THE BONN JOURNAL OF ECONOMICS

Volume V(2) December 2016

Thesis

Corinna Hartung
Present-Biased Preferences and the Health-Education Gradient

Leonie Hug
The Real Effects of Hedge Fund Activism

Nicolas Kaufung
Spillover effects of financial deregulation on income and income characteristics

Lukas Nüse
Sovereign debt reduction through privatization: The case of Portugal

Adrian Waltenrath
LASSO-Based Forecasting of Financial Time Series on the Basis of News Headlines

Department of Economics
University of Bonn

Legal Information

Publisher	The Bonn Journal of Economics Adenauerallee 24-42 53113 Bonn http://bje.uni-bonn.de bje@uni-bonn.de
V. i. S. d. P.	JProf. Tobias Berg University of Bonn
Place of Publication	Bonn, Germany
ISSN	2195-3449

The Bonn Journal of Economics

Founded in 2012

VOLUME 5
ISSUE 2

Department of Economics
University of Bonn

The Bonn Journal of Economics

Contents

Present-Biased Preferences and the Health-Education Gradient
Corinna Hartung 1

The Real Effects of Hedge Fund Activism
Leonie Hug 17

Spillover effects of financial deregulation on income and income characteristics
Nicolas Kaufung 30

Sovereign debt reduction through privatization: The case of Portugal
Lukas Nüse 47

LASSO-Based Forecasting of Financial Time Series on the Basis of News Headlines
Adrian Waltenrath 58

Present-Biased Preferences and the Health-Education Gradient

Corinna Hartung*

1 Introduction

The positive and persistent relationship between education and health is well-established and observed across different countries, time periods, and health measures. A large body of literature documents that well-educated individuals are prone to live longer and enjoy better health than less educated ones (e.g. Grossman 2006; D. Cutler and Lleras-Muney 2008). This phenomenon is referred to as the health-education gradient stressing the gradual link between both factors (Deaton 2002). The potential pathways explaining this link, however, are still subject of an ongoing debate. Three mutually non-exclusive pathways are proposed in the literature: First, education might improve health (Grossman 2006). Second, health might affect educational attainment (J. Currie and Stabile 2003). Finally, an unobserved, third variable might determine education and health, and thus create an indirect link between both. Apart from genes, personality traits, and early childhood investments, individual patience is often mentioned as a potential candidate for such a latent factor (e.g. Fuchs 1982; Ippolito 2003). This study investigates whether part of the correlation between health and education can be explained by heterogeneity in patience. Patience is defined by how much an individual values future outcomes. As education and also health-investment decisions involve a tradeoff among costs and benefits at different points of time, patience (i.e. the relative valuation of outcomes at different points of time) is often assumed to influence both education and investments in health. If this holds true empirically, at least part of the link between health and education might be traced back to heterogeneous time preferences.[1]

The analysis is based on two data sets comprising 1,503 and 500 individuals representative for the adult population living in Germany. Both data sets provide information on individual time preferences elicited in paid choice experiments, a

*Corinna Hartung received her degree (M.Sc.) from the University of Bonn in 2016. The present article refers to her master thesis under supervision of JProf. Dr. Pia Pinger Corinna.Hartung@econ.lmu.de.

[1]In this study time preferences are used to quantify the individual extent of patience.

wide range of health outcomes and health-related behavior, education, as well as a broad set of background variables. The association between time preferences and the health-education gradient has been investigated in several previous studies (e.g. Fuchs 1982; D. M. Cutler and Lleras-Muney 2010; Ippolito 2003; van der Pol 2011). These studies, however, suffer from a variety of drawbacks concerning either the methods to assess time preferences, e.g. they rely on non-incentivized questions, (Fuchs 1982; D. M. Cutler and Lleras-Muney 2010; Ippolito 2003) or limited availability of background variables such as cognitive ability or personality traits (van der Pol 2011). The present study intends to fill this gap.

The analysis consists of three steps: First, the concept of the β-δ quasi-hyperbolic discounting model is applied to estimate individual-specific time preference parameters. The quasi-hyperbolic model takes into account that people discount differently in the short and long-run (Phelps and Pollak 1968; Laibson 1997; O'Donoghue and Rabin 1999). I investigate whether individuals exhibit present-biased (i.e. have a taste for immediate gratification), future-biased (i.e. overvalue future rewards) or time-consistent preferences. Second, I observe whether health and education are each in isolation related to time preferences. I find that future-biased preferences as well as the long-run discount factor δ are significantly related to some health measures, while present-biased preferences appear to play at most a minor role in predicting health. A similar pattern emerges when investigating the relationship between education and time preferences. Third, the explanatory power of time preferences on the link between health and education is estimated. The results indicate that less than 10% of the education-health gradient can be traced back to time preferences. The findings are robust to different sets of control variables and alternative methodologies.

The remainder of this study is structured as follows. The subsequent section 2 presents the theoretical framework. Section 3 describes the data. In section 4 the empirical approach is introduced. Section 5 presents the results. Various methodological limitations are discussed in section 6. Section 7 concludes.

2 Theoretical Considerations

Grossman (1972) and Ben-Porath (1967) have constructed theoretical frameworks that can be used to explain heterogeneity in the investment level in health and education, respectively. In both models individuals face a trade-off between costs and benefits across different time periods. In Grossman‘s model individuals are assumed to choose between current health investment costs and future rewards derived from a better health status (Sloan and Hsieh 2012). In the same vein, in Ben-Porath‘s model the acquisition of education is assumed to be costly in terms of e.g. material resources and opportunity costs of forgone wages, but to generate future profit in terms of higher income (Cahuc and Zylberberg 2004). As both models comprise decisions over time periods, discounting behavior is crucial in determining the optimal individual level of investment. Grossman (1972) and Ben-Porath (1967) assume that individuals discount at an exponential rate, while in the present study the concept of quasi-hyperbolic discounting is applied allowing for time-inconsistent behavior. Hence, starting with the equilibrium conditions I introduce β-δ preferences in the models and derive predictions for choice behavior which will be tested thereafter.[2]

Grossman's model In equilibrium the optimal investment level is determined when marginal returns (LHS) and marginal costs (RHS) of one additional unit of investment in each period t are equal.

$$\underbrace{\gamma_t = \frac{W_t}{\pi_{t-1}} \frac{\partial h}{\partial H}}_{\text{marginal returns}} = \underbrace{r + \rho_t - \tilde{\pi}_{t-1}}_{\text{marginal costs}} \tag{1}$$

The health stock $\mathrm{H_t}$ yields a flow of healthy days given by the concave function $\mathrm{h_t}$=f($\mathrm{H_t}$). Marginal returns are defined by the additional income $\mathrm{W_t}$ earned due to more time spent in good health $\mathrm{h_t}$ adjusted by the marginal costs of health investment $\pi_{\text{t-1}}$. Marginal costs are determined by the discount rate r, the depreciation rate ρ and changes in prices for health investment $\tilde{\pi}_{t-1}$.[3] Ceteris paribus, an increase in the discount rate r results in an increase in marginal costs and thus lowers the optimal demand for health $\mathrm{H_t}$.

Ben-Porath's model Individuals are assumed to choose the optimal schooling length x^* such that life time income is maximized given their ability θ, discount rate r and the period of retirement T. Optimal schooling length is given by:

$$x^*(\theta) = \frac{1}{r}\, ln\left(\frac{\theta - r}{\theta}\right) + T \tag{2}$$

The first order condition predicts that an increase in the discount rate r results in a decline in the optimal schooling length.

$$\frac{\partial x^*}{\partial r} = -\frac{1}{r^2}\, ln\left(\frac{\theta - r}{\theta}\right) - \frac{1}{\theta - r}\frac{1}{r} < 0 \tag{3}$$

[2]For the complete models see Grossman (1972) and Cahuc and Zylberberg (2004).
[3]This represents the pure investment model by Grossman (1972) neglecting psychological costs of health investment.

Hypotheses The relationship between the discount rate r and the quasi-hyperbolic discount factors β and δ is defined by equation (4)

$$\beta\delta = \frac{1}{1+r} \quad \Leftrightarrow \quad \frac{1}{\beta\delta} - 1 = r \tag{4}$$

$$\frac{\partial r}{\partial \beta} = -\frac{\delta}{(\beta\delta)^2} < 0 \quad \text{and} \quad \frac{\partial r}{\partial \delta} = -\frac{\beta}{(\beta\delta)^2} < 0 \tag{5}$$

From this, the following three hypotheses can be derived: (1.) Present-biased individuals demand a lower health capital stock and less education than time-consistent individuals given the same value for δ. (2.) Future-biased individuals demand a higher health capital stock and more education than time-consistent individuals given the same value for δ (3.) Individuals that are more patient in the long-run (large δ) demand a higher health stock and more education than less patient ones given the same value for β.

3 Data

My analysis uses two data sets. The first database involves a sub-sample of participants from the German Socio-Economic Panel (SOEP). The SOEP is a representative longitudinal survey that started in 1984 and provides detailed information about the socioeconomic status of approximately 22,000 individuals living in 12,000 households in Germany (for more details see Wagner, Frick, and Schupp 2007). In the 2006-wave, 1,503 individuals took part in a paid time preference experiment. This sample will be denoted as SOEP. The second data set was collected by the SOEP administration as part of the annual "pretesting" in 2005 and is denoted as PRE. In contrast to the SOEP data set, the PRE data set is cross-sectional and contains a separate sample. After completing the standard SOEP questionnaire, 500 individuals took part in a paid time preference experiment. Both samples, the SOEP and PRE sample, are drawn so as to be representative of the adult population in Germany.

3.1 Experimental Measure of Time Preferences

Experimental Procedure Time preferences are elicited in paid experiments consisting of binary choice tables. In each row, subjects are asked to decide between a smaller monetary amount X at time point t and a delayed, but larger amount of money Y at time t+τ. X is kept constant throughout each choice list, whereas Y increases from row to row. The magnitude of the increase in Y is determined by a choice list specific implied interest rate, which increases from one row to the next. Before the experiment starts, participants learn about the payment procedure ensuring that individuals have an incentive to reveal their true preferences. The switching point from the early to the delayed payment determines the individual‘s time preferences. Individuals who switch to the delayed payment at a larger Y value the future less than individuals switching earlier i.e. they are assumed to be less patient.

Treatments The experimental treatments differ in terms of the magnitude of stakes, the time horizon of payment and also the implied interest rates. Table 1.2 gives an overview of sub-samples and details of choice lists. The SOEP sample is divided into three sub-groups (S1-S3). In the SOEP, individuals are asked to complete two different choice lists, while in the PRE sample subjects completed three choice lists. The treatments comprise different experimental designs, which follow the notation in Dohmen et al. (2012). In the so-called "overlapping" design (OD), the time intervals between early and delayed payment differ across choice lists, but the payment of the early reward is provided at the same point of time in all choice lists. In a "shifted" design (SD) the time intervals between early and delayed payment are identical across choice lists, but the early payment is delivered at different points of time. A combination of shifted and overlapping design, comprising at least three choice lists, is denoted as "overlapping-shifted" design (OSD).

3.2 Survey Data

Participants in both data sets completed a detailed questionnaire. Subjects in the PRE sample were interviewed only once in 2005, while most of the participants in the SOEP study complete the survey regularly on an annual basis as they are part of the panel structure. Next to information about demographic characteristics, the questionnaire covers a broad range of topics, such as education and health.

Education Educational attainment is classified into high and low. High education is defined as having obtained a qualification to enter university, which is equivalent to "Fachabitur" or "Abitur" in the German schooling system.

Health Behavior and Outcomes Health-related behavior is measured by the following five factors: smoking habits, frequency of alcohol consumption, nutrition, physical activity and the body mass index (see Table 1.1). These parameters are evaluated both separately and aggregated using an index composed of the standardized mean of these five parameters. To assess the health status, I apply the physical and the mental component summary score, denoted as PCS and MCS. The PCS and MCS scores result from a factor analysis based on the response to the international validated SF-12 questionnaire (H. H. Andersen et al. 2007). As a second measure for general health status, I construct an indicator which takes the value one if self-assessed health status is very good or good and else zero. Additionally, I include a variable indicating whether a physician has ever diagnosed the onset of hypertension or any cardiovascular disease.

Covariates To make the samples (S1-S3, PRE) comparable I control for a large set of background variables, which are shown to determine health and educational outcomes. Next to demographic characteristics such as age, gender and the marital status, I account for regional fixed effects and differences across sub-samples including dummies for each sub-group. Additionally, I control for personality traits, risk preferences and cognitive ability.[4]

3.3 Sample Construction

The final sample used for the analysis comprises 881 individuals from the SOEP data set and 273 individuals from the PRE database. Individuals, who *always* choose the early payment in *at least one* choice list, are excluded from the analysis, because β and δ cannot be calculated for these persons (see section 4). This reduces the sample by about 40%. Furthermore, I exclude 23 individuals with missing information about their educational degree. For the SOEP data set, a cross-sectional data set based on the most recent information available is constructed. To this end, I mostly take data from the 2013-wave, but missing values are replaced with the most recent available data from the 2005-2012 waves. As described above, information on individuals in the PRE sample is cross-sectional and elicited only once in 2005.

4 Empirical Approach

Given quasi-hyperbolic time preferences, an individual i is assumed to be indifferent between a sooner smaller payment X and a larger delayed payment Y, if the utility of payment X is equal to the discounted utility of Y:

$$u(X) \;=\; \beta_i^{\mathbf{1}_{t=0}}\ \delta_i^{\tau}\ u(Y_{i,t,\tau}) \tag{6}$$

[4]Personality traits are measured by the concept of the Big Five, a taxonomy containing five superordinate dimensions, which subsume a variety of characteristics—openness to experience, conscientiousness, extraversion, agreeableness and neuroticism (Borghans et al. 2008). The dimensions are surveyed in the data sets using a short version (BFI-S) of common inventories (Dehne and Schupp 2007). Confirmatory factor analysis is applied to extract a latent factor for each dimension. Individual risk preferences are measured on a 11-point Likert-scale. Further, to proxy cognitive ability, I use two ultra-short IQ tests, which are based on the Wechsler Adult Intelligence Scale (WAIS).

where τ denotes the time horizon between payment X and payment Y. Parameter t indicates the time interval between the time of the experiment and the early payment X. The indicator function $\mathbf{1}_{t=0}$ takes the value 1 if the sooner payment X is provided immediately, otherwise zero. Payment X is kept constant throughout each choice list, while the value Y is determined endogenously by the individual switching behavior in each choice list. Parameter β and δ are identified using ordinary least squares regression analysis. To this end, I first transform equation (6) into a linear expression, by taking the logarithm. Second, I assume that the utility function is linear over the experimental outcomes.[5]

$$log(X) \ - \ log(Y_{i,t,\tau}) \ = \ log(\beta_i) \ \mathbf{1}_{t=0} \ + \ log(\delta_i) \ \tau \tag{7}$$

The experimental design does not allow to infer the exact point of indifference between the payment X and Y, but rather specifies upper and lower bounds of the time preference parameters, which are determined by (i) the highest value of Y for which an individual chooses X over Y (ii) the lowest value of Y that an individual chooses over X. To obtain point estimates for β and δ, I use the average ($\mathrm{Y_{mid}}$) between these two values of Y. It is further assumed that an individual, who switches in the first row in a choice list, weakly prefers X today over the value of X discounted τ months ($X \geq \beta\delta^{\tau}X$). This assumption is necessary to identify $\mathrm{Y_{mid}}$ for "immediate-switchers". β and δ are estimated for each individual separately using OLS regression.

$$log(X) - log(Y_{mid_{i,t,\tau}}) = log(\beta_i)\mathbf{1}_{t=0} + log(\delta_i)\tau + \epsilon_i \qquad \epsilon_i \sim \mathcal{N}(0,\sigma^2) \quad \forall i \tag{8}$$

Thereafter, I group individuals into three classes based on the value of β: present-biased, future-biased and time-consistent. An individual is considered present-biased (future-biased) if all values of the interval $\beta_i \pm 0.2$ sub-sample S.D. are smaller (larger) than one, otherwise the individual is considered time-consistent.

The returns to education are estimated first without controlling for time preferences:

$$\mathrm{H_i} \ = \ \alpha_0 \ + \ \alpha_1 \ \mathrm{E_i} \ + \ \alpha_2 \ \mathrm{X_i} + \epsilon_i \qquad \epsilon_i \sim \mathcal{N}(0,\sigma^2) \quad \forall i \tag{9}$$

where $\mathrm{H_i}$ denotes the health measure of individual i, $\mathrm{E_i}$ defines an indicator for a high or low level of education and $\mathrm{X_i}$ represents a set of control variables. The returns of education to health are identified by the coefficient of education α_1.

[5] Since the utility function is not known, one has to make identifying assumptions about the functional form of u(·). Rabin (2000) demonstrates that the assumption of linear utility/risk neutrality over small stakes outcomes is consistent with expected utility theory. However, some empirical studies find that individuals exhibit risk aversion even over small stakes (S. Andersen et al. 2008; Harrison, Lau, and Rutström 2010). Hence, I control for risk preferences in the analysis.

Equation 9 is then re-estimated including time preferences:

$$\mathrm{H_i} = \gamma_0 + \gamma_1\ \mathrm{E_i} + \gamma_2\ \delta_i +\ \gamma_3 \text{ present bias}_\mathrm{i} + \gamma_4 \text{ future bias}_\mathrm{i} + \gamma_5\ \mathrm{X_i}\ +\ \epsilon_i \tag{10}$$

the discount factor δ is a continuous variable whereas present-bias$_\mathrm{i}$ and future-bias$_\mathrm{i}$ are binary variables taking a value of 1 if individual i is classified as present-biased or future-biased, else zero. A potential effect of inter-temporal preferences on the relationship between health and education, is examined by computing the percentage change in the returns to education

$$\text{Effect}\ =\ 1\ -\ \frac{\gamma_1}{\alpha_1} \tag{11}$$

To account for different distributional properties of the outcome variables, I use the OLS regression model for continuous variables and a discrete choice probit model for binary responses.

5 Results

Estimates of Time Preferences Table 1.3 and Figure 1.1 summarize the estimates of the discount factor δ for each sub-sample. The average monthly discount factor δ ranges from 0.976 to 0.99. Three points should be emphasized: (1.) A large heterogeneity in δ is observed, indicating that individuals differ in their discounting behavior. (2.) The distributions of δ in S1, S2 and the PRE sub-sample are relatively similar, while the distribution of δ in S3 differs. If the estimation of δ in the PRE data set is based only on the data obtained in the shifted design (SD), the distribution of δ is similar to that obtained with S3, suggesting that the experimental design affects δ (3.) Contrary to theory the estimates of δ can take values greater than 1. This is discussed below. Table 1.4 presents the corresponding estimates of parameter β. The mean value ranges between 0.98 and 1.002. Standard deviation, minimum and maximum values of β differ substantially across sub-samples with a larger variance in S1 and PRE than in S2 and S3. This can be explained by different constraints imposed on the calculation of β and δ across sub-samples. Differences in the constraints come from: (i) the use of different experimental designs (OD, SD and OSD) (ii) different time horizons and (iii) the assumption that an individual, switching in the first row of a choice list, weakly prefers X today over the value of X discounted τ months ($X \geq \beta\delta^\tau X$). This assumption implies that in S2 and S3 the constraint $\beta\delta \leq 1$ is imposed, while in S1 and PRE $\beta\delta^6 \leq 1$ is applied (see Table 1.5). Since $\delta^6 < \delta^1$ if $\delta < 1$, the maximum estimate of β can take larger values in S1 and PRE than in S2 and S3 for a given δ, which is in fact observed in Table 1.4. Further, note that the values of β and δ are negatively correlated due to the estimation technique ($p_{corr} = -0.42$).

In most theoretical considerations it is assumed that $\delta \leq 1$ and $\beta\delta \leq 1$. However, previous empirical research also reports a discount factor greater 1 (Frederick, Loewenstein, and O'Donoghue 2002, p.279). Figure 1.2 illustrates that a value of $\delta > 1$ leads to the counter-intuitive result that an individual prefers a *lower* delayed payment X to a *higher* sooner payment X in the long-run. This suggests that the switching decision of individuals with $\delta > 1$ might be determined by absolute values rather than by interest rates. The data indeed supports this hypothesis revealing that the absolute value of the difference in switching amounts across choices ($|Y_{\text{choice list 1}} - Y_{\text{choice list 2}}|$) is significantly lower for individuals with $\delta > 1$ than for individuals with $\delta < 1$ ($p < 0.001$; Mann-Whitney-U-test).

