)	(139.815)	(0.032)
Sex	9.780	0.049	-25.491	0.010
	(39.650)	(0.102)	(45.505)	(0.014)
Age	3.065*	0.015***	-0.248	0.002**
	(1.686)	(0.005)	(2.019)	(0.001)
Homeowner	49.995	-0.117	133.903***	0.013
	(40.087)	(0.105)	(44.979)	(0.015)
Primary				
1. Secondary	16.786	0.109	31.231	0.014
	(59.035)	(0.169)	(46.882)	(0.025)
2. Tertiary	247.487***	0.464**	206.972***	0.069***
	(65.150)	(0.180)	(55.948)	(0.026)
Working				
1. Unemployed	-399.251***	-0.929***	-179.338**	-0.130***
	(62.864)	(0.138)	(80.789)	(0.024)
2. Non-Working	-123.284***	-0.060	-154.783***	-0.009
	(44.391)	(0.122)	(53.147)	(0.018)
Single				
1. Married	52.308	0.142	3.518	-0.011
	(58.643)	(0.159)	(68.372)	(0.022)
2. Other	-92.691	-0.321*	-52.610	-0.041
	(65.786)	(0.178)	(72.636)	(0.025)
No Child				
1. One Child	-34.773	0.043	-59.778	-0.014
	(52.006)	(0.129)	(61.905)	(0.018)
2. Two Children	-141.202*	-0.227	-141.053	-0.059**
	(81.884)	(0.180)	(105.314)	(0.026)
3. Two+ Children	-227.673*	-0.540*	66.069	-0.126***
	(135.129)	(0.286)	(102.084)	(0.042)
Rural Area	-45.102	-0.008	-88.704*	-0.014
	(45.501)	(0.122)	(45.673)	(0.016)
Constant	-47.529	-0.374	387.933***	-0.014
	(97.550)	(0.270)	(116.579)	(0.040)
Sigma	460.260***	NA	NA	0.161***
	(34.794)			(0.007)
Adjusted R^2	NA	NA	0.147	NA
Log pseudolikelihood	-10478.447	-1174.316	-6955.626	-5.407
Observations	2024	2024	948	1976
Censored	702			693

Notes: We report coefficients on the latent variable in case of the tobit models. See footnotes table 7. Standard errors in parentheses and clustered on household level. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Foreign Education and Domestic Productivity

Table B.1: Robustness of Dynamic Model (reduced instrument count)

	Sys. GMM (1)	Sys. GMM (2)	Sys. GMM (3)	Sys. GMM (4)
TFP growth$_{t-1}$	0.182***	0.178***	0.212***	0.145***
	(0.0543)	(0.0534)	(0.0545)	(0.0482)
ln TFP rel. US$_{t-1}$	-0.0268***	-0.0380***	0.00161	-0.0179
	(0.00916)	(0.00870)	(0.0182)	(0.0162)
ln Fstud$_{t-1}$	0.0344***	0.0284**	0.0355***	0.0376**
	(0.00983)	(0.0120)	(0.0114)	(0.0159)
ln Age2029$_{t-1}$	-0.00875	-0.0122	-0.00810	-0.0117
	(0.0133)	(0.0115)	(0.0154)	(0.0153)
ln HumanCap$_{t-1}$		0.0401**		0.0301*
		(0.0200)		(0.0179)
ln Imports$_{t-1}$			-0.0195*	-0.0157*
			(0.00993)	(0.00851)
Constant	-0.145	-0.148	0.279***	0.153
	(0.177)	(0.124)	(0.0739)	(0.132)
AR(1) Test	0.000	0.000	0.000	0.000
AR(2) Test	0.623	0.743	0.991	0.967
Hansen's J test	0.699	0.199	0.398	0.169
Number of Instruments	35	39	36	40
Number of Countries	102	94	101	93
Observations	1911	1765	1853	1707

Notes: The lagged dependent variable, *TFPrel.US*, *Fstud* and *HumanCap* are treated as potentially endogenous and are instrumented using their own second to fourth lag. *Age2029* is treated as predetermined and instrumented by its own first to third lag. Lagged imports are not instrumented. For all instrumented variables the collapse option is used to reduce instrument count (see Roodman, 2009*a*). AR(1) and AR(2) are Arellano-Bond test for serial correlation. Hansens J tests the null hypothesis for violation of over-identification restrictions. For all tests p-values are reported. Standard errors are robust using the Windmeijer (2005) procedure and are reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Robustness of Dynamic Model (constant country sample)