Table 1.6 summarizes the fraction of individuals with present-biased, future-biased and time-consistent preferences in each sub-sample. Across sub-samples there are 36,8% present-biased and 24% future-biased individuals. Surprisingly, the fraction of present-biased subjects in S1 and S2 is about twice as large as in S3 and PRE. There are at least two explanations: (i.) the sample composition of sub-samples differs systematically (ii.) the classification is affected by the experimental design, as OD is applied to S1 and S2, whereas SD or OSD are applied to S3 and PRE, respectively. I find no systematic differences across sub-samples in terms of risk preferences or cognitive ability. In the PRE sample, the same individuals take part in an OD and SD. Hence, to test the second explanation I use the PRE sample to calculate β for both the OD and SD design. If the classification does not depend on the design, a similar fraction of present-biased, future-biased, as well as time-consistent individuals should be observed in SD and OD. The results indicate that the classification strongly depends on the experimental design e.g. the fraction of present-biased individuals is 57% with OD, but only 24% with SD (see Row 5 and 6 in Table 1.6). Note that the fractions obtained in OD and SD for the PRE sample are similar to those calculated for S1 to S3 using the corresponding design.

Time Preferences and Health Table 1.7 displays the effect of time preferences on health investments, namely on the BMI, alcohol consumption, nutrition, smoking habits, sports, and an aggregated health investment index. According to the predictions derived from Grossman's model in section 2, individuals with present-biased (future-biased) preferences are expected to invest less (more) in their health than time-consistent individuals given the same value of δ. I find that neither present-biased nor future-biased preferences do affect significantly health investments, with the exception of BMI ($p < 0.1$). The discount factor δ, however, is significantly correlated to the aggregated health investment index ($p < 0.01$), as well as to sport activities, smoking habits and eating habits. An 0.01 point increase in the value of δ, for example, reduces the probability to be a smoker by about 2 percentage points and, simultaneously, increases the probability to keep a healthy diet by around 2 percentage points. Table 1.8 presents the corresponding results for health outcomes i.e. physical (PCS) and mental health (MCS), the onset of hypertension and any cardiovascular disease (CVD), as well as self-reported health (SRH). Again, present-biased (future-biased) individuals are expected to have an adverse (better) health status compared to time-consistent individuals given the same value of δ. In contradiction to hypothesis 1 in section 2, present-biased preferences seem to favor health in terms of CVD, MCS and PCS scores. The effect, however, is only significant for CVD. In line with hypothesis 2, individuals with future-biased preferences are significantly better off in terms of the PCS score, SRH and CVD. SRH is the only parameter to which δ is positive significantly related. Overall, the findings suggest that health investment is particularly linked to the discount factor δ, while health outcomes appear to be mainly linked to future-biased preferences.

Time Preferences and Education Table 1.9 presents the effect of time preferences on education. I find that education is positive significantly associated with future-biased preferences and with the long-term discount factor δ. A 0.01 point increase in δ is linked to a 3 percentage point increase in the likelihood to acquire a school degree qualifying for university entry. In the same vein, future-bias is associated with a 9 percentage point increase in the probability to obtain an university entry qualification. However, present-biased preferences, in turn, seem not to be predictive in terms of education attainment.

Time Preferences and the Health-Education Gradient The final part of the analysis reports the association between health and education and presents the effect of time preferences on the gradient. As expected, well-educated people enjoy a better health status and invest significantly more in their health than low educated people with the exception of alcohol consumption (see Table 1.10). In line with previous reports, high educated people consume more frequently alcohol than low-educated ones (Robert Koch-Institut 2012). Table 1.11 summarizes the effect of time preferences on the gradient. Each row presents the returns to education derived from a regression using different health measures as dependent variable. As before, there are two specifications used to identify the set of covariates: A small set controlling for demographic characteristics and a large one, which includes additional measures for cognitive ability, risk preferences, and personality traits. For each of the two specifications the returns to education are estimated twice: without (Columns 1 and 5) and with (Columns 2 and 6) the inclusion of the estimates of time preferences. The effect of time preferences on the health-education gradient is reported in Columns 3 and 7. The measure denotes the percentage change in the returns to education, which can be traced back to time preferences. Overall, the results indicate that, if at all, only a small part of the correlation between education and health can be traced back to time preferences. The inclusion of inter-temporal preferences in the regression reduces the returns to education at most by 9 percent for SRH. For most of the health variables the effect is around 1 to 5 percent. For MCS, BMI and also alcohol consumption a negative change in the coefficient of education is observed. In this case time preferences can be regarded as confounding variables. Finally, I test whether the results are robust to changes in the set of covariates. One potential confounding variable is income. Income can be considered as mediating factor i.e. education affects income, but income, in turn, might determine health, since higher income facilitates the access to supplementary, self-paid health services. Another important factor is parental socioeconomic background which is known to affect health and education of the next generation (e.g. A. Currie, Shields, and Price 2007; D. M. Cutler and Lleras-Muney 2010). I find that the effect of time preferences on the health-education gradient is robust to the inclusion of income and parental education.[6]

6 Discussion

The findings suggest that less than 10% of the health-education gradient can be traced back to heterogeneous time preferences. In line with a study by van der Pol (2011) the effect is strongest for self-reported health (SRH). As SRH describes a health status it is at least partly determined by genetic factors, while health investment is under direct behavioral control. Hence, one would expect that the impact of time preferences on the association between health-related behavior and education is stronger than on the link between SRH and education. This counter-intuitive result might, however, be explained by the fact that SRH is a very broad measure of health, clearly not only affected by "objective" health measures, but also by psycho-social factors, such as social deprivation self-expectations or, personal characteristics (Eriksson, Undén, and Elofsson 2001).

One often-mentioned concern about eliciting time preferences in a paid experiment is that individuals might be skeptical about the delivery of the delayed payment and, thus, may put a premium on payments available immediately so that they appear to be less patient than they actually are. There is, however, little scope for this concern in the present study as all payments are sent by mail after the experiment, in the sense of "front-end delay" (Coller and Williams 1999). Moreover, individuals in the SOEP data set are in a long-term relationship with the agency, since they are annually interviewed by them (Dohmen et al. 2012).

An unexpected result of this study is that the classification of individuals into future-biased, present-biased and time-consistent preferences depends strongly on

[6]The estimates are not presented in this article, but can be provided on request.

the experimental design. This could be shown nicely using the PRE sample in which the same people took part in an OD, as well as SD. I find that about twice as many individuals are classified as present-biased in OD than in SD. This raises the question which experimental treatment, if any, uncovers the "true" time preferences. Only 40% of the individuals in the PRE sample, exhibit the same preferences in OD and SD. Hence, the predictive power of time preferences might be weakened due to attenuation bias if people are incorrectly classified as time-consistent or time-inconsistent (Dohmen et al. 2012).

Another important issue not clarified yet is whether individuals discount monetary and non-monetary rewards, e.g. health and educational investments, at the same rate. A recent study by Augenblick, Niederle, and Sprenger (2015) suggests that individuals discount differently in real-effort tasks and in monetary choices. They report that a larger share of individuals behaves time-inconsistent in real-effort tasks and in monetary choices and that the correlation between a measure of present-biased preferences in monetary and real effort tasks is close to zero. However, one has to be careful with the results as they are based on a small sample (N=75).

Another concern might be that different constraints are imposed on the calculation of β and δ across sub-samples. As described above, in some treatments the estimates are restricted to $\delta \leq 1$ and $\beta\delta \leq 1$, while in others less restrictive constraints are imposed on the calculation. I find that the results are not driven by differences in the constraints, by re-estimating β and δ under the condition that $\delta \leq 1$ and $\beta\delta \leq 1$ in all sub-samples using nonlinear least square estimation.[7]

7 Conclusion

The goal of the present study is to investigate whether part of the health-education gradient can be traced back to time preferences. Grossman‘s model for health demand and Ben-Porath‘s model for human capital predict that patience causes health and education to vary in the same direction and, thus might create an indirect link between both factors (Ben-Porath 1967; Fuchs 1982; Grossman 1972). Consistent with theoretical predictions I find that time preferences are related to health and education. The impact of time preferences on the gradient, however, is limited. There are at least three explanations for this finding: (i.) In fact, only a small part of the gradient can be explained by the individual extent of patience. (ii.) The experimental design used to measure the individual extent of patience by means of choice lists does not provide a reliable proxy for patience causing a biased estimate of the impact of patience on the gradient. (iii.) Omitted variables bias the results. As the results are robust to changes in the set of covariates, omitted

[7] The estimates are not presented in this article, but can be provided on request.

variables bias seems to be only a minor issue in this study. The calculation of "true" time preferences, however, seems to be rather challenging. As discussed above, I find that the experimental design affects substantially the classification of individuals into present-biased, future-biased and time-consistent preferences. The reason of this design-dependent classification is, however, not clear yet. One explanation–supported by my results–might be that individuals refer to absolute values rather than interest rates when making their switching choices. In line with that explanation, I observe that individuals tend to switch rows at prominent numbers (see Figure 1.1). Another explanation might be that individuals are sensitive to time horizon effects in a way that is difficult to explain with any standard discounting model (Dohmen et al. 2012). Overall, the study suggests the necessity to rethink the experimental design for measuring time preferences, as well as the standard discounting models. Moreover, the findings highlight the importance to get a better understanding of inter-temporal choice behavior to assess political relevant questions as the relationship between health and education.

Bibliography

Andersen, Hanfried H. et al. (2007). "Computation of standard values for physical and mental health scale scores using the SOEP version of SF-12v2". In: *Schmollers Jahrbuch* 127.1, pp. 171–182.

Andersen, Steffen et al. (2008). "Eliciting Risk and Time Preferences". In: *Econometrica* 76.3, pp. 583–618. ISSN: 0012-9682. DOI: \url{10.1111/j.1468-0262.2008.00848.x}.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger (2015). "Working Over Time: Dy-namic Inconsistency in Real Effort Tasks". In: *The Quarterly Journal of Economics* 130, pp. 1067–1115.

Ben-Porath, Yoram (1967). "The production of human capital and the life cycle of earnings". In: *The Journal of Political Economy* 75.4, Part 1, pp. 352–365.

Borghans, Lex et al. (2008). "The economics and psychology of personality traits". In: *Journal of human Resources* 43.4, pp. 972–1059.

Cahuc, Pierre and André Zylberberg (2004). *Labor economics*. MIT Press.

Coller, Maribeth and Melonie B. Williams (1999). "Eliciting individual discount rates". In: *Experimental Economics* 2.2, pp. 107–127.

Currie, Alison, Michael A. Shields, and Stephen Wheatley Price (2007). "The child health/family income gradient: Evidence from England". In: *Journal of Health Economics* 26.2, pp. 213–232. ISSN: 01676296.

Currie, Janet and Mark Stabile (2003). "Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children?" In: *American Economics Review* 93.5, pp. 1813–1823. DOI: \url{10.3386/w9098}.

Cutler, David M. and Adriana Lleras-Muney (2010). "Understanding differences in health behaviors by education". In: *Journal of Health Economics* 29.1, pp. 1–28. ISSN: 01676296. DOI: \url{10.1016/j.jhealeco.2009.10.003}.

Cutler, David and Adriana Lleras-Muney (2008). "Education and Health: Evaluating Theories and Evidence". In: *Making Americans Healthier: Social and Economic Policy as Health Policy*. Ed. by J House, R Schoeni, G Kaplan, and H Pollack. Vol. 2008. DOI: \url{10.3386/w12352}.

Deaton, Angus (2002). "Policy implications of the gradient of health and wealth". In: *Health Affairs* 21.2, pp. 13–30.

Dehne, Max and Jürgen Schupp (2007). "Persönlichkeitsmerkmale im sozio-oekonomischen panel (Soep)–Konzept, umsetzung und empirische eigenschaften". In: *Research Notes* 26, pp. 1–70.

Dohmen, Thomas et al. (2012). "Interpreting time horizon effects in inter-temporal choice". In: *IZA* DP 6385.

Eriksson, Ingeborg, Anna-Lena Undén, and Stig Elofsson (2001). "Self-rated health. Comparisons between three different measures. Results from a population study". In: *International journal of epidemiology* 30.2, pp. 326–333.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002). "Time discounting and time preference: A critical review". In: *Journal of Economic Literature* 40.2, pp. 351–401.

Fuchs, Victor (1982). "Time Preference and Health: An Exploratory Study". In: *Economic aspects of health*. Ed. by Victor R. Fuchs. University of Chicago Press. DOI: \url{10.3386/w0539}.

Grossman, Michael (1972). "On the Concept of Health Capital and the Demand for Health". In: *Journal of Political Economy* 80.2, pp. 223–255. ISSN: 00223808, 1537534X. URL: %5Curl%7Bhttp://www.jstor.org/stable/1830580%7D.

— (2006). "Education and nonmarket outcomes". In: *Handbook of the Economics of Education* 1, pp. 577–633.

Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström (2010). "Individual discount rates and smoking: Evidence from a field experiment in Denmark". In: *Journal of Health Economics* 29.5, pp. 708–717. ISSN: 01676296.

Ippolito, Richard A. (2003). "Education versus savings as explanations for better health: evidence from the Health and Retirement Survey". In: *George Mason Law & Economics Research Paper* 3-4.

Laibson, David (1997). "Golden eggs and hyperbolic discounting". In: *The Quarterly Journal of Economics* 112.2, pp. 443–477.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing it now or later". In: *American Economic Review*, pp. 103–124. ISSN: 0002-8282.

Phelps, Edmund S. and Robert A. Pollak (1968). "On second-best national saving and game-equilibrium growth". In: *The Review of Economic Studies* 35.2, pp. 185–199.

Rabin, Matthew (2000). "Risk aversion and expected-utility theory: A calibration theorem". In: *Econometrica* 68.5, pp. 1281–1292. ISSN: 0012-9682.

Robert Koch-Institut (2012). "Daten und Fakten:Ergebnisse der Studie »Gesundheit in Deutschland aktuell 2010. RKI Berlin". In:

Sloan, Frank A. and Chee-Ruey Hsieh (2012). *Health economics*. MIT Press.

van der Pol, Marjon (2011). "Health, education and time preference". In: *Health Economics* 20.8, pp. 917–929. ISSN: 10579230. DOI: \url{10.1002/hec.1655}.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007). "The German Socio-Economic Panel study (SOEP)-evolution, scope and enhancements". In:

Online Appendix to: Present-Biased Preferences and the Health-Education Gradient by Corinna Hartung

Table 1.1: Overview of Health Behavior and Outcome Variables

	Abbreviation	Type	N	Description
Health Status				
Mental Health	MCS	C	1151	scale from 0 to 100 with 100 indicating a perfect mental health status
Physical Health	PCS	C	1151	scale from 0 to 100 with 100 indicating a perfect physical health status
Self-Reported Health	SRH	B	1154	1 = good/very good health, else 0
Cardiovascular Disease	CVD	B	737	1 = onset of a stroke, apoplex or cardiopathy, else 0
Hypertension		B	737	1 = onset of hypertension, else 0
Health Investment				
Body-Mass-Index	BMI	C	1143	kg/m^2
Smoker		B	1154	1 = currently smoking, else 0
Alcohol		C	1154	a higher value indicates more frequent alcohol consumption (beer, wine and spirits); exploratory FA; z-standardized
Sports		C	1153	a higher value indicates more frequent sportive activity; average of two items in the survey, scaled to a range from 0 to 1
Nutrition		B	1153	1 = healthy diet, else 0
Health Invest. Index		C	1154	a higher value indicates a higher investment into health; standardized mean of nutrition, smoking and alcohol habits, sports and the BMI

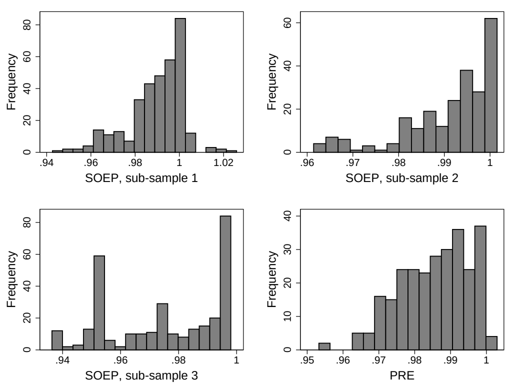
Note: Type indicates whether the variable is binary (B) or continuous (C). N denotes the number of observations. A short description of the variables is provided.

Table 1.2: Overview of Experimental Treatments

Data Set	Sample	N	IR	Design	Choice List 1	Choice List 2	Choice List 3
SOEP	S1	338	2.5%	OD	today vs. 12 months	today vs. 6 months	–
SOEP	S2	236	2.5%	OD	today vs. 12 months	today vs. 1 months	–
SOEP	S3	307	5%	SD	today vs. 1 months	12 months vs. 13 months	–
PRE		273	2.5%	OSD	today vs. 6 months	today vs. 12 months	6 months vs. 12 months

Note: N denotes the sample size. IR refers to the annual interest rate used in the first row of the choice list. Design presents the experimental approach, where OV stands for overlapping, SD for shifted and OSD for overlapping-shifted design. The time of the early and delayed payment is reported for each choice list.

Figure 1.1: Distribution of Delta



Note: The graphs show the histogram of parameter δ in each sub-sample. δ is calculated using mid point values $\mathrm{Y}_{\mathrm{mid}}$. The distributions of δ in S1, S2 and in the PRE sample are relatively similar, while the distribution of δ in S3 differs. The distribution of S3 is characterized by three spikes, that refer to switching values of Y at 20€ 0, 20€ 5 and 21€ 0. This suggests that individuals tend to switch in rows where the delayed payment surpasses prominent numbers.

Table 1.3: Estimates of Delta

Data Set	Sub-Sample	Sample Size	Mean	S.D.	Min.	Max.
SOEP	S1	338	0.99	0.012	0.943	1.026
SOEP	S2	236	0.991	0.01	0.961	1.002
SOEP	S3	307	0.976	0.02	0.936	0.998
PRE		273	0.986	0.01	0.953	1.003

Note: The table presents mean, standard deviation, minimum and maximum value of parameter δ for each sub-sample. δ is calculated using OLS and the average switching point $\mathrm{Y}_{\mathrm{mid}}$.

Table 1.4: Estimates of Beta

Data Set	Sub-Sample	Sample Size	Mean	S.D.	Min.	Max.
SOEP	S1	338	0.98	0.076	0.727	1.399
SOEP	S2	236	0.994	0.011	0.967	1.039
SOEP	S3	307	0.999	0.018	0.938	1.066
PRE		273	1.002	0.05	0.826	1.235

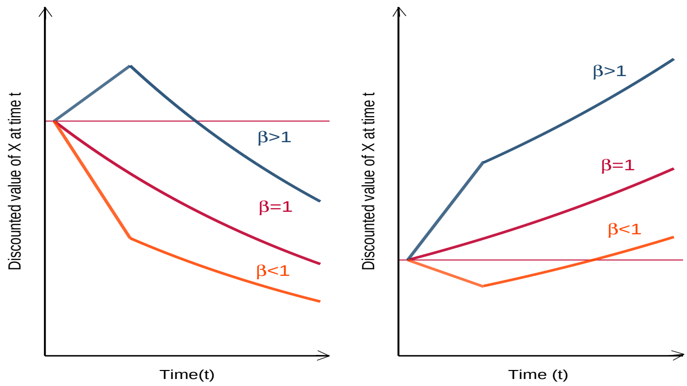
Note: The table presents mean, standard deviation, minimum and maximum value of parameter β for each sub-sample. β is calculated using OLS and the average switching point $\mathrm{Y}_{\mathrm{mid}}$.

Table 1.5: Constraints on Estimates of Time Preferences

	Sub-Samples			
Assumption	**S1**	**S2**	**S3**	**PRE**
$\delta \leq 1$	$-$	$-$	$+$	$-$
$\beta\delta \leq 1$	$-$	$+$	$+$	$-$

Note: The table presents whether the assumptions $\delta \leq 1$ and $\beta\delta \leq 1$ hold for each sub-sample. "-" denotes that the assumption is violated; "$+$" denotes that the assumption holds.

Figure 1.2: Scheme of Quasi-Hyperbolic Discounting



Note: The graph on the left hand side illustrates how a present value of X evolves over time given $\delta < 1$.The graph on the right hand side depicts the situation for $\delta > 1$. Both graphs differentiate between future-biased ($\beta > 1$), time-consistent ($\beta = 1$) and present-biased preferences ($\beta < 1$).

Table 1.6: Classification in Present-Bias, Time-Consistent and Future-Bias

Data Set	Sub-Sample	Design	Sample Size	Present-Bias	Time-Consistent	Future-Bias
SOEP	S1	OD	338	47.34%	31.36%	21.30%
SOEP	S2	OD	236	51.27%	34.32%	14.41%
SOEP	S3	SD	309	25.73%	48.21%	26.06%
PRE		OSD	273	23.81%	42.86%	33.33%
PRE		OD	273	57.51%	26.01%	16.48%
PRE		SD	273	23.81%	42.86%	33.33%

Note: The table presents the fraction of present-biased, time-consistent and future-biased individuals in each sub-sample. Additionally, beta is calculated using both OD and SD in the PRE sample. The results indicate that in the classification depends strongly on the experimental design. Note that the estimates of β are identical in the SD and OSD. This can, however, been shown by plugging in numbers into the vectors $\mathbf{1}_{t=0}$ and τ in equation 8 and solving for the OLS estimator β_{OLS}=$(X'X)^{-1}X'y$.

Table 1.7: Association: Health Investment and Time Preferences

Variables	Health Investment											
	BMI		**Alcohol**		**Nutrition**		**Smoker**		**Sports**		**Health Invest. Index**	
Present-Bias (=1)	0.504	0.637*	-0.0250	-0.0377	-0.00445	-0.0226	0.00527	0.00867	0.0248	0.0200	-0.0259	-0.0619
	(0.31)	(0.35)	(0.05)	(0.05)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.06)	(0.07)
Future-Bias (=1)	-0.183	-0.0986	-0.00249	-0.0164	-0.00244	-0.000184	0.0107	-0.000552	-0.00766	0.0186	-0.0140	0.0335
	(0.35)	(0.38)	(0.06)	(0.06)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.07)	(0.08)
Delta	-0.00512	-0.0747	0.0141	0.00550	0.0185**	0.0186**	-0.0214**	-0.0222**	0.0179**	0.0188*	0.0575***	0.0670***
	(0.10)	(0.11)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	1141	968	1152	978	1151	977	1152	978	1151	977	1152	978
Demographic Characteristics	X	X	X	X	X	X	X	X	X	X	X	X
Personality Traits, Ability & Risk	–	X	–	X	–	X	–	X	–	X	–	X
Log lik.	-3306.5	-2811.9	-1223.5	-1015.2	-385.0	-305.1	-624.4	-529.2	-574.4	-459.2	-773.2	-631.1
R-squared	0.128	0.146	0.142	0.172					0.0676	0.120	0.134	0.163

Note: OLS regression (BMI, alcohol, sports, health investment index), Probit regression (nutrition and smoker). For nutrition and smoker, coefficients reported are average marginal effects. **Dependent Variables**: BMI, frequency of alcohol consumption (z-standardized), nutrition , smoker, sports, health investment index. **Demographic Characteristics**: age, age squared, gender, marital status, region, sub-sample fixed effects. **Personality Traits, Ability and Risk**: Big 5-factors, cognitive ability scores, risk; values are standardized.

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Association: Health Status and Time Preferences

Variables	Health Status						Diagnosis			
	MCS		PCS		SRH		CVD		Hypertension	
Present-Bias (=1)	0.476	0.354	0.186	0.429	-0.0260	-0.000935	-0.0572**	-0.0613**	0.0303	0.0339
	(0.61)	(0.64)	(0.64)	(0.70)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Future-Bias (=1)	0.145	0.361	1.117*	1.606**	0.102***	0.120***	-0.0717**	-0.0681**	-0.0714	-0.0580
	(0.69)	(0.70)	(0.67)	(0.72)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)
Delta	-0.0151	-0.222	0.128	0.0715	0.0276***	0.0189*	-0.0126	-0.0117	-0.0124	-0.0123
	(0.20)	(0.21)	(0.20)	(0.22)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	1149	975	1149	975	1152	978	737	609	737	609
Demographic Characteristics	X	X	X	X	X	X	X	X	X	X
Personality Traits, Ability & Risks	–	X	–	X	–	X	–	X	–	X
Log lik.	-4118.8	-3430.1	-4114.3	-3475.6	-679.2	-545.9	-292.2	-228.6	-428.3	-342.7
R-squared	0.0202	0.120	0.292	0.302						

Note: OLS regression (MCS, PCS); Probit regression (SRH, CVD, hypertension). For SRH, CVD, hypertension, coefficients reported are average marginal effects. **Dependent Variables**: MCS, PCS, SRH, hypertension, CVD. **Demographic Characteristics**: age, age squared, gender, marital status, region, sub-sample fixed effects. **Personality Traits, Ability & Risk**: Big 5-factors, cognitive ability scores, risk preferences; values are standardized. Robust standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Association: Education and Time Preferences

	Education	
Present-Bias (=1)	0.0229	0.0328
	(0.03)	(0.03)
Future-Bias (=1)	0.0646*	0.0911**
	(0.04)	(0.04)
Delta	0.0295***	0.0274**
	(0.01)	(0.01)
Observations	1152	978
Demographic Characteristics	X	X
Personality Traits, Ability & Risks	–	X
Log lik.	-667.5	-539.3

Note: Probit Regression; average marginal effects are reported. **Dependent Variable**: education (1 = high). **Demographic Characteristics**: age, age squared, gender, marital status, region, sub-sample fixed effects. **Personality Traits, Ability & Risk**: Big 5-factors, cognitive ability scores, risk preferences. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: Health-Education Gradient

Variables	**Mean and (sd)**				**P-Value**
	Low Education		High Education		
Health Outcome					
PCS	47.26	(10.3)	51.84	(9.67)	0.000
MCS	51.62	(9.13)	52.9	(7.94)	0.095
SRH	0.41	(0.49)	0.61	(0.49)	0.000
CVD	0.21	(0.41)	0.12	(0.32)	0.004
Hypertension	0.42	(0.5)	0.28	(0.45)	0.000
Health Investment					
BMI	26.63	(4.74)	25.30	(4.46)	0.000
Smoker	0.29	(0.46)	0.22	(0.41)	0.011
Alcohol	-0.07	(1.02)	0.17	(0.94)	0.000
Sports	0.41	(0.41)	0.66	(0.38)	0.000
Nutrition	0.09	(0.29)	0.15	(0.36)	0.004
Health Investment Index	-0.13	(1.02)	0.31	(0.89)	0.000

Note: p-values of Mann-Whitney U-test for differences in means are reported.

Table 1.11: Main Analysis: Effect of Time Preferences on Health-Education Gradient

Variables	Small Set of Covariates				Large Set of Covariates			
	Education w\o Time	**Education with Time**	**Change**	**N**	**Education w\o Time**	**Education with Time**	**Change**	**N**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PCS	2.559 (0.000)	2.528 (0.000)	**1.21%**	1149	2.0936 (0.001)	2.0303 (0.002)	**3.03%**	975
MCS	1.291 (0.018)	1.292 (0.019)	**-0.04%**	1149	0.51 (0.401)	0.541 (0.376)	**-6.18%**	975
SRH	0.124 (0.000)	0.118 (0.000)	**5.08%**	1152	0.095 (0.003)	0.087 (0.007)	**8.89%**	978
Hypertension	-0.088 (0.019)	-0.086 (0.023)	**2.74%**	737	-0.104 (0.013)	-0.101 (0.016)	**2.39%**	609
CVD	-0.058 (0.044)	-0.056 (0.050)	**3.52%**	737	-0.059 (0.063)	-0.055 (0.081)	**5.88%**	609
BMI	-1.145 (0.000)	-1.164 (0.000)	**-1.61%**	1141	-1.347 (0.000)	-1.359 (0.000)	**-0.87%**	968
Sports	0.215 (0.000)	0.212 (0.000)	**1.3%**	1151	0.167 (0.000)	0.164 (0.000)	**2.12%**	977
Smoker	-0.109 (0.000)	-0.106 (0.000)	**3%**	1152	-0.11 (0.000)	-0.108 (0.000)	**2.45%**	978
Nutrition	0.068 (0.003)	0.065 (0.004)	**4.76%**	1151	0.053 (0.027)	0.05 (0.034)	**5.06%**	977
Alcohol	0.147 (0.001)	0.145 (0.001)	**0.94%**	1152	0.084 (0.088)	0.086 (0.085)	**-1.54%**	978
Health Investment Index	0.458 (0.000)	0.449 (0.000)	**1.89%**	1152	0.43 (0.000)	0.42 (0.000)	**2.27%**	978

Note: Each row reports the results from a regression with a different health measure as dependent variable. A probit regression is conducted for SRH, hypertension, CVD, smoker, nutrition. OLS regression conducted for PCS, MCS, BMI, sports, alcohol, and health investment index. Columns 1 and 5 report the coefficient of education without controlling for time preferences and columns 2 and 6 report the corresponding coefficient of education when controlling for time preferences. Columns 3 and 7 state the change in the health-education gradient that can be traced back to time preference. Columns 4 and 8 report the number of observations in each regression. **Small Set of Covariates**: age, age squared, gender, marital status, region, sub-sample fixed effect. **Large Set of Covariates**: additionally controlled for Big 5-factors, cognitive ability scores, risk preferences.

The Real Effects of Hedge Fund Activism

Leonie Hug*

1 Introduction

Since the late 1990s hedge fund activism has been a symbol for stronger shareholder rights and active monitoring. Recently, the number of hedge funds, the events of activism and the invested capital have surged substantially (Krishnan et al., 2015).

However, hedge fund activism is a divergent topic. Opponents argue that activism increases the financial risk of the target and induces a wealth transfer from other stakeholders to shareholders instead of creating value. Yet these arguments are based on personal experiences. Therefore, researchers criticize the subjective assumptions and use empirical tests to examine the relation. Brav et al. (2008), Becht et al. (2009), Clifford (2008), Gantchev et al. (2015) find about 5-10% short-term average abnormal returns around the activism event. Hence, contrary to former shareholder activism, hedge fund activism as an advanced corporate governance strategy generates value for shareholders.

The essential question regarding hedge fund activism is whether it creates value beyond the pure financial effect on shareholders. Do active hedge funds lead to economic real effects which outlast short-term market reactions? Do they generate spillover effects on other stakeholders?

Four different effects are identified based on recent empirical studies. Hedge fund activism improves the target firm's productivity and competitive positioning (Brav et al. (2015) and Aslan and Kumar (2016)). Innovation efficiency is enhanced by focusing on the target firm's primary expertise (Brav et al. (2014)). The demanded reforms by active hedge funds affect other stakeholders such as debtholders, employees and supply chain firms as well (Aslan and Kumar (2016), Brav et al. (2015) and Sunder et al. (2014)) . The improved performance of the target firm has positive

*Leonie Hug received her degree (B.Sc.) from the University of Bonn in 2015. The present article refers to her bachelor thesis under supervision of Prof. Dr. Hendrik Hakenes, which was submitted in August 2015.

and negative externalities for industry competitors Aslan and Kumar 2016 (Aslan and Kumar (2016) and Gantchev et al. (2015)).

The next section develops hypotheses about hedge fund activism. Section 3 analyzes the challenges and approaches to identify a causal link and suggests an alternative method. The four identified real effects of hedge fund activism are examined in section 4. The last section concludes.

2 Hypotheses Development

Using the empirical results of previous research, hypotheses regarding the real effects of hedge fund activism will be developed (see table 1).

Activists strategically target firms that are undervalued and do not achieve their full economic potential due to the agency problem of free cash flow (Brav et al., 2008 and Jensen, 1986). Brav et al. (2008) and Clifford (2008) report improvements in operating performance after the hedge fund's intervention. Improving the efficiency of assets in place and reallocating firm assets to better-suited new owners, are channels which enhance productivity. Hedge funds intend to focus on sectors in which targets benefit from a comparative advantage and thus can exploit rents better (Aslan and Kumar, 2016).

Hedge fund activists might not invest in innovative projects, as they are unpredictable, take a long time and are expensive. Nevertheless, it can be argued that comparable to hedge funds' ability to efficiently reallocate corporate assets, they may be successful in reallocating innovative assets (e.g. patents) and human capital (Brav et al., 2014).

Klein and Zur (2011) state that a fall in the target's cash and assets due to a rise in shareholder payout causes a higher credit risk. Shareholders' gain comes at the expense of debtholders in the form of mean negative bond returns in these cases. However, Sunder et al. (2014) emphasize if hedge funds monitor the target firm's governance structure more closely and reduce agency cost related inefficiencies, financial fragility is weakened and debtholders can benefit as well. Effective monitoring and better incentive-driven executive compensation, induced by activism, discipline managers and reduce managerial slack (Brav et al., 2008). They monitor workers better, which can enhance labor productivity. Hedge fund activism and the associated restructuring of the firm could lead to a lay off of workers and a cut of wages yet. The effect on customers and suppliers is ambiguous. Target firms may charge supra-competitive prices and lower their costs along the vertical supply chain due to their improved bargaining power. However, suppliers may benefit if target firms demand more inputs due to an increased market share and output. Customers may benefit if the target firm transfers some of the cost cuts.

Previous activism in an industry may threaten the target's industry peers to be targeted next by hedge funds. The threat disciplines them and induces reforms in order to prevent activism. The market forecasts these policy enhancement and values the threatened peers higher. This, in addition to the performance reform, leads to a lower ex-post probability for the peers to be targeted (Gantchev et al., 2015). The target's reformed business strategy, better governmental expertise and improved financial efficiency negatively affect peer firms post-activism (Aslan and Kumar, 2016).

3 Empirical Strategy

3.1 Obstacles

Measuring a positive effect on the target firm after hedge fund activism does not allow to assume directly that the investors' activism alone is the origin of such an improvement.

Simultaneity, omitted variables and unobserved heterogeneity cause endogeneity. Simultaneity implies that activism might affect firm aspects and improve firm performance, but certain firm characteristics could trigger activism at the same time. It could be that hedge fund managers are only superior stockpickers who target firms for which they believe they will improve the most or which are in a structural adjustment process anyway (Aslan and Kumar (2016)). This implies that real effects could happen irrespectively of the hedge fund's engagement. Omitted variables and unobserved heterogeneity arise if some (imperceptible) factors are correlated with activism as well as the expected firm development and which are not included in the analysis. Sources as heterogeneous firm characteristics (e.g. liquidity or institutional ownership) have to be considered as an independent variable, as they are likely correlated with the explanatory variable, which dilutes the activism effect.

Unobservable actions derive, because researchers normally only have access to the publicly available data from the hedge fund's Schedule 13D filings. Greenwood and Schor (2009) report that hedge funds often do not entirely reveal their strategies and that final results frequently contrast original objectives. The estimation could be weakened through this missing data problem and the identification of the real effects could be exacerbated.

Moreover, the term activism and its definition is not fixed. Section 13D of the 1934 Securities Exchange Act necessitates hedge funds to disclose within ten days of acquiring any form of publicly traded securities of a firm if it exceeds the 5% threshold and they intend to change corporate structures. There can be cases of activism, for which less than 5% of the target's shares might be efficient for investors to influence the target firm and they could try to stay below the 5% treshold to avoid disclosure (Bebchuk and Jackson, 2012). "Wolf packs", another difficulty in this context, are a group of hedge funds who collaborate to target a firm together. Although "wolf packs" have to reveal that they are acting in concert if they accumulate a combined stake of more than 5%, it can happen that they coordinate actions unconcealed and thus escape regulatory and public notice (Kahan and Rock, 2006). The studied papers do not explicitly deal with the interaction of "wolf packs".

A selection bias can occur, as hedge funds target certain firms and do not select their targets randomly. It is questionable if targeted and nontargeted firms are similar.

3.2 Approaches

Prevalent Approaches

The standard average treatment effect describes the activists' effect on the target firms (treatment group) if hedge funds randomly selected a firm and did not choose one on purpose. The control group is a subset of firms which are similar to the target firms, but which are not being targeted. However, the treatment effect is rather restricted for hedge fund activism as hedge funds do not engage in activism randomly, but choose their target wisely. Hence, the counterfactual question is more relevant: Conditioned on hedge funds' choice of targets, would the same effect have occured without activism? Studies use the Schedule 13D filings as a proxy for activism.

Aslan and Kumar (2016), Brav et al. (2015), Brav et al. (2014) and Gantchev et al. (2015) employ a difference-in-differences analysis (DiD), among other methods, to control for endogeneity. The pre- to post-intervention change after activism for the target firms relative to the nontargeted firms is examined. The parallel trend assumption between the treatment and control group implies that the average change in nontargeted firms presents the counterfactual situation for the targeted firms if they were not targeted. Hence, changes that would have occurred irrespectively of the activism can be identified due to the control group. A general DiD regression is presented here to explain the model.

$$y_{i,t} = \beta_0 + \beta_1 \cdot Target_i + \beta_2 \cdot Post_{i,t} + \beta_3 \cdot Target_i \times Post_{i,t} + \gamma \cdot Control_{i,t} + \varepsilon_{i,t} \tag{1}$$

y may be a performance, market share or innovation variable of target firm i at year t. The dummy term $Target$ is one if the firm is targeted by hedge fund activism in a given year and zero otherwise. *Post* equals one if the target firm or its matched control firm is t years after the intervention year and zero otherwise. The interaction term $Target_i \times Post_{i,t}$ is the variable of interest, which measures the effect of hedge fund activism. The inclusion of relevant control variables (e.g. market capitalization and firm age), industry and time fixed effects and clustered standard errors is necessary (Aslan and Kumar, 2016, Brav et al., 2014 and Gantchev et al., 2015).

Investors which possess a beneficial ownership of more than 5% and only have an investment intention can file a shorter Schedule 13G instead of a Schedule 13D. Passive investors cannot influence target's policies and might be stock-pickers. Brav et al. (2015) examine events in which hedge funds change from a 13G filing to a 13D filing and therefore become active investors. Hedge funds do not change their ownership stakes in the target firm due to this shift. Thus, this allows for a thorough identification of the treatment effect. If hedge funds are not only superior stockpickers, the switch should be associated with a performance improvement of the target firm.

Alternative Proposal

A natural experiment is a suitable method to measure the effects of an exogenous shock for the indenpendent variable of interest and filter the treatment effect. The EU filed a Transparency Directive (Directive 2004/109/EC) in 2004, which had to be implemented by the member states by January 20, 2007. The directive aims at harmonizing the disclosure rules concerning beneficial ownership and at forming a unified European financial market. Germany, the first country implementing the directive, reduced the disclosure threshold from 5% to 3%. The disclosure window dropped from seven calendar days to four trading days. These amendments increase the transparency of voting control for German stock operations, however they are not beneficial for hedge funds according to Bebchuk and Jackson (2012). Outside blockholders who monitor and discipline the firm's management have to pay for the cost alone, but receive only part of the benefits together with other shareholders. Shortening the disclosure window will cut their returns as public information about the acquisition of activists will cause the share price to rise. If they do not earn a net benefit, they will acquire a smaller stake or not engage in activism at all. The Transparency Directive and the following changes in Germany can be exploited as an exogenous shock to filter the effects of hedge fund activism by employing a natural experiment. The focus is made on the shortening of the disclosure window. The EU is seen as one integrated market. The following difference-in-difference-in-differences (DDD) analysis could be employed to measure the effect of a tightening of the disclosure window on hedge fund activism between target firms and nontargeted firms in Germany relative to other European countries, which keep a longer disclosure window, before and after the reform in 2007:

$$\begin{aligned} y_{i,t} = {} & \beta_0 + \beta_1 \cdot Target_i + \beta_2 \cdot Post_{i,t} + \beta_3 \cdot GER_i + \beta_4 \cdot Target_i \times GER_i \\ & + \beta_5 \cdot Target_i \times Post_{i,t} + \beta_6 \cdot Post_{i,t} \times GER_i + \beta_7 \cdot Target_i \times Post_{i,t} \times GER_i \\ & + \gamma \cdot Control_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{2}$$

$Target$ equals one if the firm is targeted by a hedge fund and zero otherwise. $Post$ is one if the target firm or its matched control firm is t years after the intervention year and zero otherwise. GER equals one if the firm is incorporated in Germany and thus experiences a shortening of the disclosure window and zero otherwise. β_7 is the coefficient of interest. Control dummies for industry, time and country specific characteristics have to be included. The Spamann (2010) anti-director rights index (ADRI) could be added as well to consider the different regulations and governance systems of the home markets (Humphrey-Jenner, 2012). It could be contemplated to exclude the UK from the analysis and focus on continental Europe to control for heterogeneity between the regions. One constraint are the fewer activism events in Europe compared to the U.S. and a smaller sample for individual countries. If Bebchuk and Jackson's (2012) theory holds, the coefficient β_7 should be negative. The reform, as an exogenous shock to hedge funds in Germany, allows for a clear identification of the effects of active intervention. If the conditions for hedge fund activism is less favorable, the real effects may be weakened and target firms may improve less.

4 The Real Effects of Hedge Fund Activism

4.1 Productivity and Competitive Positioning

Hedge fund activism generates real changes in the target firm's productivity and product market positioning. Brav et al. (2015) employ a hedge fund activism events sample from 1994 to 2007 and match plant level data (Census data) to measure the target firms' productivity to the sample. The major measure for plant performance is total factor productivity (TFP), which is more precise than firm level measures. TFP is estimated as the residual from a log-linear Cobb-Douglas production function by three-digit SIC industry and year. TFP follow a "V" shape course. Targets' plants are more productive compared to the control plants three years before the activism event. Due to bad governance productivity declines until the year of intervention. The underperformance to their own former level attracts activist investors. Three years after the intervention, targets' plants' productivity has significantly risen to a level higher than in $t-3$ and the control plants. On average, plants are 5.2% to

11.8% of the standard deviation from years t to $t+3$ more productive. Therefore productivity does not only improve in the short-term, but until three years after the intervention. The productivity increase is highest if activists intend to change "Business strategy", 27% of the standard deviation, and "Sale", 31% of the standard deviation. Brav et al. (2015) employ a probit regression to verify the reallocation hypothesis. Target plants which are less productive are more likely to be put up for sale after hedge fund activism relative to control plants. Hedge funds clearly aim to sell the target in 20% of the cases and disinvest in nonessential assets in 15% of the time. Activist investors sell underperforming plants to firms which manage the plants more efficiently in order to concentrate on the core business in approximately 23% of the cases between the activist event and three years thereafter. Nontargeted plants are only sold in 13% of the cases. These findings suggest that hedge fund activism enhances firm productivity through the reallocation of assets rather than improving the assets in place.