	Pooled OLS	LSDV	Sys. GMM	Sys. GMM	Sys. GMM	Sys. GMM
	(1)	(2)	(3)	(4)	(5)	(6)
TFP growth$_{t-1}$	0.290***	0.240***	0.243***	0.235***	0.238***	0.216***
	(0.0412)	(0.0389)	(0.0440)	(0.0434)	(0.0480)	(0.0461)
ln TFP rel. US$_{t-1}$	-0.00622***	-0.118***	-0.0199***	-0.0252***	-0.0168*	-0.0157**
	(0.00147)	(0.0173)	(0.00679)	(0.00460)	(0.00899)	(0.00635)
ln Fstud$_{t-1}$	0.00216*	0.0184***	0.0197**	0.0184**	0.0242**	0.0263**
	(0.00110)	(0.00560)	(0.00813)	(0.00757)	(0.00944)	(0.01000)
ln Age2029$_{t-1}$	-0.00183	0.00604	-0.0207**	-0.0143*	-0.0200**	-0.0143**
	(0.00114)	(0.0119)	(0.00893)	(0.00767)	(0.00862)	(0.00664)
ln HumanCap$_{t-1}$				0.0232*		0.0229*
				(0.0126)		(0.0121)
ln Imports$_{t-1}$					-0.00353	-0.00868*
					(0.00497)	(0.00490)
Constant	0.00950	-0.382**	0.138*	0.00545	0.176**	0.144
	(0.0111)	(0.189)	(0.0758)	(0.0821)	(0.0792)	(0.0949)
AR(1) Test			0.00	0.00	0.00	0.00
AR(2) Test			0.79	0.81	0.71	0.78
Hansen's J test			0.19	0.26	0.29	0.33
Number of Instruments			96	98	97	99
Number of Countries	93	93	93	93	93	93
Observations	1707	1707	1707	1707	1707	1707
Adjusted R^2	0.224	0.271				

Notes: The selection of the constant country sample is determined by data availability. Pooled OLS suffers from endogeneity bias as fixed effects are excluded while LSDV suffers from Nickell (1981) bias. System GMM is consistent and unbiased. The lagged dependent variable, ln *TFP rel. US*, *Fstud* and *HumanCap* are treated as potentially endogenous and are instrumented using their own second lag. *Age2029* is treated as predetermined and instrumented by its own first lag. Lagged imports are not instrumented. For the instrumentation of *Fstud*,*Age2029* and *HumanCap* the collapse option is used to reduce instrument count (see Roodman, 2009*a*). AR(1) and AR(2) are Arellano-Bond test for serial correlation. Hansens J tests the null hypothesis for violation of over-identification restrictions. For all tests p-values are reported. Standard errors are reported in parentheses. For column (1) and (2) HAC standard errors are reported clustered on country-level. For column (3) to (5) robust standard errors are reported using the Windmeijer (2005) procedure. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Robustness for Developing and Industrialized Countries (full model)

	LSDV		LSDVc		System GMM	
	Developing	High	Developing	High	Developing	High
	(1)	(2)	(3)	(4)	(5)	(6)
L.tfp1_gr	0.235***	0.194***	0.292***	0.257***	0.158***	0
	(0.0400)	(0.0660)	(0.0271)	(0.0530)	(0.0517)	(.)
ln TFP rel. US$_{t-1}$	-0.140***	-0.124***	-0.139***	-0.124***	-0.0217	0.248
	(0.0188)	(0.0225)	(0.0139)	(0.0205)	(0.0175)	(0.161)
Log Fstud_t-1	0.0161***	-0.00161	0.0143***	-0.00160	0.0373**	0.0113
	(0.00604)	(0.00531)	(0.00387)	(0.00585)	(0.0143)	(0.00819)
Log Age 20-29_t-1	-0.0198	-0.00220	-0.0183	-0.00190	-0.0172	0.0915
	(0.0146)	(0.0136)	(0.0145)	(0.0136)	(0.0144)	(0.0741)
HumanCap_t-1	-0.0110	0.0255	-0.0120	0.0267	0.0220	0.653
	(0.0252)	(0.0364)	(0.0187)	(0.0274)	(0.0192)	(0.414)
Imports_t-1	0.00537*	-0.00229	0.00527**	-0.00257	-0.0124	-0.0984
	(0.00310)	(0.00326)	(0.00237)	(0.00383)	(0.00752)	(0.0724)
Constant	-0.165	0.00723			0.173	-0.543
	(0.234)	(0.241)			(0.144)	(0.396)
AR(1) Test					0.000	0.007
AR(2) Test					0.791	0.791
Hansen's J test					0.552	1.000
Number of Instruments					40	40
Number of Countries	69	24	69	24	69	24
Observations	1258	449	1258	449	1258	449
Adjusted R^2	0.276	0.460				