Aslan and Kumar (2016) examine the effect of hedge fund activism on targets' product market positioning in regard to market shares and price-cost markups with a sample lasting from 1996 to 2008. They calculate the effect with a fixed-effect analysis (modified version of (1)), a quasi-natural experiment and an endogenous switching model. Three years after the activism event, targets can expand their market share by 3.7% and increase their price-cost markups by 6.2% of the standard deviation determined by the fixed-effects estimator. Further implications in the cross section by the three models show that targets, which are less leveraged, more liquid, more profitable and have a greater size, experience even larger market share improvements and profits.

4.2 Innovation Efficiency

Innovation-intensive sectors as the high tech or pharma industry see high numbers of hedge fund activism events. Brav et al. (2014) study the relation between active interventions and corporate innovation. They restrict their sample from 1991 to 2010 to innovative firms only. This condition demands at least one patent in any year before the hedge fund's active engagement. Innovative firms account for 30% of the whole sample. Brav et al. (2014) do not explicitly verify if the results of the restricted innovative sample are consistent with the complete sample. Innovation is estimated by R&D expenditures as inputs and patent quality and quantity as outputs around the activism event. Patent quality can be measured with the number of consecutive citations, the patent's originality and the patent's generality (see table 2). Patent quantity is the number of patent applications of a firm which are accepted within one year. Brav et al. (2014) employ a DiD method for which the input and output measures are the dependent variable.

Targets' total R&D expenditures significantly decline by $20.58 million compared to nontargeted firms within five years after the intervention. This is, however, proportionate to the drop in capital expenditures due to sales post-activism (Brav et al., 2008 and Brav et al., 2015). While the input measure falls, the ouput measures, patent quantity and quality rise. Prior to the intervention, target firms register around the same amount of patents as matched firms. Post-activism, they register approximately 15.3% more patent applications and obtain 14.9% more citations than matched firms. The target's patents are cited by a wider technology class of patents as well post-activism - its generality rises.

Brav et al. (2014) identify three mechanisms for the innovation efficiency improvement: limiting the diversity of the firm's patent set and patent as well as human capital reallocation. Target firms which had a more diverse set of patents before the event, but then concentrate on their central technological expertise, see higher innovation productivity improvements. Hedge funds find more suitable owners for the innovative assets, which is consistent to Brav et al. (2015) and hedge funds' expertise in reallocating assets.

Human capital plays a pivotal role for a firm's innovation and performance. Inventors are assigned to firms where they can best maximize their productivity. Before the intervention, inventors leave and enter the target firm less often relative to matched peers. During the following five years targets catch up to their peers and inventors leave the target firm 10.4% more and enter the target 7.3% more than before. "Stayers" file 0.35 more patents and receive 1.98 more citations per patent than matched peers. "Leavers' " new patents at another firm get cited four times more often than the control group (see table 2).

4.3 Other Stakeholders

The majority of the literature deals with the effects of hedge fund activism on shareholders. However, active intervention influences other stakeholders as debtholders, employees and vertical supply chain firms as well.

Sunder et al. (2014) focus on debtholders and analyze bank loan contracts of target firms. Their sample covers the years from 1995 to 2009. The authors regress the natural logarithm of loan spreads on five independent variables resembling hedge funds' strategies prior to the intervention and an interaction term with a post-activism indicator. The interest spread of a loan rises by 33% if hedge funds demand a merger. This reflects the general assumption that creditors dislike takeovers and assume a higher credit risk. Target firms with a lower takeover risk before the intervention experience significant higher increases in interest spreads. Moreover, interest spreads rise significantly by 14% if hedge funds demand a higher payout to shareholders. Thus, the target firms take on more leverage and debtholders rate the financial risk higher. Creditors rate the hedge fund's restriction of the targets' manager to engage in empire-building mergers positively and interest spreads decrease by 20%. In addition, interest spreads decrease by approximately 11% when the hedge fund replaces the CEO to impede entrenched management. The findings demonstrate that hedge fund activism is two-sided for debtholders. They benefit if inefficiencies are weakened through closer monitoring, whereas they suffer if the financial risk of the target firm is increased.

Hedge fund activism is often not favorable to internal stakeholders. Brav et al. (2008) report a cut in CEO compensation by one million dollar within a year of the intervention and a rise in CEO turnover rate by about 10% for target firms. Pay-for-performance, which is enforced by hedge funds, equips managers with higher incentives to monitor workers as well as enhances the governance structure of the target firm. Brav et al. (2015) examine the effect of activism on wages and work hours for workers below the executive position by employing plant-level data. Labor productivity (output per hour) increases by 8.4% to 9.2% at the target plants within three years post-activism. The productivity-adjusted per-hour wages drop by 7.3% within three years after the event due to nearly stable wages. Hedge fund activism reduces inefficiencies through closer monitoring and cutting down an excess of labor. However, employees do not benefit from the rise in labor productivity, except for highly unionized industries, but rather labor rents are passed to shareholders after activism.

Aslan and Kumar (2016) identify no significant change of the abnormal returns of suppliers and customers on average post-activism by employing an event-study. The positive effect due to the target firm's increased market share as well as output and the negative effect due to the target firm's improved bargaining power seem to offset each other for customers and suppliers.

4.4 Externalities

The improved performance of the target firm can influence its competitors through the exposure to be targeted next and product market competition.

Gantchev et al. (2015) employ a sample of hedge fund activism events from 2000 to 2011 and use a DiD method. The activism threat disciplines and motivates peer firms to improve their position in the same direction as the targets in order to avoid being targeted. Peers which have similar characteristics than the target firm feel more threatened than peers with different fundamentals. Gantchev et al. (2015) report a 1.5% larger rise in mean book leverage, a 0.3% higher payout yield, a 1.5% increase in ROA and 0.5% lower capital expenditures for threatened peers with high baseline target propensity than for peers with low baseline target propensity between $t-1$ and $t+1$.

The market forecasts the proactivity of peer firms and values the threatened peers higher. Peers with high baseline target propensity gain larger monthly abnormal returns of 1.1% relative to low target propensity peers. The improvements occur within the threat event quarter, indicating that hedge fund activism is the cause. Gantchev et al. (2015) do not find a sign of price reversal. Using a linear probability model, the authors discover a feedback effect: the policy improvements and the gain in value lead to a lower ex-post-probability for the peers to be targeted, because it is more expensive for the hedge funds to acquire a substantial stake in the firm after the higher returns.

In contrast to these positive externalities, Aslan and Kumar (2016) reveal negative externalities for the rivals' operational performance. They examine all rivals relative to target firms in regard to the product market competition, whereas Gantchev et al. (2015) differentiate among nontargeted firms with high and low probability to be targeted and thus focus exclusively on the effect of activism threat, separating this from competitive and other factors.

Aslan and Kumar (2016) employ univariate and multivariate tests to measure the abnormal returns for rivals. Their market-adjusted abnormal return decreases by 1.8% within the five-day horizon post-activism. Proposals concerning changes in business strategy experience the most negative abnormal returns of 3.6% (2.9% in cross-sectional analyses). Aslan and Kumar (2016) use a fixed-effects model to measure rivals' operational performance changes in excess of the target's performance. Rival firms' cash flows decrease by about 2% (2.8%) and EBITDA by 2.4% (2.7%) relative to the target firms' cash flow after the hedge fund activism event in the short-term (long-term). R&D investment and TFP for rivals are 1.3% smaller and 4% lower, respectively, compared to targets in the third post-event year. Using the endogenous switching model, the authors measure the counterfactual situation: If the hedge fund had not targeted the firm, the rivals would have had higher market shares and markups.

In other tests, Gantchev et al. (2015) report significant positive returns of 0.9% for rivals without filtering the activism threat effect. As mentioned before, Aslan and Kumar (2016) measure negative returns for rivals due to competitive effects. The positive impact of threat and the negative impact of product market competition may balance each other. The net effect is not clear, as both studies focus on different channels.

5 Conclusion

The findings contradict the assumption that hedge fund activism exclusively generates financial effects. The developed hypotheses are confirmed: Active hedge funds create real effects which go beyond the short-term market reaction and beyond the target firm as well. The reallocation of assets, especially unproductive ones, enhances the target firm's productivity and the competitive positioning (Brav et al., 2015 and Aslan and Kumar, 2016). Focusing on the primary firm expertise and redeploying patents as well as human capital to better suited owners improve the innovation efficiency of the target firm (Brav et al., 2014). Other stakeholders are able to enjoy the benefits of hedge fund activism only under certain circumstances. Debtholders benefit from activism strategies which monitor the target firm closer, whereas they suffer from demands which may increase the financial risk (Sunder et al., 2014). Employees improve their labor productivity, however they are not able to share the gain, except for workers in highly unionized industries, as their wages stay flat (Brav et al., 2015). Target firms extract a greater surplus from supply chain firms with a lower bargaining power (Aslan and Kumar, 2016). Moreover, active activism threat disciplines other industry competitors and lower their probability to be targeted (Gantchev et al., 2015). Nevertheless, the improved competitive standing of the target firm weakens the rivals' performance (Aslan and Kumar, 2016).

It will be interesting to see how hedge fund activism will develop. Will the industry remain to grow and will more hedge funds push into the market? Will they still be able to create real effects?

Bibliography

Aslan, Hadiye and Praveen Kumar (2016). "The product market effects of hedge fund activism". In: *Journal of Financial Economics* 119.1, pp. 226–248.

Online Appendix to: The Real Effects of Hedge Fund Activism by Leonie Hug

Table 2.1: Hypotheses

Hypothesis	Prediction	Rationale
Productivity & competitive positioning	Improvement	Reallocation of assets, focus on firm's core expertise
Innovation efficiency	Improvement	Limiting patent diversity, reallocation of patents and human capital
Other stakeholders	Ambiguous:	
	Positive	Closer monitoring mitigates inefficiencies and increases labor productivity, target firm demands more inputs and transfer cost cuts
	Negative	Credit risk increases for certain strategies, productivity-adjusted wages drop, target firm gains a higher bargaining power
Externalities	Ambiguous:	
	Positive	Activism threat disciplines rivals
	Negative	Target's improved competitive position weakens rivals' performance

Table 2.2: List of Variables

Variable	Definition	Source
Market share	Ratio of the target's annual sales to the absolute sales by the target firm and the matched competitor	Aslan and Kumar (2014)
Price-cost markups	Empirical standard cost minimization model, ratio of output price to the marginal cost	Aslan and Kumar (2014)
Core business	If 20% or more of the target's patent stock comes from this division	Brav et al. (2014)
Generality	The more expanded technology classes of patents cite a patent, the greater is its generality	Brav et al. (2014)
Originality	The more expanded technology classes of patents a patent cites, the greater is its originality	Brav et al. (2014)
Diversity	One minus the Herfindahl-Hirschman index of patents filed by a firm in the past three years across 2-digit technological classes defined by the NBER patent database	Brav et al. (2014)
Leaver	Inventor who leaves the target firm in a given year and invents at least one patent in the target firm before the activism and one patent in a new firm afterwards	Brav et al. (2014)
Stayer	Inventor who stays at the target firm in a given year and invents at least one patent in the target firm before and after the activism	Brav et al. (2014)

Spillover effects of financial deregulation on income and income characteristics

Nicolas Kaufung*

1 Introduction

As the Great Recession started to enfold almost the entire world economy, financial regulation attracted a lot of attention in the media and politics. Till now, there are loud calls for stronger regulation and significant governmental intervention. Yet, in many areas of financial intermediation, there is mixed or only little evidence about the implications of banking regulation on economic outcomes. One important aspect of the general debate concerning the optimal degree of governmental intervention is the assumptive tradeoff between financial stability and banking efficiency. According to this hypothesis, higher concentration in the banking sector leads to less efficiency and therefore hinders economic growth (Besanko and Thakor 1992; Guzman 2000; Boot and Thakor 2000). At the same time, market power resulting from concentration leads to higher profits and thereby increases financial stability (Keeley 1990; Besanko and Thakor 2004; Perotti and Suarez 2002; Hellmann, Murdock, and Stiglitz 2000). A deregulation wave that took place in the US during the second half of the 20th century established the unique opportunity to analyze the impact of banking competition on economic performance. Starting in the 1970s, state governments began to slacken the legislation of bank expansion through M&A as well as through de novo branching within and across state borders. In contrast to other settings, the staggered process, combined with the analogousness of deregulation across states, gives economists the potential to analyze the impact of banking competition controlling for unobserved state differences and aggregate shocks.

The existing literature analyzes the impact of deregulation on economic characteristics like growth (Jayaratne and Philip E. Strahan 1996), entrepreneurship (Black and Philip E Strahan 2002; Kerr and Nanda 2009), income-inequality (Beck, Levine, and Levkov 2010) or innovation (Cornaggia et al. 2015). However, all studies focus on

*Nicolas Kaufung joins the Boston Consulting Group GmbH as an Associate in January 2017 nicolas.kaufung@gmail.com.

the impact of deregulation within a given state. To determine the overall impact of financial deregulation and banking competition it is necessary to understand whether deregulation affects other states in addition. It might be the case that positive impacts in a given state come at the expense of negative impacts in surrounding states. If so, previous results overestimate the beneficial impact of deregulation. In contrast, if positive impacts spill over to surrounding states, then previous results underestimate the positive impact for the US. In addition, such patterns would give cause for considerations about the optimal level of legislation for financial regulation. Given that there are spillover effects of deregulation, it is unclear whether state governments actually exhibit incentives in line with the universe of affected subjects. In conclusion, the existence of spillover effects would indicate a need for financial legislation on a higher level – i.e. on the federal level in the US.
This study tackles this issue by analyzing spillover effects of financial deregulation on geographical neighbor states, focussing on the impact on personal income and the income distribution.

2 Relevant Deregulations

2.1 Intra-State Banking

The structure of the US banking sector has dramatically changed during the last century. Most of the states used to prohibit intra-state branching – that is, a BHC was not allowed to open up another branch of the existing bank at a different location or to acquire another bank and organize it as a branch. However, except for unit banking states, it was possible to organize Multibank Holding Companies (MBHC) that were allowed to own several banks as long as their businesses were organized independent of each other. The resulting operational structure prohibited the realization of a variety of scale effects (Amel and Liang 1992).
The restrictive legislation was mainly profitable for small banks in rural areas whose local monopolies were protected. In contrast, larger BHCs from major cities were more expansion minded and formed a strong lobby for the relaxation of regulation. The first states softened branching regulations in the 1930s. However, in 1970, still only 12 states allowed intra-state banking. Starting in 1970, 39 states followed and allowed intra-state branching until 1999.[1] First, most states allowed branching through M&A only. Opening up new branches was just allowed after the legalization of de novo banking, which was delayed in most states (Jayaratne and Philip E. Strahan 1996). In the meantime, between those two deregulations, banks were still able to extract a higher surplus of the acquired target by using the monopoly power

[1]For a complete list of intra-state branching regulation dates see 3.1 in Appendix 5. For a graphical illustration of the geographical pattern of intra deregulation see 3.1 in Appendix 6.

resulting from the high entry barriers.
Beginning in the late 1990s, the first analyses of the reforms were conducted. The descriptive findings of the immediate impact on the banking sector are that intra-state banking reforms led to a huge wave of BHCs entering local banking markets via de novo banking or M&A (Amel and Liang 1992), especially small banks were subject of consolidation into bigger BHCs (Calem 1994) and many former independent banks of MBHCs were reorganized as branches of another bank within the MBHC (McLaughlin 1995).

2.2 Inter-State Banking

Inter-state banking, that is bank branching through de novo banking and M&A across state-borders, was generally prohibited by federal law after the Bank Holding Company Act and its Douglas Amendment was implemented in 1956. The Bank Holding Company Act was passed to prohibit BHCs to take a share in non-banking activities. However, its Douglas Amendment also prohibited BHCs headquartered in one state to acquire banks or open up new branches in another state.
The first state to relax this legislation was Maine, where the acquisition of incumbent banks through MHBC's from other states was legalized in 1978. However, they decided to open up their banking sector only to MBHCs headquartered in states that opened their banking sector as well – I will refer to this legislation as "reciprocity of inter-state banking" (Savage 1993). In fact, most states used the reciprocal design for the inter-state reform.[2] Since Maine was the first state to deregulate, and because of the reciprocal design, the relaxation had no real impact until the 1980s, when other states started to pass similar laws. Afterwards, in a fast staggered process, all states except Hawaii allowed inter-state M&A. As for intra-state deregulation, the relaxation of inter-state de novo banking was somewhat delayed (P. Strahan 2003). In 1994 federal government passed the Inter-state Banking and Branching Efficiency Act (IBBEA) that allowed nation-wide inter-state banking from 1997 on.
McLaughlin (1995) analyze the immediate impact on the banking sector itself after inter-state banking deregulation. The descriptive findings are that the reactions to inter-state banking deregulation were, in comparison to intra-state banking, lagged and that BHCs that made use of the relaxed regulation expanded to geographical neighbor states in most cases.

[2]For a list of which states deregulated when and how see 3.1 in Appendix 5. For a graphical illustration of the geographic pattern of inter-state banking deregulation see 3.2 in Appendix 6.

3 Empirical Analysis

3.1 Motivation

While there is a wide literature analyzing the direct impact of intra- and inter-state banking deregulation, little is known on whether these deregulations affected surrounding states. The existing literature emphasizes many positive consequences of deregulation: increasing growth of GSP, personal income (Jayaratne and Philip E. Strahan 1996) and entrepreneurial activity (Black and Philip E Strahan 2002; Kerr and Nanda 2009) as well as shrinking income inequality (Beck, Levine, and Levkov 2010). However, to determine the total impact on the US it is insufficient to consider the isolated impact on a state itself.

Especially inter-deregulation had a direct impact on neighboring states. The existing literature indicates that bank expansion was mainly focused on neighbor states (McLaughlin 1995). In fact, some states even restricted the expansion to banks headquartered in neighbor states (Savage 1993). Following this logic, one would expect inter-state banking deregulation to have a direct impact on the banking sector of contiguous states – and through the banking sector on the real economy. For instance, after California allowed inter-state banking, banks headquartered in Nevada were able to expand to California. This obviously affects the access of Californian citizens and firms to banking services. At the same time, the investment opportunities of Nevada's banks changed.

In contrast to inter-state, intra-state banking does not directly affect the banking sector of neighbor states. In this case, the channel through which spillover effects can occur is the real economy. The literature paints a picture of strong improvements in the access of (new) firms to financial services (Kerr and Nanda 2009). These led to increased foundations of establishments. It is possible that entrepreneurs decided to move from one state to another to get access to improved financial institutions. This would have a direct impact on business foundations in states with non-improving financial institutions. Furthermore, intra-state banking increased the relative wages of unskilled to skilled workers (Beck, Levine, and Levkov 2010). This, however, decreases the relative returns to education which could especially foster migration of high-skilled workers (Borjas, Bronars, and Trejo 1992).

To determine the total impact of financial deregulation, it is therefore necessary to find out whether and how financial deregulation affected other states. The following sections tackle this issue by analyzing the impacts of financial deregulation on geographical neighbor states.

3.2 Data

The main analysis is based on the data of Beck, Levine, and Levkov (2010). Observations are on the state-year level and include information about economic as well as social characteristics of 49 states of the US. Following the existing literature, the data excludes observations from South Dakota and Delaware because of their special position in the credit card sector. It covers the years from 1976 to 2006 and consists of 1,519 observations. It also includes the year of intra-state banking deregulation, which is defined as the year where M&A restrictions were lifted. This study focuses on geographical spillover effects, therefore I drop observations from Alaska and Hawaii leading to 1,457 observations. To analyze the effects of inter-state banking deregulation I add the dates where inter-state M&A restrictions were lifted from P. Strahan (2003).
I extend the data with measures of the degree of financial deregulation of neighbor states. For this purpose I use geographical neighbor-ship data provided by *The State Border Dataset* by Thomas J. Holmes from the University of Minnesota.
To get a more precise picture of the effects of deregulation on personal income, I use data provided by the *Current Population Survey (CPS)* march supplement for the years 1977 to 2007 which include information about the years 1976 to 2006. The original data includes 5,174,724 observations. I restrict the sample to observations that have non-missing, positive household- and personal-weights (5,174,516 obs.) as well as non-missing income data (3,550,900 obs.). Furthermore, to make the results comparable to the existing literature, I focus on individuals aged between 25 to 54 years (1,995,437 obs.) with non-missing ethnic and schooling data that do not live in group quarters (1,971,870 obs.). Following the existing literature I truncate the data at the 1st an 99th percentile of the income distribution to deal with outliers. This leaves me with a final dataset containing 1,931,435 observations. The *CPS* allows me to calculate income characteristics like income distribution measures (i.e. percentiles of the income distribution or the Gini-coefficient), as well as income by sub-population (gender, age, education) and migration patterns. I calculate these characteristics and merge them to the state-year level data.
The *CPS* includes a sub-sample of the total population. Therefore, income variables calculated using the *CPS* are only proxies for the actual state level income statistics. To get more precise aggregate statistics I use data provided by the *Bureau of Economic Analysis.*
To make the results compareable to Beck, Levine, and Levkov (2010) I transform all monetary variables to 2001 US$ values. 3.2 provides summary statistics for all relevant variables of the given state-level data.

3.3 Empricial Strategy

For the empirical analysis I mainly focus on classic difference-in-difference (*DiD*) estimators of the following form:

$$log(dep\_var_{it}) = \alpha + indep\_var_{it}B + \epsilon_{it} \tag{1}$$

This specification allows to analyze the impact of the independent variables captured in a row vector $indep\_var_{it}$ on an arbitrary dependent variable dep_var_{it}. Here, i refers to the state and t to the year of interest. $indep\_var_{it}$ represents different deregulation measures. For instance, those could be dummy variables indicating deregulation in the state itself or the fraction of geographical neighbors that already deregulated prior to date t. B is a column vector capturing the impact of the independent variables. In case that $indep\_var_{it}$ consists only of dummy variables, this is a classic *DiD* estimator. It compares the average change of the dependent variable (before vs. after deregulation) for deregulating vs. not deregulating states – the estimated mean differences are captured by B. Taking the natural logarithm of the dependent variable enables us to interpret the estimates (B) as percentage changes in dep_var after deregulation in the case of a dummy variable. For general scales, the estimate refers to the percentage change in dep_var after an increase of one unit in the independent variable of interest. Note that the estimator assumes that the impact of different independent variables is additive and constant.
Every *DiD* estimator relies on two important assumptions: i) No contaminating events and ii) Parallel trends. Assumption i) refers to events taking place at the same time as deregulation which might be the actual cause of the measured effects. From a statistical perspective it is impossible to rule out contaminating events. Assumption ii) refers to time-varying, unobservable characteristics that affect the dependent variable in a different way for deregulating vs. non-deregulating states or in a different way for different degrees of neighbor deregulation. To check the assumption of parallel trends, several pre-event parallel trends tests were implemented that validated the correctness of the assumption.
To further improve the reliability of the estimates, I include state fixed effects to control for unobserved, time-invariant characteristics of states and year fixed effects to control unobserved, state-invariant characteristics of specific years. Including fixed effects leads to the following *DiD* specification:

$$log(dep\_var_{it}) = indep\_var_{it}B + \delta_t + \gamma_i + \epsilon_{it} \tag{2}$$

In this specification δ_t captures year fixed effects and γ_i state fixed effects. Additionally to state and year fixed effects, there could be regional business cycles that drive the results. To control for those I use the classification of Jayaratne and Philip E. Strahan (1996) to divide the US in 4 broad regions denoted $R_j,\ j \in \{1,2,3,4\}$ and estimate the *DiD* specification with time fixed effects that are allowed to vary across regions:

$$log(dep\_var_{it}) = indep\_var_{it}B + \sum_{j=1}^{4} \mathbb{I}(i \in R_j)\delta_{jt} + \gamma_i + \epsilon_{it} \quad (3)$$

I further utilize three different sets of control variables that Beck, Levine, and Levkov (2010) use to make the results comparable. The first one is the empty set. The second one (called Xs) includes the proportion of blacks, the proportion of high-school dropouts and the proportion of female headed households. The third set ($Xs2$) includes the second set, the unemployment-rate and the GSP-growth-rate.
Clustering standard errors is necessary in situations where estimation errors are not independent across observations. In the given setting, it is likely that estimation errors are not independent in one state across time or in a given period across states (or even both). However, the cluster-level is always somewhat arbitrary in natural settings since estimation errors might also be correlated in bigger regions or within a state only for a given set of counties. We are not able to determine the optimal cluster. Following Bertrand, Duflo, Mullainathan, et al. (2004) I am going to use state level clusters in most specifications which is the most conservative way of dealing with correlated estimation errors without decimating the set of independent observations too much.
If not stated otherwise, I refer to the specification including state and year fixed effects, control set Xs and standard errors clustered on the state level.

3.4 Baseline

This section tries to replicate the results of Beck, Levine, and Levkov (2010). The main finding in Beck, Levine, and Levkov (2010) is that intra-state deregulation decreased income inequality by raising low incomes while keeping high incomes unchanged. The results using the restricted sample do not confirm this result. The baseline regression includes only a dummy variable $Intra_{it}$ that is equal to one in the years after intra-state banking deregulation. The estimates are presented in 3.3. They imply an increase in low incomes (11% for the 10th percentile, $p < 0.05$) as well as a decrease in the Gini-coefficient (-2.4%, $p < 0.01$). However, also high incomes decline significantly in all specifications (2.7% for the 90th percentile, $p < 0.05$). There is no statistically significant impact on the mean income, but the point estimate indicates a decrease of 1.4%. The impact on the median income is neither of statistical nor of economic significance. The described pattern is robust to the different control sets as well as regional time fixed effects. However, they disappear after controlling for state-trends.
Beck, Levine, and Levkov (2010) also claim that only intra-deregulation affects the

Gini-coefficient in a joint test of intra- and inter-deregulation. To verify this result I add a dummy variable $Inter_{it}$ to the regression that is equal to one after inter-state banking deregulation. Again, the estimates for the restricted sample do not confirm their result. As 3.5 suggests, both, intra- as well as inter-deregulation, had a significant impact on the Gini-coefficient. The estimates of intra deregulation are similar to the ones of the previous specification. While intra-deregulation increases low incomes and decreases high incomes, inter-deregulation seems to affect only the lower and middle part of the income distribution. The results indicate a significant 1.9% ($p < 0.01$) increase in mean incomes and a 3.6% ($p < 0.01$) increase of the median income. Again, the findings are robust to different control sets and regional time fixed effects. Additionally, the estimates for $Inter$ remain significant after controlling for trends, whereas the estimates for $Intra$ lose statistical as well as economic significance.

3.5 Spillover Effects

The previous section described the baseline effects of intra- and inter-state banking deregulation. It remains necessary to find out whether there are spillover effects of financial deregulation to determine the total impact. The exact definition of a "spillover effect" is somewhat different for intra- and inter-state banking deregulation. Intra-deregulation had no direct impact on the banking sector of other states. Therefore, spillover effects measure only second order impacts of deregulation. In contrast, inter-deregulation directly affected the banking sector of other states – it increased their prospects of expansion. I use a broad definition of spillover effects that captures all impacts arising through reforms in other states.

The staggered process of deregulation allows for different empirical strategies. For instance, one could use the number or the fraction of deregulated neighbors as an additional independent variable in the regression analysis. The first strategy assumes that each neighbors deregulation has an impact that is independent of the total number of neighbors. The second strategy assumes that the impact of one neighbor deregulation is inversely proportional to the total number of neighbors, and therefore that the total impact of neighbor deregulation is constant. The linear structure of both estimation strategies further implies that the effect of one neighbor deregulation is independent of the number of neighbors that already deregulated. I am going to use the second strategy because the assumption that the influence of one particular neighbor state is decreasing in the total number of neighbors seems natural. This strategy is further appealing because one can interpret the estimates in an absolute fashion – that is, the estimate is equal to the total effect of the deregulation of all neighbors. Note, however, that this interpretation heavily relies on the linear structure of the estimator.

Let Ω_i be the set of all neighbors of state i and ω_i the number of neighbors. I also define contiguous neighbors with a border length of 0 miles (for instance Utah and New Mexico) as geographical neighbors and therefore include them in Ω_i. I define the following variables for each observation:

$$n\_Inter_{it} = \sum_{j\in\Omega_i} \frac{Inter_{jt}}{\omega_i} \tag{4}$$

$$n\_Intra_{it} = \sum_{j\in\Omega_i} \frac{Intra_{jt}}{\omega_i} \tag{5}$$

Including the measures of neighbor deregulation, the *spillover specification* reads as follows:

$$\begin{aligned} log(dep\_var_{it}) = \quad & \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} \\ & + \beta_4 n\_Inter_{it} + \delta_t + \gamma_i + \epsilon_{it} \end{aligned} \tag{6}$$

3.7 presents the results of the regression including the measures of the degree of deregulation of geographical neighbor states. The results imply that the qualitative findings about the impact of intra- and inter-state banking deregulation on the Gini-coefficient, mean, low and high incomes are robust to controlling for spillover effects. However, the impact of inter-state deregulation on the Gini-coefficient declines and gets less significant (-0.9%, $p < 0.1$). Additionally I find significant spillover effects. The estimates for n_Intra w.r.t. the income distribution are qualitatively the same as for $Intra$, however, they are bigger in magnitude leading to a significant decrease in the mean income. The Gini-coefficient decreases by about 4.1% ($p < 0.05$) which is about three times the impact of deregulation in the state itself (-1.7%, $p < 0.05$). The 10th percentile increases by 7.7% ($p < 0.1$) after intra-deregulation and 22.2% ($p < 0.05$) after neighbor intra-deregulation while the 90th percentile decreases by 2.5% ($p < 0.05$) after own intra-deregulation and 4% ($p < 0.1$) after neighbor intra-deregulation. In contrast, there are negative but insignificant spillover effects of inter-state deregulation on the Gini-coefficient. However, mean, median and high incomes significantly increase after neighbor inter deregulation (5.4%, 7.4% and 4.6%, respectively, all $p < 0.01$). As in the previous specifications, these results are robust to all control specifications except state-trends.
To further illustrate the impacts on the income distribution 3.3 plots the estimates of the spillover effects regression with every 5th income-percentile as dependent variable and for each independent variable of interest (i.e. $Intra$, $Inter$, n_Intra and n_Inter). The $Intra$ graph in 3.3 looks qualitatively very similar to the one in Beck, Levine, and Levkov (2010). However, in contrast to their estimates, I do find a significant negative impact on higher income-percentiles. The pattern of n_Intra is similar to the pattern of $Intra$, that is, the impacts of intra-state banking deregulation of neighbors and in the state itself are qualitatively similar. Low incomes increase while high incomes decrease. However, the point estimates for

spillover effects are even bigger and more precise for lower incomes, which indicates that neighbor deregulation has an even higher impact than own deregulation. While the estimates of $Intra$ are in the range of 9% for the 5th and -3% for the 80th percentile, the estimates for n_Intra range from 40% for the 5th to -5% for the 80th percentile.
In contrast to intra-state deregulation, inter-state deregulation has a positive point estimate for all income percentiles. However, only the ones from 25th to 60th percentiles are significant on a 5% level with an average point estimate of about 3%. This pattern further explains the positive impact of inter deregulation on mean income. The graph for n_Inter indicates significant positive estimates for the 35th to the 90th percentile which are slightly bigger in magnitude than the ones for $Inter$ (average about 6%).

3.6 Reciprocity of Inter-Deregulation

As discussed in the previous sections, inter-state banking deregulation had a reciprocal design in many states.[3] Banks headquartered in neighbor states were allowed to acquire a bank in a deregulated state only if their home state allowed similar acquisitions. Following this logic, the first deregulation should not have any impact (Amel and Liang 1992). Furthermore, the effect of a deregulation should be increasing in the number of other states that already deregulated. As a result, the previous estimates for $Inter$ and n_Inter are likely to be biased. For illustration suppose that every state used the reciprocal design and that only direct neighbors are relevant for spillover effects. If a state deregulated, but none of its neighbors, we should not observe any impact and therefore an estimate of 0 for $Inter$. Only after the first neighbor state deregulates, we would observe an effect. This impact gets completely picked up by n_Inter. However, what is happening is the result of both deregulations – that is, both directions are important: banks from state x are now allowed to buy banks in neighbor states as well as banks in neighbor states are now allowed to buy banks in state x. As McLaughlin (1995) points out, most targets of M&A after inter-state deregulation were banks in geographical neighbor states. Therefore, I include an interaction term of $Inter$ and n_Inter to control for the reciprocal structure and to isolate spillover effects from interaction effects.
Using the interaction I am able to measure the isolated impact of own deregulation ($Inter$), which should be equal to zero if all states used the reciprocal design. Furthermore, n_Inter captrues the spillover effects of inter-deregulation, but not any direct impacts of neighbor deregulation on the banking sector in the state itself. Finally, the interaction measures the effects that just arise through the combination of both ($Inter \cdot n\_Inter$, to which I will refer as $Rec.Inter$) – it measures the higher

[3]For a list of which states actually implemented the reciprocal legislation see 3.1 in Appendix 5.

banking competition in the state itself as well as the increased prospects of expansion. Note that this strategy implicitly assumes that all states used the reciprocal legislation, which was not the case. However, further splitting the independent variables according to which states actually used the reciprocal design makes any identification with the given sample size impossible.
Including the interaction, the *reciprocity specification* reads as follows:

$$\begin{aligned} dep\_var_{it} = \quad & \alpha + \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} \\ & + \beta_5 (Inter_{it} \cdot n\_Inter_{it}) + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it} \end{aligned} \tag{7}$$

3.9 shows the estimates for different dependent variables using this specification. As soon as I control for the reciprocity of inter-state deregulation, all estimates for inter-deregulation in the state itself become insignificant. The estimates of neighbor inter-deregulation for the mean income and the 90th percentile decrease and lose statistical significance. However, the median income still increases by almost 6% ($p < 0.01$) after neighbor inter-deregulation. The estimates further indicate strong impacts of *Rec.Inter* on the mean income, which increases by 4% ($p < 0.01$) after all neighbors and the state itself deregulated inter-state banking. In contrast to the results for intra-state deregulation, these estimates are robust to all control sets – even state-level trends. The pattern indicates that the spillover effects measured in the last section are partially driven by inter-deregulation coming in effect after neighbor deregulation because of the reciprocal structure. However, since acquisitions are allowed in both directions as soon as both states deregulate, we are not able to distinguish between the effects of increased competition due to the threat of acquisitions from outside or increasing opportunities due to the possibility of acquiring banks in another state. Furthermore, the estimates for n_Inter could be driven by more distant neighbors, or the fact that actually not all states used the reciprocal design of deregulation. In contrast to inter-state deregulation, the results stay basically unchanged for intra-deregulation. Here we find strong spillover effects that have a similar pattern to the deregulation in the state itself. Again, these results are robust to all control sets, but not to state-trends.
It seems like the interaction term of $Inter$ and n_Inter is necessary to get an unbiased picture of the effects of inter-state banking deregulation as well as its spillover effects. 3.4 plots the impact on the income distribution for the reciprocity specification. One can see that the impact of $Inter$ and n_Inter is very limited as soon as we control for their interaction. The graph for $Inter$ exhibits not a single significant estimate. The one for n_Inter indicates some increases in the middle part of the income distribution. However, this pattern is not very strong and, as described above, could be driven by other factors. The objective of this study is to find out whether and how financial deregulation affected surrounding states. Since I am not able to identify any clean spillover effects of inter-state banking deregulation, the remainder focuses on intra-state banking deregulation and its spillover effects.

3.7 Low-Skill Labor Demand

The patterns of the impact of intra-deregulation and its spillover effects on the income as well as the wage distribution indicate strong increases in low incomes as well as significant reductions in high incomes. One channel which would result in such a pattern could be an increasing relative demand for unskilled labor. This section provides a test on whether intra-state banking deregulation affected the relative wage of unskilled to skilled workers. A simple comparison of the incomes of low- and high-education employees is not sufficient, since low and high skill subjects might not be homogenous in other time-varying aspects as well. To control for characteristics like experience, gender and race I use a two step procedure following Beck, Levine, and Levkov (2010).
In a first step I regress the wage of high skill subjects[4] on experience, gender and race, their interactions and 3 polynomials (denoted as a row vector N_{it}) separately for every year using the *CPS* data.

$$log(wage_{it}) = N_{it}\beta_t + \epsilon_{it} \tag{8}$$

I use the estimates of this regression to predict the wages of the entire sample - that is low- as well as high-skill employees. One can interpret the residual of low-skill workers as an estimate of the education wage gap that abstracts from differences in experience, race and gender. The predicted value of a low-skill subject is the wage it had with the same characteristics but a higher level of education. The resulting residual is the difference of the wage of low and high skill subjects that can not be explained by gender, experience or race.

$$r_{it} = log(wage_{it}) - N_{it}\beta_t \tag{9}$$

In a last step, I aggregate the residuals for low skilled workers to the state-year level (i.e. calculate the mean residual using the individual weights) and analyze whether the aggregated residuals changed after the reforms using the standard regression design discussed above. The results for different control sets are presented in the 3.12. Recall that the wage gap is defined as the natural logarithm of the actual wage of low-skill workers less the predicted logarithm of the wage of low-skill workers using the estimates for high-skill workers. Since $r_{it} = log(wage_{it}) - log(\hat{wage}_{it}) = log(wage_{it}/\hat{wage}_{it})$, the estimate can be interpreted as a percentage change in the relative wages of unskilled to skilled workers. The estimates of 3.12 indicate an average 2.8% and 5.3% ($p < 0.05$ for all specifications) increase in the relative wages after intra- and neighbor intra-deregulation, respectively. Again the impact of total neighbor deregulation is bigger than the impact of deregulation in the state itself. The average relative wage is about 60%. Therefore, these estimates indicate an

[4]High-skill is defined as 12 or more years of education.

increase of about 1.7 percentage-points after intra-deregulation and 3.2 percentage-points after all neighbors deregulated intra-state banking. The results are strong evidence for the outlined hypothesis that intra-state banking deregulation increased the relative demand for unskilled labor and thereby increased the relative wage of low- to high-skill workers in the state itself as well as surrounding states.

4 Conclusion

The baseline results of this thesis are in conflict with the findings of the existing literature. Beck, Levine, and Levkov (2010) show that especially intra-state banking deregulation led to decreasing income inequality. They claim that this effect was driven by an increased relative demand for low-skill labor that increased low incomes while keeping high incomes unaffected. The analysis using the sample restricted to continental United States paints a different picture. Indeed, the relative wage of low- to high-skill workers increased. However, this effect was not only driven by increasing low incomes, but also by decreasing high incomes.

Furthermore, Beck, Levine, and Levkov (2010) do not find any impact of inter-state banking deregulation on the income distribution. Again, analysis restricted to continental United States do not confirm this result. This paper provides evidence for inter-state banking deregulation decreasing income inequality by increasing middle incomes.

Additionally to distributional impacts, both deregulations differ in the impact on mean incomes. While inter-state banking deregulation significantly increased mean incomes, intra-state banking exhibits a negative, insignificant point estimate in most specifications, and therefore had at best no impact on the mean income.

The existing literature emphasizes that intra-state banking deregulation, which occurred prior to inter-state banking deregulation, led to consolidations of local monopolies and many small banks. In contrast, inter-state banking mainly increased competition that was already strong. These observations, combined with the evidence provided by this thesis, are in line with a very short-time horizon version of the Kuznets hypothesis. Prior to intra-state banking deregulation, mainly high income subjects benefited from financial intermediation of local banking monopolies with bad screening and monitoring abilities and therefore a high need for guarantees and collateral. After intra-state banking deregulation, monopoly rents decreased and the access of low income subjects to financial intermediation improved causing lower income inequality. Further competition through inter-state banking deregulation affected the mean income but had less sever re-distributional impacts. However, a more careful analysis of which banking services and whose access to financial intermediation actually improved after deregulation is needed to fully understand the differences in the impacts of the two deregulations and the relationship between

financial development and income inequality.
In terms of spillover effects, this study provides evidence for a strong relationship between financial deregulation and economic outcomes of neighbor states. The reciprocal structure of inter-state banking deregulation makes a clean identification of spillover effects impossible. However, the results imply that especially because of this reciprocal design economic outcomes where dependent on regulation of other states. For intra-state banking deregulation the qualitative impact of neighbor deregulation is similar to the one of intra-deregulation in the state itself – low incomes increase while high incomes decrease and both effects are driven by wages and salary income. Furthermore, as for intra-deregulation in the state itself, neighbor deregulation leads to a shrinking education wage gap that provides evidence in favor of an increasing relative demand for low-skill labor. The point estimates indicate that the impact of neighbor deregulation is even higher than the impact of own deregulation. However, the size of the estimated total impact of neighbor deregulation heavily relies on the linear structure of the estimator. Therefore, every comparison of the magnitude of the impact of own vs. neighbor deregulation needs to be treated with caution.
The results are highly robust to pre-event parallel trend tests as well as placebo tests. Additionally, the geographical structure was substantiated by showing that impacts decrease in distance and increase in initial migration activity. Further tests of different definitions of neighborliness imply that the results are qualitatively robust to neighbor weights which are proportional to GSP, population or border length. However, only GSP weights were able to improve the precision of the estimates which indicates that the size of the economy is an important determinant of the magnitude of spillover effects.
The described findings directly imply that financial legislation can have huge impacts on economic outcomes of neighbor states. This raises the question whether the incentives of policymakers are actually in line with the interests of the universe of affected people, firms and institutions. Since this might not necessarily be the case, higher level politics seem to be inevitable for democratic decision making that incorporates the interests of each affected subject.