Notes: All regression include time and country fixed effects. System GMM is consistent and unbiased but LSDVc might be preferable if N is moderate (cf. Judson and Owen, 1999). The lagged dependent variable, *TFPrel.US*, *Fstud* and *HumanCap* are treated as potentially endogenous and are instrumented using their own second to fourth lag. *Age2029* is treated as predetermined and instrumented by its own first to third lag. Lagged imports are not instrumented. The collapse option is used to reduce instrument count (see Roodman, 2009*a*). LSDV suffers from Nickell (1981) bias. LSDVc refers to the bias corrected LSDV estimation using the procedure from Bruno (2005*a*). AR(1) and AR(2) are Arellano-Bond test for serial correlation. Hansens J tests the null hypothesis for violation of over-identification restrictions. For all tests p-values are reported. Standard errors are reported in parentheses. For column (1) and (2) robust standard errors are reported using the Windmeijer (2005) procedure. For column (3) and (4) HAC standard errors are reported clustered on country-level. Significance of estimates in column (5) and (6) is derived by bootstrapping the variance-covariance matrix using 50 repetitions. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4: Robustness Domestic Productivity and Foreign Education (static model)

	(1) FE	(2) FE	(3) FE with AR(1)	(4) FE with AR(1)	(5) FE with Drisc.-Kraay SE	(6) FE with Drisc.-Kraay SE
Log Fstud$_{t-1}$	0.0218***	0.0187***	0.00988***	0.0116***	0.0218***	0.0187***
	(0.00593)	(0.00665)	(0.00367)	(0.00375)	(0.00682)	(0.00558)
Log Age 20-29$_{t-1}$	0.0174	0.0152	0.00926	0.0137	0.0174*	0.0152
	(0.0117)	(0.0103)	(0.00985)	(0.0118)	(0.00965)	(0.0108)
HumanCap$_{t-1}$		-0.00660		0.00400		-0.00660
		(0.0247)		(0.0156)		(0.0122)
Imports$_{t-1}$		0.00239		-0.00809***		0.00239
		(0.00455)		(0.00219)		(0.00290)
Constant	-0.392**	-0.380***	-0.188*	-0.108		
	(0.177)	(0.137)	(0.102)	(0.120)		
Observations	2000	1785	1898	1692	2000	1785
Adjusted R^2	0.152	0.148				

Notes: Regression 1, 2, 5, and 6 include time and country fixed effects. Regressions 3 and 4 only include country fixed effects. The maximum lag length chosen to correct for serial correlation in regressions 5 and 6 is 2. Standard errors in regression 1 and 2 are clustered on country level and are therefore heteroskedasticity and autocorrelation consistent. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Robustness Domestic Productivity and Foreign Education (static model sample split)

	FE		FE with AR(1)		FE with Drisc.-Kraay SE	
	Developing (1)	High (2)	Developing (3)	High (4)	Developing (5)	High (6)
Log Fstud$_{t-1}$	0.0175**	-0.000605	0.0108**	0.00472	0.0175***	-0.000605
	(0.00709)	(0.00671)	(0.00446)	(0.00639)	(0.00585)	(0.00429)
Log Age 20-29$_{t-1}$	0.00591	-0.0227	0.0119	-0.0254	0.00591	-0.0227**
	(0.0154)	(0.0146)	(0.0152)	(0.0164)	(0.0157)	(0.00940)
HumanCap$_{t-1}$	-0.0155	0.00536	0.00775	-0.0117	-0.0155	0.00536
	(0.0282)	(0.0320)	(0.0199)	(0.0220)	(0.0133)	(0.0237)
Imports$_{t-1}$	0.00263	-0.00855**	-0.00646**	-0.0233***	0.00263	-0.00855**
	(0.00438)	(0.00342)	(0.00256)	(0.00406)	(0.00319)	(0.00328)
Constant	-0.223	0.511*	-0.118	0.882***		
	(0.222)	(0.249)	(0.152)	(0.230)		
Observations	1313	472	1244	448	1313	472
Adjusted R^2	0.138	0.386				

Notes: Regression 1, 2, 5, and 6 include time and country fixed effects. Regressions 3 and 4 only include country fixed effects. The maximum lag length chosen to correct for serial correlation in regressions 5 and 6 is 2. Standard errors in regression 1 and 2 are clustered on country level and are therefore heteroskedasticity and autocorrelation consistent. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: Selected Fields of Study by Place of Origin in 2010