Bibliography

Amel, Dean F and J Nellie Liang (1992). "Relationship between Entry into Banking Markets and Changes in Legal Restrictions on Entry, The". In: *Antitrust Bull.* 37, p. 631.

Beck, Thorsten, Ross Levine, and Alexey Levkov (2010). "Big bad banks? The winners and losers from bank deregulation in the United States". In: *The Journal of Finance* 65.5, pp. 1637–1667.

Bertrand, Marianne, Esther Duflo, Sendhil Mullainathan, et al. (2004). "How Much Should We Trust Differences-in-Differences Estimates?" In: *The Quarterly Journal of Economics* 119.1, pp. 249–275.

Besanko, David and Anjan V Thakor (1992). "Banking deregulation: Allocational consequences of relaxing entry barriers". In: *Journal of Banking & Finance* 16.5, pp. 909–932.

— (2004). *Relationship banking, deposit insurance and bank portfolio choice.* Tech. rep. EconWPA.

Black, Sandra E and Philip E Strahan (2002). "Entrepreneurship and bank credit availability". In: *The Journal of Finance* 57.6, pp. 2807–2833.

Boot, Arnoud WA and Anjan V Thakor (2000). "Can relationship banking survive competition?" In: *The journal of Finance* 55.2, pp. 679–713.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo (1992). "Self-selection and internal migration in the United States". In: *Journal of Urban Economics* 32.2, pp. 159–185. ISSN: 0094-1190. DOI: http://dx.doi.org/10.1016/0094-1190(92)90003-4. URL: http://www.sciencedirect.com/science/article/pii/0094119092900034.

Calem, Paul (1994). "The impact of geographic deregulation on small banks". In: *Business Review* Nov, pp. 17–31.

Cornaggia, Jess et al. (2015). "Does banking competition affect innovation?" In: *Journal of financial economics* 115.1, pp. 189–209.

Guzman, Mark G (2000). "Bank structure, capital accumulation and growth: a simple macroeconomic model". In: *Economic Theory* 16.2, pp. 421–455.

Hellmann, Thomas F, Kevin C Murdock, and Joseph E Stiglitz (2000). "Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough?" In: *American economic review* 90.1, pp. 147–165.

Jayaratne, Jith and Philip E. Strahan (1996). "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation". English. In: *The Quarterly Journal of Economics* 111.3, pp. 639–670. ISSN: 00335533. URL: http://www.jstor.org/stable/2946668.

Keeley, Michael C (1990). "Deposit insurance, risk, and market power in banking". In: *The American Economic Review* 80.5, pp. 1183–1200.

Kerr, William R and Ramana Nanda (2009). "Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship". In: *Journal of Financial Economics* 94.1, pp. 124–149.

McLaughlin, Susan (1995). "The impact of interstate banking and branching reform: evidence from the states". In: *Current Issues in Economics and Finance* 1.2.

Perotti, Enrico C and Javier Suarez (2002). "Last bank standing: What do I gain if you fail?" In: *European Economic Review* 46.9, pp. 1599–1622.

Savage, Donald T (1993). "Interstate banking: A status report". In: *Fed. Res. Bull.* 79, p. 1075.

Strahan, Philip (2003). "The real effects of US banking deregulation". In: *Federal Reserve Bank of St. Louis Review* 85.July/August 2003.

Online Appendix to: Spillover effects of financial deregulation on income and income characteristics by Nicolas Kaufung

5 Tables

Table 3.1: Dates of Intra and Inter-State Banking Deregulation

Notes: This table shows the year of intra-state and inter-state banking deregulation. I define the year of deregulation as the year when M&A restrictions were lifted in both cases. An "X" for Inter Reciprocity indicates that the state used the reciprocal design of inter-state banking legislation.

State	Intra M&A	Inter M&A	Inter Reciprocity
Alabama	1981	1988	X
Arizona	1960	1987	
Arkansas	1994	1990	X
California	1960	1988	X
Colorado	1991	1989	
Connecticut	1980	1984	X
District of Columbia	1960	1986	X
Florida	1988	1986	X
Georgia	1983	1986	X
Idaho	1960	1986	
Illinois	1988	1987	X
Indiana	1989	1987	X
Iowa	1999	1992	X
Kansas	1987	1993	X
Kentucky	1990	1984	X
Louisiana	1988	1988	X
Maine	1975	1988	
Maryland	1960	1986	X
Massachusetts	1984	1984	X
Michigan	1987	1987	X
Minnesota	1993	1987	X
Mississippi	1986	1989	X
Missouri	1990	1987	X
Montana	1990	1994	X
Nebraska	1985	1991	X
Nevada	1960	1986	
New Hampshire	1987	1988	
New Jersey	1977	1987	X
New Mexico	1991	1990	
New York	1976	1982	X
North Carolina	1960	1986	X
North Dakota	1987	1992	X
Ohio	1979	1986	X
Oklahoma	1988	1988	
Oregon	1985	1987	
Pennsylvania	1982	1987	X
Rhode Island	1960	1985	X
South Carolina	1960	1987	X
Tennessee	1985	1986	X
Texas	1988	1988	
Utah	1981	1985	
Vermont	1970	1989	X
Virginia	1978	1986	X
Washington	1985	1988	X
West Virginia	1987	1989	X
Wisconsin	1990	1988	X
Wyoming	1988	1988	

Table 3.2: Summary Statistics

Variable	**Description**	**Mean**	**Std.Dev.**
Regulation			
Intra_reform	Date of intra-deregulation	1,980.68	4.843
Inter_reform	Date of inter-deregulation	1,987.49	2.324
Inter	Dummy equal to one after inter-deregulation	0.629	0.483
Intra	Dummy equal to one after intra-deregulation	0.721	0.449
n_Intra	Mean of intra of geographical neighbors	0.71	0.372
n_Inter	Mean of inter of geographical neighbors	0.605	0.462
Rec.Inter	Interaction of Inter & n_Inter	0.583	0.476
CPS			
Gini	Gini-coefficient	0.431	0.023
P10	10th Percentile of Income Distribution	6,551	2,803
P50	50th Percentile of Income Distribution	35,686	14,179
P90	90th Percentile of Income Distribution	82,893	29,832
BEA – Income Source Data (Entire Sample)			
Personal Income	Per Capita Personal Income	24,959	5,309

*: Per total salary employment of the given industry

Table 3.3: Impact of Intra Deregulation on Income Characteristics

Notes: This table shows estimates of the impact of intra-state banking deregulation on different income characteristics. *Intra* is a dummy variable which is equal to one in the years after deregulation. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$. For regression results using different control sets, trends and regional time fixed effects see 3.4 in Appendix 5.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Gini	Personal Income PC	P10	P50	P90
Intra	-0.0240***	-0.0141	0.108**	-0.000262	-0.0269**
	(0.00661)	(0.00908)	(0.0442)	(0.0101)	(0.0100)
Observations	1,457	1,457	1,457	1,457	1,457
R-squared	0.375	0.939	0.763	0.672	0.727
Number of statefip	47	47	47	47	47
Controls	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No
Trend	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.4: Impact of Intra Deregulation

Notes: This table shows estimates of the impact of intra-state banking deregulation on different income characteristics. *Intra* is a dummy variable which is equal to one in the years after deregulation. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Gini	Personal Income PC	P10	P50	P90	Gini	Personal Income PC	P10	P50
Intra	-0.0254***	-0.0134	0.109**	0.000461	-0.0283**	-0.0208***	-0.0198*	0.0853*	-0.00525
	(0.00746)	(0.0109)	(0.0473)	(0.0119)	(0.0111)	(0.00623)	(0.0101)	(0.0425)	(0.0110)
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.359	0.935	0.762	0.639	0.718	0.398	0.946	0.778	0.695
Number of statefip	47	47	47	47	47	47	47	47	47
Controls	No	No	No	No	No	Xs2	Xs2	Xs2	Xs2
Regional Time FE	No	No	No	No	No	No	No	No	No
Trend	No	No	No	No	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Gini	Personal Income PC	P10	P50	P90	Gini	Personal Income PC	P10	P50
Intra	-0.00672	0.00489	0.0178	0.0184	0.000600	-0.0147**	-0.00687	0.0724*	0.00078
	(0.00598)	(0.00730)	(0.0428)	(0.0115)	(0.00811)	(0.00610)	(0.00900)	(0.0380)	(0.0107)
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.513	0.964	0.808	0.733	0.817	0.613	0.979	0.842	0.887
Number of statefip	47	47	47	47	47				
Controls	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Trend	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Impact of Intra and Inter Deregulation on Income Characteristics

Notes: This table shows estimates of the impact of intra- and inter-state banking deregulation on different income characteristics. *Intra* and *Inter* are dummy variables which are equal to one in the years after deregulation. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$. For regression results using different control sets, trends and regional time fixed effects see 3.6 in Appendix 5.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Gini	Personal Income PC	P10	P50	P90
Intra	-0.0226***	-0.0156*	0.105**	-0.00310	-0.0271***
	(0.00642)	(0.00918)	(0.0439)	(0.00996)	(0.0101)
Inter	-0.0184***	0.0193***	0.0385	0.0359***	0.00363
	(0.00541)	(0.00676)	(0.0459)	(0.0108)	(0.00851)
Observations	1,457	1,457	1,457	1,457	1,457
R-squared	0.381	0.939	0.764	0.677	0.727
Number of statefip	47	47	47	47	47
Controls	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No
Trend	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Impact of Intra and Inter Deregulation

Notes: This table shows estimates of the impact of intra- and inter-state banking deregulation on different income characteristics. *Intra* and *Inter* are dummy variables which are equal to one in the years after deregulation. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$.

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90	(6) Gini	(7) Personal Income PC	(8) P10	(9) P50
Intra	-0.0241***	-0.0147	0.106**	-0.00200	-0.0285**	-0.0201***	-0.0202*	0.0863**	-0.00673
	(0.00730)	(0.0111)	(0.0471)	(0.0119)	(0.0112)	(0.00612)	(0.0102)	(0.0419)	(0.0108)
Inter	-0.0172***	0.0168**	0.0418	0.0315***	0.00205	-0.0119**	0.00720	-0.0158	0.0232**
	(0.00543)	(0.00679)	(0.0453)	(0.0112)	(0.00860)	(0.00527)	(0.00651)	(0.0429)	(0.00973)
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.364	0.936	0.763	0.643	0.718	0.400	0.946	0.778	0.697
Number of statefip	47	47	47	47	47	47	47	47	47
Controls	No	No	No	No	No	Xs2	Xs2	Xs2	Xs2
Regional Time FE	No	No	No	No	No	No	No	No	No
Trend	No	No	No	No	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90	(6) Gini	(7) Personal Income PC	(8) P10	(9) P50
Intra	-0.00545	0.00260	0.0155	0.0150	-0.000427	-0.0136**	-0.00767	0.0699*	-0.00110
	(0.00577)	(0.00689)	(0.0419)	(0.0108)	(0.00781)	(0.00599)	(0.00915)	(0.0375)	(0.0108)
Inter	-0.0136**	0.0246***	0.0246	0.0369***	0.0110	-0.0167***	0.0126**	0.0389	0.0294**
	(0.00531)	(0.00744)	(0.0473)	(0.0120)	(0.00943)	(0.00591)	(0.00586)	(0.0480)	(0.0105)
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.516	0.965	0.808	0.739	0.818	0.616	0.979	0.842	0.888
Number of statefip	47	47	47	47	47				
Controls	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Trend	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Impact of Intra and Inter Deregulation on Income Characteristics – Spillover Specification

Notes: This table shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on different income characteristics. $Intra$ and $Inter$ are dummy variables which are equal to one in the years after deregulation. n_Intra and n_Inter are the fraction of geographical neighbors that already deregulated. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$. For regression results using different control sets, trends and regional time fixed effects see 3.8 in Appendix 5.

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90
Intra	-0.0169***	-0.0141	0.0770*	-0.00551	-0.0245**
	(0.00594)	(0.00975)	(0.0413)	(0.0106)	(0.0102)
Inter	-0.00906*	0.0153**	-0.00381	0.0246**	0.00204
	(0.00511)	(0.00680)	(0.0478)	(0.0104)	(0.00950)
n_Intra	-0.0412***	-0.0328*	0.222**	-0.00819	-0.0401*
	(0.0124)	(0.0192)	(0.0874)	(0.0194)	(0.0207)
n_Inter	-0.0173	0.0537***	0.0456	0.0741***	0.0459***
	(0.0113)	(0.0150)	(0.0916)	(0.0171)	(0.0154)
Observations	1,457	1,457	1,457	1,457	1,457
R-squared	0.404	0.941	0.770	0.684	0.732
Number of statefip	47	47	47	47	47
Controls	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No
Trend	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on different income characteristics. *Intra* and *Inter* are dummy variables which are equal to one in the years after deregulation. *n_Intra* and *n_Inter* are the fraction of geographical neighbors that already deregulated. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$.

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90	(6) Gini	(7) Personal Income PC	(8) P10	(9) P50
Intra	-0.0176***	-0.0141	0.0777*	-0.00596	-0.0260**	-0.0154***	-0.0174	0.0652*	-0.0076
	(0.00629)	(0.0114)	(0.0432)	(0.0124)	(0.0112)	(0.00573)	(0.0106)	(0.0383)	(0.0113
Inter	-0.00748	0.0117*	0.00150	0.0183*	-0.000302	-0.00527	0.00751	-0.0379	0.0169*
	(0.00515)	(0.00675)	(0.0464)	(0.0106)	(0.00945)	(0.00464)	(0.00658)	(0.0423)	(0.00951
n_Intra	-0.0462***	-0.0244	0.211**	0.00545	-0.0373*	-0.0392***	-0.0374*	0.206**	-0.0103
	(0.0132)	(0.0219)	(0.0908)	(0.0201)	(0.0207)	(0.0121)	(0.0209)	(0.0823)	(0.0207
n_Inter	-0.0160	0.0518***	0.0501	0.0725***	0.0468***	-0.00724	0.0334***	-0.0465	0.0527**
	(0.0113)	(0.0156)	(0.0934)	(0.0173)	(0.0156)	(0.0102)	(0.0109)	(0.0797)	(0.0143
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.392	0.937	0.769	0.651	0.723	0.418	0.947	0.783	0.701
Number of statefip	47	47	47	47	47	47	47	47	47
Controls	No	No	No	No	No	Xs2	Xs2	Xs2	Xs2
Regional Time FE	No	No	No	No	No	No	No	No	No
Trend	No	No	No	No	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Lev

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90	(6) Gini	(7) Personal Income PC	(8) P10	(9) P50
Intra	-0.00399	-0.000254	0.00843	0.0109	-0.00278	-0.0122**	-0.00766	0.0623*	-0.0020
	(0.00594)	(0.00694)	(0.0430)	(0.0110)	(0.00766)	(0.00567)	(0.00912)	(0.0355)	(0.0108
Inter	-0.00875	0.0140**	-0.00617	0.0216*	0.000697	-0.0107*	0.0107*	0.00654	0.0228**
	(0.00587)	(0.00680)	(0.0510)	(0.0113)	(0.00854)	(0.00600)	(0.00630)	(0.0479)	(0.0110
n_Intra	-0.00147	0.0168	0.102	0.0262	0.0344**	-0.0279**	-0.0264	0.161**	0.0153
	(0.00993)	(0.0133)	(0.0732)	(0.0198)	(0.0154)	(0.0107)	(0.0184)	(0.0740)	(0.0192
n_Inter	-0.0273***	0.0486***	0.0999	0.0689***	0.0333**	-0.0220*	0.0333**	0.110	0.0593**
	(0.0101)	(0.0152)	(0.0880)	(0.0190)	(0.0145)	(0.0120)	(0.0159)	(0.0785)	(0.0200
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.520	0.966	0.810	0.748	0.822	0.623	0.979	0.845	0.889
Number of statefip	47	47	47	47	47				
Controls	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Trend	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Lev

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.9: Impact of Intra and Inter Deregulation on Income Characteristics – Reciprocity Specification

Notes: This table shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on different income characteristics. *Intra* and *Inter* are dummy variables which are equal to one in the years after deregulation. n_Intra and n_Inter are the fraction of geographical neighbors that already deregulated. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \beta_5(Inter_{it} \cdot n\_Inter_{it}) + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$. For regression results using different control sets, trends and regional time fixed effects see 3.10 in Appendix A.

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Gini	Personal Income PC	P10	P50	P90
Intra	-0.0170***	-0.0135	0.0769*	-0.00521	-0.0242**
	(0.00596)	(0.00957)	(0.0413)	(0.0105)	(0.0102)
Inter	-0.00761	-0.000592	0.000293	0.0172	-0.00579
	(0.00754)	(0.00769)	(0.0492)	(0.0130)	(0.00928)
n_Intra	-0.0415***	-0.0304	0.221**	-0.00706	-0.0389*
	(0.0126)	(0.0194)	(0.0872)	(0.0194)	(0.0207)
n_Inter	-0.0145	0.0226	0.0536	0.0596***	0.0306
	(0.0130)	(0.0184)	(0.109)	(0.0205)	(0.0231)
Rec.Inter	-0.00368	0.0403***	-0.0104	0.0188	0.0199
	(0.0130)	(0.0145)	(0.0948)	(0.0211)	(0.0215)
Observations	1,457	1,457	1,457	1,457	1,457
R-squared	0.404	0.942	0.770	0.684	0.732
Number of statefip	47	47	47	47	47
Controls	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No
Trend	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.10: Impact of Intra and Inter Deregulation on Income Characteristics – Reciprocity Specification

Notes: This table shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on different income characteristics. *Intra* and *Inter* are dummy variables which are equal to one in the years after deregulation. n_Intra and n_Inter are the fraction of geographical neighbors that already deregulated. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \beta_5 (Inter_{it} \cdot n\_Inter_{it}) + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$.

VARIABLES	(1) Gini	(2) Personal Income PC	(3) P10	(4) P50	(5) P90	(6) Gini	(7) Personal Income PC	(8) P10	(9) P50
Intra	-0.0177***	-0.0133	0.0773*	-0.00546	-0.0256**	-0.0156***	-0.0166	0.0658*	-0.0072
	(0.00634)	(0.0113)	(0.0432)	(0.0123)	(0.0112)	(0.00577)	(0.0103)	(0.0383)	(0.0112
Inter	-0.00520	-0.00689	0.0105	0.00606	-0.0105	-0.000809	-0.0156*	-0.0561	0.00598
	(0.00771)	(0.00811)	(0.0484)	(0.0141)	(0.00987)	(0.00769)	(0.00788)	(0.0473)	(0.0121
n_Intra	-0.0465***	-0.0217	0.210**	0.00724	-0.0358*	-0.0398***	-0.0343	0.209**	-0.00878
	(0.0133)	(0.0221)	(0.0910)	(0.0201)	(0.0207)	(0.0123)	(0.0210)	(0.0819)	(0.0206
n_Inter	-0.0116	0.0153	0.0677	0.0484**	0.0267	0.00151	-0.0120	-0.0822	0.0313*
	(0.0135)	(0.0197)	(0.114)	(0.0216)	(0.0236)	(0.0131)	(0.0139)	(0.100)	(0.0180
Rec.Inter	-0.00581	0.0473***	-0.0228	0.0312	0.0261	-0.0112	0.0582***	0.0458	0.0274
	(0.0128)	(0.0140)	(0.0981)	(0.0207)	(0.0212)	(0.0133)	(0.0159)	(0.0903)	(0.0210
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.392	0.938	0.769	0.651	0.723	0.419	0.948	0.783	0.701
Number of statefip	47	47	47	47	47	47	47	47	47
Controls	No	No	No	No	No	Xs2	Xs2	Xs2	Xs2
Regional Time FE	No	No	No	No	No	No	No	No	No
Trend	No	No	No	No	No	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Lev

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.11: Impact of Intra and Inter Deregulation on Income Characteristics – Reciprocity Specification continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Gini	Personal Income PC	P10	P50	P90	Gini	Personal Income PC	P10	P50
Intra	-0.00399	0.000745	0.00673	0.0115	-0.00212	-0.0122**	-0.00741	0.0625*	-0.00202
	(0.00594)	(0.00668)	(0.0428)	(0.0110)	(0.00768)	(0.00569)	(0.00901)	(0.0356)	(0.0108)
Inter	-0.00876	-0.000492	0.0184	0.0131	-0.00895	-0.00922	-0.00337	-0.00364	-0.0209
	(0.00785)	(0.00724)	(0.0501)	(0.0143)	(0.00902)	(0.00870)	(0.00700)	(0.0550)	(0.0164)
n_Intra	-0.00147	0.0201	0.0966	0.0281	0.0366**	-0.0281**	-0.0247	0.162**	0.0150
	(0.0100)	(0.0134)	(0.0720)	(0.0199)	(0.0156)	(0.0107)	(0.0187)	(0.0734)	(0.0193)
n_Inter	-0.0274**	0.0201	0.148	0.0523**	0.0143	-0.0191	0.00480	0.0899	0.0554**
	(0.0114)	(0.0182)	(0.100)	(0.0233)	(0.0240)	(0.0128)	(0.0204)	(0.0938)	(0.0239)
Rec.Inter	1.93e-05	0.0363***	-0.0615	0.0211	0.0241	-0.00373	0.0365**	0.0263	0.00491
	(0.0123)	(0.0125)	(0.0896)	(0.0214)	(0.0199)	(0.0126)	(0.0166)	(0.0992)	(0.0252)
Observations	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457	1,457
R-squared	0.520	0.966	0.810	0.748	0.822	0.623	0.979	0.845	0.889
Number of statefip	47	47	47	47	47				
Controls	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs	Xs
Regional Time FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Trend	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Cluster	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Level	State Lev

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.12: Impact on the Relative Wage of Unskilled Workers

Notes: This table shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on the relative wage of low- to high-skill workers. $Intra$ and $Inter$ are dummy variables which are equal to one in the years after deregulation. n_Intra and n_Inter are the fraction of geographical neighbors that already deregulated. The underlying regression specification is a log-linear regression of the following form: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \beta_5(Inter_{it} \cdot n\_Inter_{it}) + \delta_t + \gamma_i + \epsilon_{it}$.