Rank	Place of Origin	Total	Business	Engin.	Life Sci.	Math*	Social Sci.	Educ.
1	China	127628	31014	25781	16081	13656	8551	2425
			24%	20%	13%	11%	7%	2%
2	India	104897	16049	40700	10699	20770	3147	734
			15%	39%	10%	20%	3%	1%
3	South Korea	72153	12266	9091	5484	3752	7215	2814
			17%	13%	8%	5%	10%	4%
4	Canada	28145	4222	1998	2083	704	3321	3377
			15%	7%	7%	3%	12%	12%
5	Japan	24842	5192	1093	1416	720	3279	845
			21%	4%	6%	3%	13%	3%
6	Mexico	13450	2905	2233	888	605	1264	471
			22%	17%	7%	5%	9%	4%
7	Turkey	12397	2306	2889	979	1240	1562	459
			19%	23%	8%	10%	13%	4%
8	Germany	9548	2358	792	716	334	1117	172
			25%	8%	8%	4%	12%	2%
9	United Kingdom	8861	1737	452	718	248	1382	416
			20%	5%	8%	3%	16%	5%
10	Brazil	8786	2486	729	536	290	817	264
			28%	8%	6%	3%	9%	3%
11	Thailand	8531	2269	1604	606	708	597	264
			27%	19%	7%	8%	7%	3%
12	Hong Kong	8034	2563	755	595	394	1109	96
			32%	9%	7%	5%	14%	1%
13	France	7716	2261	1034	509	255	594	93
			29%	13%	7%	3%	8%	1%
14	Indonesia	6943	2548	1305	368	437	437	174
			37%	19%	5%	6%	6%	3%
15	Malaysia	6190	1331	1758	650	371	495	186
			22%	28%	11%	6%	8%	3%
16	Kenya	5384	926	571	625	264	431	248
			17%	11%	12%	5%	8%	5%
17	Pakistan	5222	1279	1212	319	559	496	141
			25%	23%	6%	11%	10%	3%
18	Venezuela	4958	431	788	317	154	302	149
			9%	16%	6%	3%	6%	3%
19	Russia	4827	1327	299	579	323	507	126
			28%	6%	12%	7%	11%	3%

Source: ?. Notes: *Includes Computer Sciences.

Table B.7: Countries Included and their Classification

Country	Classification	Country	Classification
Algeria	LM	South Korea	HM
Argentina	UM	Kyrgyzstan	LM
Armenia	LM	Latvia	HM
Australia	H	Lesotho	L
Austria	H	Luxembourg	L
Azerbaijan	LM	Macau	HM
Bahamas	H	Macedonia	LM
Bangladesh	L	Madagascar	L
Belarus	UM	Malaysia	LM
Belgium	H	Mali	L
Bolivia	LM	Malta	HM
Botswana	UM	Mauritius	LM
Brazil	UM	Mexico	HM
Brunei Darussalam	H	Moldova	KM
Bulgaria	LM	Morocco	LM
Cameroon	LM	Mozambique	L
Canada	H	Namibia	LM
Cape Verde	LM	Netherlands	H
Chile	LM	New Zealand	H
China	L	Nicaragua	L
Costa Rica	LM	Norway	H
Cuba	LM	Pakistan	L
Cyprus	H	Panama	LM
Czech Republic	LM	Paraguay	LM
Denmark	H	Peru	LM
Dominican Rep.	LM	Philippines	LM
Ecuador	LM	Poland	LM
Egypt	L	Portugal	HM
El Salvador	LM	Russia	HM
Estonia	UM	Romania	LM
Ethiopia	L	Syria	LM
Finland	H	Senegal	LM
France	H	Singapore	H
Gabon	UM	Slovenia	HM
Germany	H	South Africa	HM
Greece	UM	Spain	H
Guatemala	LM	Sudan	L
Guinea	L	Swaziland	LM
Honduras	L	Sweden	H
Hong Kong	H	Switzerland	H
Hungary	HM	Tanzania	L
Iceland	H	Tajikistan	LM
India	L	Thailand	LM
Indonesia	L	Tunisia	LM
Iran	LM	Turkey	LM
Ireland	H	Uganda	L
Italy	H	Ukraine	LM
Japan	H	United Kingdom	H
Jordan	LM	Uruguay	UM
Kazakhstan	LM	Venezuela	UM
Kenya	L	Zambia	L

Notes: Classification refers to income groups (*Low* (L), *Lower Middle* (LM), *Upper Middle* (UM), and *High* (H)) defined by the *World Bank* in 1991 or the first year data is available. The group of developing countries combines income groups L, LM, and UM. The table lists all 102 countries used in regressions. The U.S. is excluded from this table.