	(1)	(2)	(3)	(4)
VARIABLES	Edu. Wage Gap	Edu. Wage Gap	Edu. Wage Gap	Edu. Wage Gap
Intra	0.0290**	0.0283**	0.0264**	0.0292**
	(0.0126)	(0.0118)	(0.0116)	(0.0128)
n_Intra	0.0586***	0.0447**	0.0421**	0.0691***
	(0.0205)	(0.0171)	(0.0165)	(0.0194)
Inter	0.00507	0.0146	0.00593	0.0172
	(0.0218)	(0.0203)	(0.0209)	(0.0210)
n_Inter	0.000589	0.0126	-0.00666	-0.0126
	(0.0415)	(0.0398)	(0.0377)	(0.0417)
Rec.Inter	-0.00777	-0.0182	-0.00657	-0.0160
	(0.0365)	(0.0347)	(0.0349)	(0.0358)
Observations	1,457	1,457	1,457	1,457
R-squared	0.576	0.589	0.592	0.624
Controls	No	Xs	Xs2	Xs
Trend	No	No	No	No
Cluster	State Level	State Level	State Level	State Level
Regional Time FE	No	No	No	Yes

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6 Graphs

Figure 3.1: Dates of Intra Deregulation

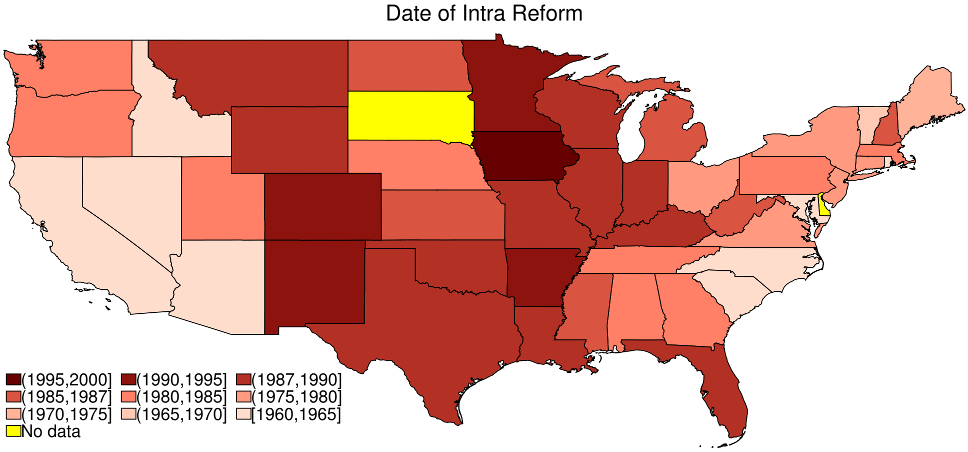


Figure 3.2: Dates of Inter Deregulation

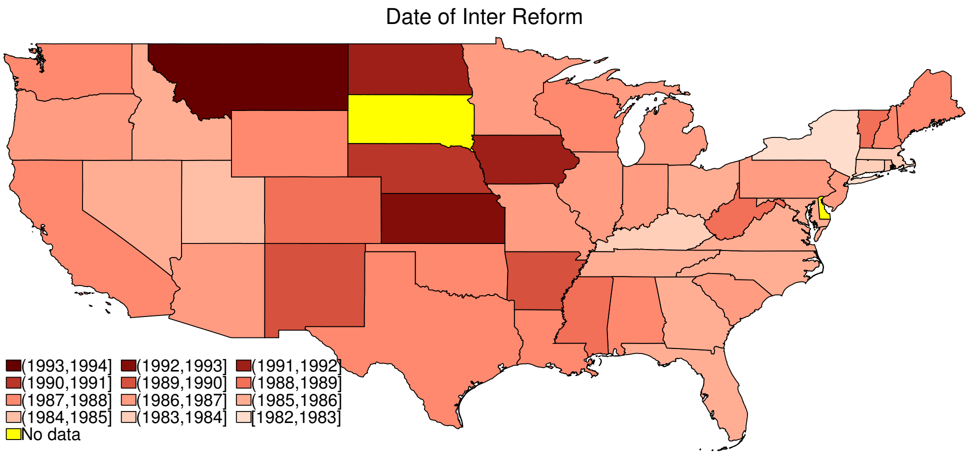


Figure 3.3: Impact of financial deregulation on percentiles of income distribution

Notes: This graph shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on the income distribution. Underlying regression specification: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$, control-set Xs, standard errors clustered on state level. x-axis plots income percentiles from P5 to P95 in steps of 5 as dependent variable of the regressions. Bars represent the coefficient of $Intra$, $Inter$, n_Intra and n_Inter (i.e. β_1 - β_4). Light blue bars are significant at 10%level, dark blue at the 5% level.

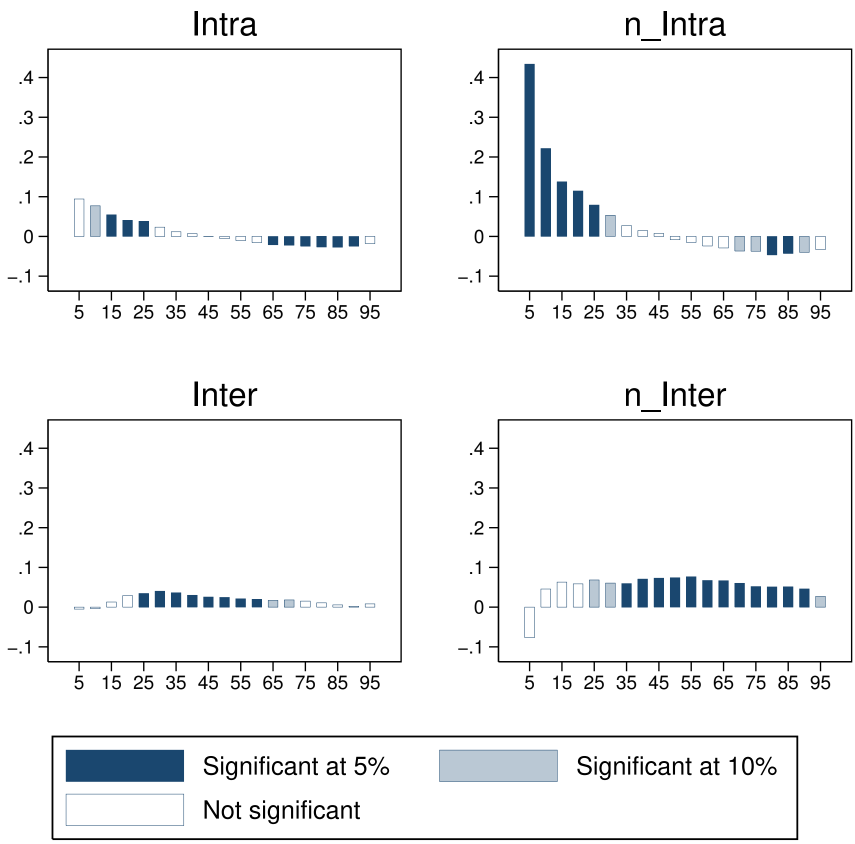
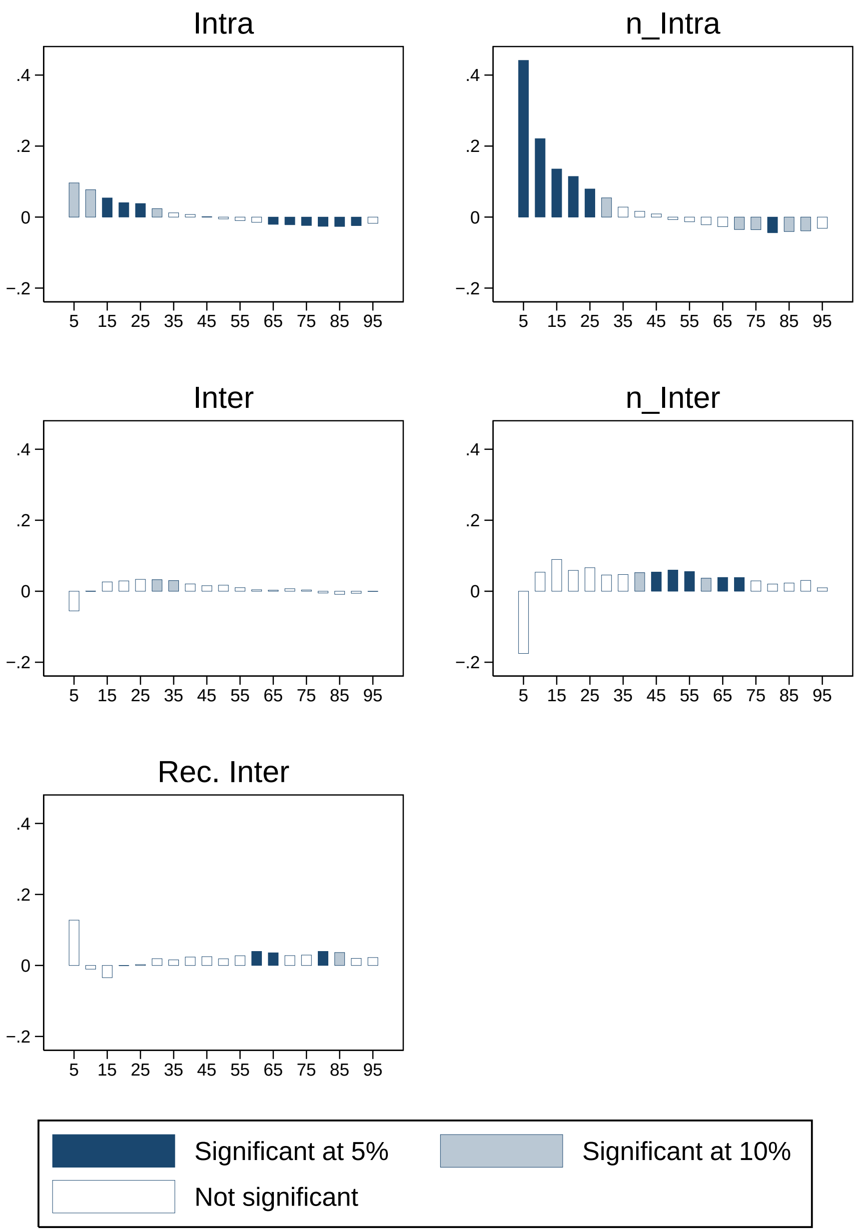


Figure 3.4: Impact on Income Distribution

Notes: This graph shows estimates of the direct impact as well as spillover effects of intra- and inter-state banking deregulation on the income distribution. Underlying regression specification: $log(dep\_var_{it}) = \beta_1 Intra_{it} + \beta_2 Inter_{it} + \beta_3 n\_Intra_{it} + \beta_4 n\_Inter_{it} + \beta_5(Inter_{it} \cdot n\_Inter_{it}) + \delta_t + \gamma_i + X_{it}\Theta + \epsilon_{it}$, control-set Xs, standard errors clustered on state level. x-axis plots income percentiles from P5 to P95 in steps of 5 as dependent variable of the regressions. Bars represent the coefficient of $Intra$, $Inter$, n_Intra, n_Inter and the interaction of $Inter$ and n_Inter ($Rec.Inter$) (i.e. β_1 - β_5). Light blue bars are significant at 10%level, dark blue at the 5% level.



Sovereign debt reduction through privatization: The case of Portugal

Lukas Nüse*

1 Introduction

As in many other countries, the Portuguese national debt sharply increased following the outbreak of the 2008 financial crisis. Despite accounting for only 71.7 percent of gross domestic product (GDP) in 2008, public debt reached 111.1 percent of GDP by 2011. Interest rates on Portuguese government bonds rose at the same time. From 3.88 percent at the start of 2008, rates on ten-year Portuguese government bonds rose to 9.19 percent by the end of April 2011. On April 7, 2011, the Portuguese government applied for financial assistance from EFSM and EFSF funds, followed subsequently by the International Monetary Fund (IMF). Lenders agreed to provide the Portuguese state with a total of 78 billion euro over the next three years.

Included in the terms of the loans, among other conditions, was the privatization of numerous enterprises with state-held shares. Revenues from privatization were supposed to reduce the debt burden to a tolerable level. Thus, the interest burden on the state budget should have decreased, thereby allowing the high budgetary deficit to reach a manageable level. This raises the question of how privatization can help achieve these objectives and under which circumstances the privatization of a state enterprise is judicious. Also to be depicted are the reasons for which the Portuguese government selected certain state enterprises for sale.

This paper is structured as follows: the first section gives an overview of enterprises which have been privatized thus far, followed by a summary of the state of research on the topic of privatization.The main body attempts to use the theoretical knowledge about privatization and its effect on public debt, in order to assess the usefulness of the individual privatization instances within the framework of the Portuguese bailout package. The six largest privatized enterprises by revenue are considered.

*Lukas Nüse received his degree in Economic (B.Sc.) from the University of Bonn in 2015. The present article refers to his master thesis under supervision of Prof. Dr. Jürgen von Hagen, which was submitted in September 2013. Today, Lukas is working in the European Department of the German Federal Ministry of Labour and Social Affairs.

Central to this paper is the comparison of profitability of a state enterprise with the costs of sovereign debt.

2 Overview: Privatized state enterprises in Portugal since 2011

The Portuguese Stability and Growth Program (SGP) 2009-13 provides the basis for the privatization under the rescue program. The SGP, which was presented in May 2010 to the European Commission, contains a list of enterprises whose state holdings are to be either partially or completely sold (see 4.1). In the May 2011 agreement with the lending parties, the Portuguese government was obligated to accelerate this privatization program. It also promised to not just partially but to completely divest from government holdings in Energias de Portugal and Redes Energéticas Nacionais, two energy firms. An overview of privatizations that have taken place in 2011 can be found in 4.2.

What is not publicly known is the criteria for which a given firm's shares were selected for privatization. Section three of this thesis attempts to retrace these privatization decisions on the basis of balance sheet data. With the sale of company ownership shares, the Portuguese state raised nearly 9.1 billion euro, 3.1 billion more than originally planned (see 4.2). The proceeds from privatization corresponded to 5.3 percent of 2014 GDP. The largest revenues were generated through the sale of shares in Energias de Portugal (EDP) and Aeroportos de Portugal (ANA), each at 3.1 billion euro. Portugal underwent a large-scale privatization effort previously in the period from 1996 to 2000. The sale of state-owned enterprises at that time generated the Portuguese government income of 16 percent of GDP (Abbas et al 2013).

Despite these high additional revenues, to date there has been no reduction in sovereign debt. The public debt of Portugal[1] rose from 196 billion euro (111.1% of GDP) in 2011 to 225 billion euro (130% of GDP) in 2014. In contrast, the deficit [2] declined in the same period from 7.4% to 4.5% of GDP. Therefore improvement was experienced primarily in reduction in primary deficit.

[1] The Maastricht debt criteria includes the positions in currency and deposits (coins in circulation), money market and capital market papers, as well as short- and long-term credit (Bundesbank).

[2] Government deficit, according to the Maastricht criteria, refers to the consolidated deficits or surpluses of the federal government, provinces, municipalities and social insurance (Bundesbank).

3 Literature Review: The macroeconomic and fiscal aspects of privatization

In principle, governments have two possible uses for revenues from the sale of state-owned enterprises. First they can immediately spend the money on regular state expenditures. This is particularly attractive for states which have difficulties refinancing. This option, however, contains the risk that government spending will continue at an unsustainably elevated level due to the loss of privatization revenue in subsequent periods (Davis et al 2000). Secondly, the revenue can be saved or used to reduce the national debt level. The latter option reduces the interest burden on the budget and could be understood as a signal to capital markets for policy change. Thus the risk premium on government bonds could possibly be further reduced (Davis et al 2000). Some findings indicate that states save the income from privatization, that is to say, use it to reduce public debt, rather than to spend it immediately (Davis et al 2000; Barnett 2000).

The initial hypothesis is to be examined whether the sale of a state-owned firm has a positive effect on the state budget. In a functioning market with perfect competition, the sale price should equal the sum of the discounted future cash flows. The state indeed has more income available in the period of privatization yet correspondingly less income in subsequent periods. From this perspective, privatization therefore has no impact on the state budget (Mansoor 1987). Under equivalent conditions, a similar assertion can be reached about the net assets of a state. The sale of assets to reduce debt leaves the net assets (i.e. equity) at the same level (Davis et al 2000, Katsoulakos and Likoyanni 2002). The interest burden saved from debt reduction corresponds to the loss of cash flows from the company to the state in future periods. These assertions, however, apply only to instances in which the discounted interest rate is equal for the private and public sectors, the profitability of a company remains unchanged after the sale and the general conditions of the market do not change. (Heller 1990; Hemming und Mansoor 1987; Katsoulakos und Likoyanni 2002). These assumptions are, however, rarely true in reality. The full text of this thesis goes into this issue in more precise detail.

In summary, the benefits of privatization tend not to come from short-term revenues, but from the long-term improvements in various macroeconomic indicators (Pinheiro und Schneider, 1994).

4 Analysis: Privatization since 2011

Considering an overview of the state-held companies sold since 2011, in the second half of this paper, the following two questions arise:

1. Using which criteria did the Portuguese government choose enterprises for privatization?
2. Were these privatizations justified at that time?

The analysis and evaluation of privatization takes place from the perspective of an entrepreneur who is faced with an investment decision. According to a business management point of view, a debt-financed investment under simplified assumptions is to be undertaken if the annual return is higher than the payable lending rates. Transferred to a government level, this means that the return on state ownership shares in an enterprise should be higher than the interest that the state must pay for its bonds. The decision to privatize is taken every time the government takes on new debts because it could privatize a company instead and use the revenue to finance the budget. These considerations form the basis of the following sections.

4.1 Methodology

Firstly, the development of various key figures from the company's annual balance sheet up to the time of privatization is presented. These indicators provide information on the state of the firm, or rather its profitability, and allow conclusions to be drawn about the decision to privatize these firms. The first focus here is on net profit, i.e. the profit after deducting all costs including interest payments. This shows whether the firm was at all profitable in the past. The second focus is on equity, i.e. the value of the assets after deduction of all liabilities, which is also an indicator of the value of a company. Subsequently, the net return of the state investment is considered. The net yield is calculated as follows:

$$Net\ yield_t = \frac{(S_t - S_{t-1})}{S_t} + \frac{(D_{t-1})}{S_t} - i_t \tag{1}$$

Net yield = share price yield + dividend return[3] *- interest on government bonds (2Y,5Y,10Y)* [4]

1. case: Return on shares < refinancing interest

$$\frac{(S_t - S_{t-1})}{S_t} + \frac{(D_{t-1})}{S_t} < i_t \tag{2}$$

If the financing interest is permanently higher than the return on shares (share price yield + dividend yield), it is sensible to sell the shares in order to pay off debt. The

[3] A dividend from Period t-1 is distributed in Period t

[4] Monthly averages are used (Banco de Portugal).

interest payments for the value of shares would be greater than their total return in monetary units. The budget would thus be sustainably relieved through such a sale.

2. case: Return on shares > refinancing interest

$$\frac{(S_t - S_{t-1})}{S_t} + \frac{(D_{t-1})}{S_t} > i_t \tag{3}$$

Conversely, a sale would have a negative effect on the budget if the return on shares is permanently above the governments refinancing rate.