C The Scrubber Rip-Off. Regulation-Based Price Discrimination: Evidence from the Acid Rain Program

C.1 Appendix C.A

The data of the EIA-767 form is partially incomprehensive and several additional information had to be collected. Most of the additional information comes from the EIA-860 form or directly from the power plants. In some cases additional information was extracted from the footnote data set to the EIA-767 and manually added to the main data set. Footnote data is only available for waves 2001 to 2005 but it is backward-looking. We primarily use data from the 2005 footnote data sheet. Stata do-files and raw data used for the construction of the final data set – including corrections – can be obtained from the authors on request. To get idea on the magnitude a short overview follows.

Data provided in footnotes or reported in EIA-860 has been used to identify FGD manufacturer in about a dozen of chases. For some chases in some waves the inservice year of the FGD unit was reported as 2099. Cross checking with other waves revealed the true year of inservice and the data was corrected accordingly. In two cases the FGD ID changed but scrubber characteristics remained unchanged. Both FGD units are located at the same plant (ID 1250) and the change coincided with the installation of a new scrubber. The new FGD unit has different characteristics. As it is important that FGD IDs remain constant, we correct the data such that the new unit gets the new ID while the old IDs remain with the old FGD units. We identified one naming mistake in the data. One plant listed in Table A of Title IV of the 1990 CAAA referred to "JACK MCDONOUGH" was not found in the data. Research revealed that the plant with ID 710 should have this name. It was corrected accordingly. Finally, cross checking with other waves revealed that there are mistakes in the measurement type of SO_2 regulation in five chases. Type of regulation changes in several waves to another measure (and back) but without a corresponding change in emission stringency. This makes little sense. For instance, there was a change in measurement type from lbs/mBtu per heat input (classified "DP") to parts per million (ppm) of SO_2 in flue gas (classified "DM"). However, the emission stringency remained at 1.2 a value that makes sense for an emission standard type of lbs/mBtu of heat input but not ppm of SO_2 in flue gas. We corrected those obvious temporary coding mistakes in the respective waves.

C.2 Appendix C.B

In most settings it is essential to control for covariates in order to correctly identify a treatment effect. The appropriate selection of control variables is therefore of upmost importance. However, in many cases observed controls are incomplete proxies for the true omitted variables. To address this problem many researchers conduct sensitivity analyzes of the treatment effects to the inclusion of observable controls. If the coefficient on the treatment variable remains stable to the inclusion of controls – so the common argument – the bias from unobservables is neglectable. Altonji, Elder and Taber (2005) and Oster (2017) describe how this approach can be misleading because omitted virable bias is proportional to coefficient movements only if the movements are scaled by the change in R^2 when controls are included. Altonji, Elder and Taber (2005) suggest that the relationship between treatment and unobservables can be recovered from the relationship between treatment and obeservables. Given this relationship it can be calculated how important the unobservables, relative to the included observables, would have to be to eliminate the treatment effect. They assume that given the full set of unobservables, the outcome variance would be fully explained, i.e. the regression would have a R^2 of 1. This my understate the robustness of results if there are e.g. measurement errors in the outcome. Oster (2017) inter alia provides a test that allows to vary the degree of relative selection (δ) and how much of the outcome variance can actually be explained (R_{max}).

In her paper Emily Oster provides empirical evidence on which values for R_{max} and δ should be assumed and how those assumption can be used to test the robustness of coefficients. First, one can vary the relative importance of unobservables. Assuming $\delta = 0$ is equivalent to assuming that there is no omitted variable bias. As in this case R_max does not influence the result, the coefficient is equal to OLS. In our case it is the upper bound of the treatment effect conditional on all observable controls. To find the lower bound of the coefficient on the treatment, one has to decide on an upper bound of δ and R_{max}. There is a clear suggestion in Altonji, Elder and Taber (2005) and Oster (2017) that δ should be bounded between [0,1]. For instance, choosing δ to be bounded by 1 is similar to assuming that observables are at least as important as unobservables. This seems to be a reasonable assumption and Oster (2017) provides empirical evidence in favor of the assumption. Therefore, given a certain R_{max} one the true bias adjusted coefficient should lie between the bounds determined by δ [0,1]. In our case it will be especially of interest if $\beta = 0$ lies within in this range. However, the lower bound of the bias adjusted coefficient range depends on our assumption on R_{max}. Oster (2017) argues that R_{max} should be a function of the R^2 from the regression ($\tilde{\mathrm{R}}$) with controls. Defining $R_{max} = \Pi\ \tilde{\mathrm{R}}$ she provides empirical evidence which cutoffs (Π) provide a reasonable range for the true coefficient. As mentioned earlier assuming a R_{max} of 1 is unrealistic in many empirical work in economics. Given the survey nature of our data we also argue that measurement errors in the outcome variable do not allow to assume a R_{max} of 1.

Table C.1 displays bias adjusted coefficients for varying assumption on R_{max} for coefficients on *TableA* and *Regulation*. Adding controls to the baseline decreases the treatment effect, i.e. drives the coefficient towards zero. In column (3) to (5) the stability of the coefficient is tested

Table C.1: Coefficient stability with varying R_{max}

	(1)	(2)	(3)	(4)	(5)
Treatment	Baseline	Controlled	Varying $R_{max}, \delta = 1$		
variable	(Std. error), [R^2]	(Std. error), [R^2]	$R_{max} = \min(2\tilde{R};1)$	$R_{max} = \min(1.5\tilde{R};1)$	$R_{max} = \min(1.3\tilde{R};1)$
TableA	44952.42	35877.85	-101212.14	15034.42	28971.38
ProxyMAC$_{Coal}$	207.73	359.80	1266.48	618.26	456.55
	Small sample without Table-A				
RegNonTableA	-1647.25	-2344.76	-5585.81	-3404.77	-2741.14

Notes: The table reports differences between the groups Table-A and Non-Table-A in columns (1) and (2). Column (3) is different, as now differences between non-Table-A installed after 1990 and all other scrubber installations are reported. The estimated differences are decomposed using the Oaxaca-Blinder procedure and have been carried out using Jann's (2008) Stata command. Regressions include the covariates included in specification VI of tables 15 or 16. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

using the method and cutoffs proposed by Oster (2017). The biased adjusted coefficients on TableA remain above 0 for all but the most highest value of R_{max}. However, assuming $R_{max} = 1$ is not sensible because the survey nature of our data makes it unlikely that it does not suffer from measurement error. Contrary, the bias adjusted coefficients on *Regulation* are all below zero. This might not be surprising as the effect is not statistically different from zero in the controlled regression. However, the procedure seems to suggest that if at all the effect is rather negative than positive. A negative coefficient on *Regulation* would support our hypothesis.

D Miscellaneous

D.1 Executive Summary

This dissertation discusses the potentials and pitfalls of empirical economic research. Several pieces of applied research illustrate the discipline's diverse use of statistical methods as well as their applicability to different topics.

Empirical economic research uses empirical evidence to test hypotheses and statistical inference to uncover general rules. However, very often several rules or causal mechanisms exist that can equally well explain the investigated outcome. This can be problematic whenever statistical inference does not yield convincing results, i.e. the degree of the study's internal validity is low. Although many different statistical tools and techniques have been developed to increase the degree of internal validity, in practice it remains difficult to claim causality. One reason for this is that the quality of statistical inference depends on the appropriateness of the chosen statistical method. Another reason is that causality requires that competing alternative explanations for an estimated statistical relationship are addressed and at best can be dismissed. Thus, an empirical study's overall quality depends critically on an author's judgement and knowledge of the environment in which the outcome is nested. Given the importance of personal perception it is not surprising that the validity of results in many studies in empirical economics is heatedly discussed in- and outside the community.

At the beginning of this dissertation the current status of this academic debate is reproduced, leading to the conclusion that there is no panacea for causal inference (section 1). Instead, it is proposed that several equally sensible strategies to strengthen causality exist and that their selection depends on the specific research question and setting. While this still allows the author to base decisions on personal perception it also stresses that justifications are required. Hence, each new study demands a tailored research agenda in which the choice of statistical methods and the existence of alternative explanations are transparently discussed. Subsequently to the discussion, three independent papers illustrate that there is indeed no blueprint procedure to conduct empirical economic research.

The first paper addresses the question whether individuals react to natural disasters by adjusting their saving behavior (section 2). The study applies statistical tools commonly used in applied microeconomic research. The research design uses quasi-experimental variation in a panel survey to infer a causal relationship between flooding and saving behavior. The study finds that from the flooding affected individuals save less in subsequent years. While the study's internal validity is rather high, the generalizability of the relationship remains to be seen. Several alternative explanations for the observed behavior are discussed and evaluated. The concluding explanation is that unusually high amounts of post-disaster financial aid induces moral-hazard-behavior. Thus, the paper makes a case for policy makers to carefully design post-disaster aid payments so as to minimize the possibility of detrimental reductions in individual precautionary efforts.

The second paper investigates the link between foreign education and domestic productivity (section 3). The paper uses aggregate data, thus encountering statistical challenges commonly occurring in applied macroeconomic research. The research design focuses on the dynamic structure of the data. The paper finds that the more students a country sends to the U.S., the higher subsequent domestic productivity growth rates will be. Additional analyses show that this effect is driven by developing countries. It is argued that the relationship is causal because foreign students transfer productivity enhancing skills from the U.S. to their home country. However, the data does not reveal whether foreign students indeed return and therefore causal inference is weaker than it could be otherwise. Measures to overcome this shortage are presented and applied. Nonetheless, the extent of the data allows for a certain generalizability of the results. In conclusion, the study suggests that foreign education poses a viable additional strategy for economic development.

Finally, the third paper addresses a research question from the field of empirical industrial organization (section 4). Specifically, the paper tests whether prices for an abatement technology are influenced by the type of environmental regulation of polluting sources. In order to test this relationship, the paper's research design combines a structural economic model with quasi-experimental empirical evidence. The paper finds that the price of abatement technology is significantly higher for those polluting sources that are participating in a permit trading scheme. Causal inference relies on the quasi-experimental nature of the data and the theoretical derivations from the structural model. However, it remains empirically challenging to exclude alternative explanations as doing so considerably strains the scope of our data. In the end, the study's results should caution policy makers to consider that regulatory instruments can have unintended side-effects hampering the diffusion and adoption of abatement technology by increasing its price.

The final section discusses the role of empirical research in the overall process of scientific progress (section 5). The importance of diverse and comprehensive empirical economic research is emphasized. Finally, it is concluded that empirical research with all its outgrowth is essential to establish something like an objective "truth" in economic science.

D.2 Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit den Möglichkeiten und Grenzen der empirischen Wirtschaftsforschung. Anhand mehrerer Forschungsarbeiten verdeutlicht sie die Anwendbarkeit statistischer Verfahren auf verschiedene Fragestellungen aus den Wirtschaftswissenschaften.

In der empirischen Wirtschaftsforschung werden Beobachtungen statistisch ausgewertet, um Hypothesen zu testen und allgemeine Regeln aufzudecken. Die Herleitung eines kausalen Zusammenhangs zwischen zwei Ereignissen gilt dabei als ein wichtiges Ziel. In der Praxis erweisen sich kausale Schlussfolgerungen allerdings als überaus schwierig. Ein Grund hierfür ist, dass der Gegenstand empirischer Wirtschaftsforschung - unsere Gesellschaft - ein komplexes und dynamisches System ist. Eine allgemeingültige Blaupause, mithilfe welcher sich kausale Zusammenhänge belegen lassen, lässt sich daher kaum entwickeln. Vielmehr hängt die Herleitung eines kausalen Zusammenhangs von den konkreten Faktoren des Einzelfalls ab. Solche Faktoren sind die Forschungsfrage, die Qualität der Daten und das Umfeld, in welchem diese erhoben wurden. Sie bestimmen anschließend das Forschungsdesign und die Auswahl eines geeigneten statistischen Verfahrens. Eine empirische Studie ist somit in vielerlei Hinsichten einzigartig. Daher ist es auch nicht verwunderlich, dass Kausalitätsbehauptungen in Bezug auf empirische Ergebnisse in den Wirtschaftswissenschaftenhäufig kontrovers diskutiert werden.

Zu Beginn dieser Dissertation wird eine aktuelle Diskussion zur Herleitung von Kausalität in der empirischen Wirtschaftsforschung wiedergegeben. Aus dieser Diskussion geht hervor, dass es momentan kein Allheilmittel für kausale Inferenz gibt, sondern stattdessen mehrere gleichermaßen sinnvolle Strategien für die Herleitung von Kausalität existieren. An die Wiedergabe und Auswertung der Diskussion schließt sich die Darstellung dreier unabhängiger empirischer Studien an. Jede dieser Studien befasst sich mit einem anderen Themengebiet der Wirtschaftswissenschaften, wobei Forschungsdesign, Auswahl der empirischen Methoden und die Art der kausalen Herleitung variieren. Die drei Studien illustrieren somit mehrere Punkte, die sich aus der Diskussion in der Einleitung ergeben haben.

Der erste empirische Beitrag in dieser Dissertation geht der Frage nach, ob Opfer von Naturkatastrophen im Anschluss an ihre Erlebnisse ihr Sparverhalten verändern (Abschnitt 2). In der Studie werden statistische Methoden verwendet, die üblicherweise in der angewandten mikroökonomischen Forschung verwendet werden. Das Forschungsdesign nutzt die durch eine Flut generierte quasi-experimentelle Variation in einer Panelbefragung aus, um einen Kausalzusammenhang zwischen Betroffenheit und Sparverhalten abzuleiten. Die Studie zeigt, dass von den Überschwemmungen Betroffene in den Folgejahren weniger sparen. Während der Grad der kausalen Schlussfolgerung hoch ist, bleibt es abzuwarten, ob sich der Zusammenhang auf andere Situationen übertragen lässt. Es werden mehrere Gründe für das beobachtete Verhalten diskutiert und gegeneinander abgewogen. Die Studie kommt zu dem Ergebnis, dass ungewöhnlich hohe Hilfszahlungen zu einem sogenannten “Moral Hazard-Verhalten, also einem verantwortungslosen Verhalten aufgrund von Fehlanreizen, führen können. In dem untersuchten Fall haben die Hilfszahlungen zu einer Verringerung im Vorsorgeverhalten bei Betroffenen geführt. Die Studie plädiert daher dafür, dass politische Entscheidungsträger etwaige Hilfszahlungen nach einer Katastrophe sorgfältig planen, um einen nachteiligen Einfluss auf individuelle Vorsorgemaßnahmen zu vermeiden.

In der zweiten Arbeit wird die Verbindung zwischen einem Studium im Ausland und heimischer Produktivität untersucht (Abschnitt 3). Das Papier verwendet dafür aggregierte Daten und befasst sich aus ökonometrischer Sicht mit bestimmten statistischen Herausforderungen, die häufig in der angewandten makroökonomischen Forschung auftreten. Das Forschungsdesign fokussiert sich auf die dynamische Struktur der Daten, um einen kausale Herleitung zu ermöglichen. Gezeigt wird, dass die Anzahl von Studenten, die ein Land in die USA schickt, sich positiv auf die Produktivitätszuwächse dieses Landes in den Folgejahren auswirkt. Des Weiteren kann gezeigt werden, dass dieser positive Zusammenhang nur für Entwicklungsländer gilt. Dies erscheint plausibel, da insbesondere die Bevölkerung aus Entwicklungsländern durch den Transfer produktivitätssteigernder Fähigkeiten aus den USA in ihr Heimatland profitieren sollte. Die Ergebnisse deuten daher an, dass es tatsächlich einen positiven Kausalzusammenhang zwischen Auslandsstudium und heimischer Produktivität gibt. Da die Daten jedoch nicht darüber informieren, ob ausländische Studenten wirklich zurückkehren, ist die kausale Inferenz schwächer, als sie es sonst sein könnte. Hingegen erlaubt der Umfang der Daten eine gewisse Generalisierbarkeit der Ergebnisse. Zusammenfassend lässt sich festhalten, dass die Förderung eines Auslandsstudiums eine sinnvolle zusätzliche Strategie für eine erfolgreiche internationale Entwicklungszusammenarbeit darstellen kann.

Das dritte und letzte Papier befasst sich mit einer Forschungsfrage, die dem Bereich der empirischen Industrieökonomik zugeordnet werden kann (Abschnitt 4). Darin wird untersucht, ob die Regulierung von Schwefeldioxidemissionen von Kohlekraftwerken die Preissetzungsstrategie von Herstellern einer geeigneten Vermeidungstechnologie beeinflusst. Um diese Beziehung zu testen, nutzt die Studie ein Forschungsdesign, das auf einem Strukturmodell und Daten mit quasi-experimenteller Variation basiert. Die Ergebnisse der empirischen Untersuchung zeigen, dass die Preise für die Vermeidungstechnologie höher sind, wenn ein Kohlekraftwerk an einem Emissionshandelssystem teilnehmen muss. Diese Entwicklung ist kontraproduktiv, da es den Anreizen eines Emissionshandelssystems, die Verbreitung von Vermeidungstechnologien zu fördern, entgegen wirkt. Die Herleitung eines kausalen Zusammenhangs beruht auf dem quasi-experimentellen Charakter der Daten sowie einem theoretischen Modell, welches den empirischen Befund ebenfalls vorhersagt. Die wichtigsten alternativen Erklärungen für das Ergebnis können ausgeschlossen werden. Einschränkend wirkt hierbei jedoch der Umfang der Daten. Dieser lässt eine rigorose Untersuchung alternativer Erklärungen nur begrenzt zu und schwächt somit den kausalen Zusammenhang etwas ab. Am Ende unterstreichen die Ergebnisse der Studie allerdings, dass politische Entscheidungsträger bei der Ausgestaltung regulatorischer Instrumente umfassend auf unbeabsichtigte Nebenwirkungen achten sollten.

Im letzten Abschnitt der Dissertation wird die Rolle empirischer Forschung im Gesamtprozess des wissenschaftlichen Fortschritts diskutiert (Abschnitt 5). Dabei wird die Bedeutung einer umfangreichen und vielfältigen empirischen Wirtschaftsforschung hervorgehoben. Abschließend wird festgestellt, dass die empirische Forschung mit all ihren Ergebnissen und Methoden notwendig ist, um eine objektive "Wahrheit" in der Wirtschaftswissenschaft zu generieren.