The period from the introduction of the euro in 2002 until the year of privatization is considered. In order to determine the net effect in the past on the national budget, the geometric mean of all monthly net yields is calculated. The difficulties with this type of valuation include, among other things, to determine the financing interest for the Portuguese state as an investor. A state is financed by bonds with different maturities and thus different interest rates. For this reason, a comparison with the return on equities is made by means of bond interest rates with two, five, and ten-year maturities. Only companies that are sufficiently large and whose privatization has already been completed under the reform program are subject to analysis. For this particular paper, two of the six total company analyses from the bachelor thesis were selected as examples.

4.2 Analysis of privatization cases

Energias de Portugal (EDP)

Energias de Portugal, one of the largest energy suppliers in Europe, was initially renationalized after the Carnation Revolution of 1974. An initial partial privatization took place in July 1997. With the sale of nearly 30 percent of shares, EDP shares were publicly traded on the stock market for the first time. In the course of the Portuguese rescue program, the last publicly held shares were sold. In December 2011, 21.35 percent of shares were sold to Chinese firm China Three Gorges Corporation for 3.45 euro per share. At 2.7 billion euro, the proceeds exceeded the expected total by 600 million euro. 4.14 percent of the shares remained in possession of the state. They were ultimately sold in February 2013 on the capital market at 2.35 euro per share, generating a total of 256 million euro. Thus, the Portuguese state has earned a total of 3.056 billion euro since 2011.

In recent years, Energias de Portugal has increasingly specialized in generating electricity from renewable energy resources. The focus on this promising market is also reflected positively in the balance sheet figures. Sales rose from 6.4 billion euro in 2002 without any significant setbacks to 16.3 billion in 2014 [5]. The net profit after 2005 [6], however, remained constant at around one billion euro. The value of the firm's equity positively increased, as did that of turnover. Equity grew from 5.5 billion euro in 2002 to 12 billion euro in 2014. It is also worth mentioning that the increase in turnover was accompanied by a decline in the number of employees. After a count of 18,455 in 2002, the number of employees decreased by an average of 3.7 percent per year, reaching 11,798 by 2014.

EDP paid a dividend to its owners during the entire period under review. Initially this amounted to 11.3 cents per share (2002) and eventually rose to 18.5 cents per share, remaining at this amount ever since 2011. Because the shares of Energias de Portugal were already traded on the stock market in the entire period prior to privatization, the stock return is based on the share price yield and dividend yield. On average, the sum of the two returns from 2002 to 2011 was 4.4 percent per year. Negative deviations from this trend particularly affect the years 2002 as well as 2008, with a negative overall return of 34 and 37 percent, respectively. The net yield reached its highest value in 2006 at a value of 51 percent. During the privatization of 2011, the combined return from the share price and dividends was only 2.8 percent, but it was able to recover thereafter and rose to 27.5 percent in 2014.

4.1 demonstrates that since 2002, the return on shares has largely exceeded the lending rate for government bonds with a two-, five-, and ten-year maturity. However, the particularly weak years of 2002 and 2008 mean that the average net yield per year is slightly negative. For government bonds with a two-year maturity the net yield is at -0.24 percent, for government bonds with a five-year maturity at -0.9 percent, and considering government bonds with a ten year maturity at -1.3 percent. For the Portuguese state, however, participation in Energias de Portugal as a debt-financed investment was nevertheless profitable until privatization. If the share price yields are not taken into account and if only the dividend yield is considered, it is concluded that the net yield per annum is an average of one percent for all maturities.

Neither the balance sheet figures nor the comparison of the stock yields with the loan interest rates for Portuguese state bonds clearly indicate why the remaining 25 percent of shares in possession of the state were selected for sale. Although the net yield is negative on average for all maturities, a clear statement is hardly possible, partly due to the very slight discrepancy from zero. The company was profitable under state participation and continuously increased its turnover and the value of its equity. In addition, since the dividend yield averaged above the lending rate for

[5] See Appendix 4 for the balance sheet figures.

[6] Due to a significant restructuring of financial assets, values prior to 2005 are not comparable.

government bonds with different maturity dates, the long-term positive effect of this privatization on the state budget is rather questionable. Despite appearing to be only one of six private enterprises in this analysis, Energias de Portugal accounted for one third of total privatization revenues.

Transportes Aéros Portugues (TAP)

Transportes Aéreos Portugueses was founded in 1945 as a state-run airline, being privatized for the first time in 1953. Following the Carnation Revolution of 1974, TAP was renationalized in 1975. There were initial considerations on the privatization of the airline starting October 2002, but no sale took place for the time being. A first privatization attempt as part of the reform program failed in December 2012 due to the potential buyer's insufficient financial solvency. In June 2015, 61 percent of the shares were successfully sold to a Brazilian-American investor for 10 million euro. He also promised to invest at least 338 million euro into the airline over the next few years. In addition, 5 percent of TAP shares were sold to employees at a preferential price.

The reason for this exceptionally low purchase price of 10 million euro can be found in the company's balance sheet. Since 2002, Transportes Aéreos Portugueses has recorded a net loss of almost 40 million euro per year. The biggest loss dropped 288 million euro in 2011, and the highest profit amounted to only 8.7 million euro in 2004 [7]. In the final year prior to privatization, the company recorded yet another net loss of 85 million euro (see 4.2). The development of the equity value reflects the permanently poor earnings performance of TAP. From 14 million euro in 2002, the value of its assets, after deduction of all debts, continually fell to a negative value of more than 500 million euro in 2014. Since Transportes Aéreos Portugueses does not pay dividends due to lack of profitability, the comparison with the interest rate for Portuguese government bonds on various maturities is omitted. Sale proceeds of 10 million euro for 61 percent of shares are so low that no significant effects are expected in any case by reducing sovereign debt. However, as shown in the literature, there are other mechanisms through which the sale of TAP could have positive effected the national budget. If the company is more profitable as a result of privatization, higher tax revenues could be expected. Moreover, transfers, such as in the form of debt guarantees, are likely to be omitted.

Overall, TAP seems to be a highly deficient state enterprise, whose privatization is long overdue. The privatization should, hence, alleviate the national budget and the state balance sheet.

[7] Transportes Aéros Portugueses calculates figures using the IFRS/IAS system only since 2006. Previously, accounting was carried out in accordance with Portuguese trade law.

5 Conclusion

The aim of this work was to explore the possible reasons for the privatization of certain companies and whether these decisions could have a positive effect on the state budget. The basic consideration for the latter analysis was a comparison of the corporate return rate with the government bond interest rate. In the event that the corporate return is lower than the interest rate, the state pays more interest on the amount of the investment than the investment yields. Sale and reduction of sovereign debt by the amount of the proceeds would thus have a positive effect on the national budget.

Four out of six companies [8] can be described as profitable or economically sound, but not characterized as sufficiently profitable. The latter may have been the reason for their privatization. It is therefore expected that the state budget will be relieved by the sale both short and long run. Exceptions are the state carrier Transportes Aéreos Portugueses (TAP) as well as the energy provider Energias de Portugal (EDP).

Since 2002, TAP has recorded an average loss of almost 40 million euro per year and has therefore never been able to distribute a dividend. A look at the development of the equity value, it becomes clear that the company should have been sold earlier. Plans to do so existed since 2002.

EDP, on the other hand, has been exceedingly profitable for a long time. The company paid dividends since at least 2002. Together with the stock price yield, this gives an average return on equities of 4.4 percent per year for the ten years leading up to privatization. Although interest rates on Portuguese government bonds, as in other companies, were significantly higher than the annual stock returns, the dividend yield alone exceeded the interest rate for bonds of various maturities by one percent per year. It is questionable whether the decision to sell the remaining state shares will have a long-term positive effect on the public budget. This case is particularly relevant in light of the fact that the proceeds from the sale of EDP shares account for one third of the total privatization income.

Some methodological problems arise in this type of analysis. For public-sector companies whose shares are not traded on a stock exchange, it is sometimes difficult to determine the company value for the periods prior to privatization. For the sake of simplicity and to calculate the dividend yield, it was assumed here that the company could have been sold at the same price in the preceding periods. This assumption, however, is a rather inaccurate approximation. In addition, effects such as increased tax revenues after such a privatization or the welfare gains resulting from increased competition are disregarded because they are difficult to quantify. Taking on these

[8] Aeroportos de Portugal (ANA), Caixa Seguros e Saúde, Correios de Portugal (CTT), and Redes Energéticas Nacionais (REN)

difficulties could be the basis for further research. On July 9, 2015, the Portuguese government began the privatization process for the rail transport enterprise CP Carga as well as for Empresa de Manutenção de Equipamento Ferroviário (EMEF) [9], responsible for the maintenance of the Portuguese rail system [10]. In the past, both companies had problems being profitable and therefore did not distribute dividends. Assuming an efficient sale price, an analysis on this topic may also come out in favor of privatization.

6 Tables

Company	Sector	Shareholding of the State
Sale of the State's shareholding in full		
BPN	Financial	100.0%
INAPA – Investimentos, Participações e Gestão, S.A.	Paper	32.7%
Edisoft	Defence	60.0%
EID		38.57%
Empordef TI		100.0%
Sociedade Portuguesa de Empreendimentos SPE. S.A.	Mining	81.1%
Hidroeléctrica de Cahora Bassa, S.A.	Energy	15.0%
Partial sale of the State's shareholding		
GALP Energia, SGPS, S.A.	Energy	8.0%
EDP – Energias de Portugal, S.A.		25.73%
REN – Redes Energéticas Nacionais, S.A.		51.08%
Estaleiros Navais de Viana do Castelo, S.A.	Shipbuilding	100.0%
CP – Carga, S.A.	Transport	100.0%
TAP, SGPS, S.A.		100.0%
CTT – Correios de Portugal, S.A.	Communications	100.0%
ANA – Aeroportos de Portugal, S.A.	Transport	100.0%
Caixa Seguros	Financial	100.0%
EMEF – Emp. de Manutenção de Equip. Ferroviário, S.A.	Transport	100.0%
Concessions		
CP - operation of routes	Transport	100.0%

Table 4.1: Privatization under the Portuguese Stability and Growth Program 2009-2013

[9] Both firms are under 100 percent state ownership
[10] An earlier Privatization was not possible due to unfavorable market conditions.

Company	Sector	Date of sale	Number of shares sold (10^6)	Equity share sold	Preis/ Aktie in €	Proceeds (10^6€)	Buyer	Remaining public equity share
CP Carga	Transport	7/9/2015	Portuguese government is initiating the privatization process					100%
EMEF	Transport	7/9/2015	Portugiesische Regierung beginnt Privatisierungsprozess					100%
Transportes Aéros Portugueses (TAP)	Transport	6/1/2015		61%		10	Gateway Consortium	39%
Empresa Geral do Formento (EGF)	Waste management	9/17/2014	11.2	100%	13.38	150	SUMA (span./port.)	0
Correios de Portugal (CTT)	Communikations	9/5/2014	47.2	31.50%	7.26	343	Stock market	0
Redes Energéticas Nacionais (REN)	Energy	6/17/2014	58.7	11%	2.68	157	Stock market	0
Caixa Seguros e Saúde	Financial	5/1/2014	320	80%	3.13	1000	Fosun Int. (chin.)	20%
Correios de Portugal (CTT)	Communikations	12/5/2013	102.8	68.50%	5.63	579	Various institutions	31.50%
Aeroportos de Portugal (ANA)	Transport	9/17/2013	40	100%	77	3080	Vinci Group (frz.)	0
Energias de Portugal (EDP)	Energy	2/19/2013	151.5	4.14%	2.35	356	Stock market	0
Hidroeléctrica de Cahora Bassa	Energy	4/1/2012		7.50%		38.4	Redes Energéticas Nacionais (REN)	0
Banco Portuguès de Negócios (BPN)	Financial	3/30/2012				40	Banco BIC Portugûes	0
Redes Energéticas Nacionais (REN)	Energy	2/2/2012	133.4	25%	2.9	387	China's State Grid Corporation	11%
Redes Energéticas Nacionais (REN)	Energy	2/2/2012	80.1	15%	2.56	205	Oman Oil Company	11%
Energias de Portugal (EDP)	Energy	12/22/2011	780.6	21.35%	3.45	2700	China Three Gorges Corporation	4.14%
Hidroeléctrica de Cahora Bassa	Energy	11/1/2011		7.50%		38.4	The government of Mosambik	7.50%
		4/7/2011	Portuguese government applies for financial assistance					

Table 4.2: Privatization since 2011

7 Figures

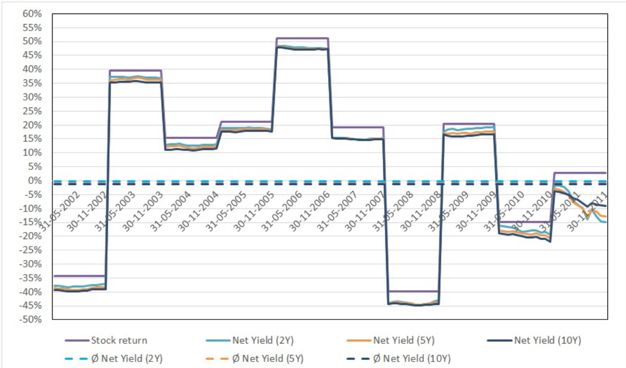


Figure 4.1: Energias de Portugal: Stock return, net yield and average net yield, 2002-2011

Figure 4.2: TAP: Net income, total equity

Online Appendix to: Sovereign debt reduction through privatization: The case of Portugal by Lukas Nüse

LASSO-Based Forecasting of Financial Time Series on the Basis of News Headlines

Adrian Waltenrath*

1 Introduction

In this paper, I carry out an interdisciplinary approach which is rather new and has drawn increasing attention recently. Since, theoretically, news articles contain all relevant information affecting stock prices, it is reasonable to use them to predict stock market volatility. All events causing stock price movements should be reflected in a news story and therefore be incorporated in the data. A challenge lies in the processing of news from their textual form to numerical quantities, which can be handled by mathematical methods to detect potential relations.

This approach contains elements from different fields of study such as finance, econometrics and computer sciences. Two subfields of the latter are especially interesting in this context: natural language processing and machine learning. Natural language processing generally deals with the processing of human language in its spoken or textual form by computers, while machine learning addresses the development of methods to categorize observations and/or recognize patterns for prediction. As pointed out by Varian (2014), machine learning offers a variety of methods which can be beneficial to econometricians in related applications. Although the focus of this paper lies on the econometrics involved, I borrow and apply methods developed by the other fields.

From the perspective of a computer scientist, the given problem is tackled in a rather basic way: I use headlines to predict market volatility by simply counting the number of occurrences for single words-stems over time. This means that I calculate the *term frequency* (TF), while term, in this paper, refers to a single word-stem.[1] Using single

*Before entering the master's programm of the University of Bonn, Adrian Waltenrath earned a B.Sc. in Economics from the University of Mannheim. He received his M.Sc. in Economics in fall 2015 and is now employed at DZ BANK AG.

[1] Since every field of study has its own special vocabulary it is not always trivial to please all of them. The expression *term frequency* (TF) is commonly used in relevant literature to describe the number of occurrences in a document or a set of documents.

words implies dividing texts into collections of words without taking into account any linguistic connections between them. This technique is common in relevant literature and called *bag-of-words* or *1-grams*. The reason to focus solely on headlines follows the argumentation that they are more to-the-point than news stories as they have a higher proportion of signal words, which might have explanatory power (Peramunetilleke and Wong 2002, Huang et al. 2010 and Nassirtoussi et al. 2015).

2 Literature Review

Due to the limitation in space and since most of the research done on this topic tackles the problem from a different perspective than I intend to, I cease from an extensive review. Detailed reviews are provided by Nassirtoussi et al. (2014), Hagenau, Liebmann, and Neumann (2013) as well as Nikfarjam, Emadzadeh, and Muthaiyah (2010). In addition, an overview can be found in appendix 8.1

3 Data Description

3.1 VIX

Data on opening and closing prices are taken from Datastream for every trading day from Jan 1, 2014 to Jul 31, 2015. The CBOE Volatility Index (VIX) measures the implied volatility of S&P 500 and is computed by the Chicago Board Options Exchange. It is quoted in percentage points and intends to estimate the annualized expected volatility of the S&P 500 within the next 30 days.[2]

3.2 News

News headlines are taken from *The New York Times*. They are gathered via the *New York Times Article Search API*[3]. To reduce noise I only gather news articles belonging to the sections *World*, *Business*, *Business Day* and *U.S.* I gather data

[2] Further information on the construction of the VIX can be accessed via http://www.cboe.com/micro/vix/vixintro.aspx. Detailed information aubout the S&P 500 can be found in the methodology document available under http://us.spindices.com/indices/equity/sp-500.

[3] The New York Times (2015): http://developer.nytimes.com/docs/read/article_search_api_v2.

from Jan 1, 2014 to Jul 31, 2015[4]. In total 243 750 articles are collected.[5]
As mentioned before, the news have to be processed in a way that maps words into numbers. To deal with the structure of natural language, a few more steps have to be performed, aiming to keep noise at a minimum: First, all headlines are converted to lower case. Second, all special (meaning all non-alphabetic) characters are deleted. The remainder can be viewed as a vector of lower-case words for every headline. From this vector I remove all so-called stopwords. The list of stopwords applied in my analysis is taken from the R-package *tm* (Feinerer and Hornik 2015) and is presented in Appendix 8.2. Next a stemming algorithm is applied. A common choice is the algorithm created by Porter (1980) which maps words back to a stem by applying transformations to the words suffix. This algorithm is commonly used in literature and has proven to work reasonably well. The algorithm is applied via the R-package *SnowballC* (Bouchet-Valat 2014), an example of its performance is shown in the appendix (reproduced online) .
Finally, the stemmed words were counted.[6] In total, there exist 47 023 different stemmed words, leaving me with the same number of potential predictors to include in the model. Since I try to forecast the closing price prior to market opening and trading hours are from 9:30 a.m. to 4:00 p.m. Eastern Time[7], I treat the period from 9:30 a.m. (the day before) to 9:29 a.m. as one time interval.

3.3 Holidays and Weekends

Holidays and weekends are ignored. That means I always use the news released within 24 hours prior to market opening, no matter if the prior day has been a trading day or not.[8]

[4]From Jan 1, 2014 onwards news releases increase heavily due to the inclusion of additional sources.

[5]Headlines are not always unique. Sometimes an update on the news is performed, leading to a repost of the same news. In addition, there are recurrent articles, each time having the same headline and different content. In order to avoid noise, articles whose exact headline occurs ten or more times over the whole period are deleted. In addition, I delete news if the same headline already occurred in a period of seven days prior to the news-release.

[6]Tables presenting the most frequently occurring words as well as empirical quantiles of frequencies are provided in Appendix 8.4.

[7]UTC−5 in winter and UTC−4 in summer. As illustrated in Appendix 8.3, time originally measured in Coordinated Universal Time (UTC) is converted to Eastern Time (ET) while the news were processed.

[8]This includes the assumption that news lying more than 24 hours in the past are already fully incorporated in the opening price, whereas news from the past 24 hours are assumed to have some predictive power for the performance over the upcoming trading day.

4 Methodology

4.1 LASSO

When trying to estimate the impact of the different (stemmed) words on financial time series, a high-dimensional problem is created. Reducing dimensions to a moderate number of explanatory variables that can be assumed to have predictive power is crucial in my analysis. This is done by applying different variations of the LASSO (*least absolute shrinkage and selection operator*): The standard LASSO which was proposed by Tibshirani (1996), the relaxed LASSO by (Meinshausen 2007) as well as the adaptive LASSO by Zou (2006). The adaptive LASSO is carried out in two variations. One uses the first stage estimates as weights during the second step (aLASSO-L), the other uses OLS estimates calculated on the subset which is selected by the first stage LASSO-regression (aLASSO-O). In Addition I analyse the performance of a simple OLS forecast based on the subset selected by the first stage LASSO-regression (LASSO-OLS). The mathematics of these procedures are described in more detail in appendix 8.5.

4.2 Parameter Selection: Cross-Validation

It is a common approach to determine the tuning parameters by cross-validation (CV). CV in general is considered, e.g., in Hastie, Tibshirani, and J. Friedman (2009) and Arlot and Celisse (2010). Given the time series character of the data at hand, its application is not trivial. The topic of CV in a time-series environment with dependent data is extensively studied by Bergmeir and Benítez (2012). Although they do not find any practical issues with standard k-fold CV, they suggest to use a blocked form of k-fold CV and to additionally control for stationarity. In my analysis non-stationarity is not an issue, since all variables are assumed to be stationary. This assumption is supported by performing augmented Dickey-Fuller tests[9] (ADF-tests) on the VIX-returns. In addition, ADF-tests are performed for the ten most frequent word stems as well as for ten more words, which are randomly drawn. All tests reject non-stationarity at 1% such that the assumption of stationarity is justified. Following Bergmeir and Benítez (2012) I implement a blocked form of k-fold CV, while dropping 20 observations at the borders of each training set. This is done to obtain approximate independence between folds. They argue that the presented method makes full use of the data[10], while – by retaining the time-series structure –

[9] Tests are performed with seven lags. This is the lag length chosen by default by the R-package *tseries* (Trapletti and Hornik 2015), which is used to compute the test-statistics.

[10] In contrast to the use of a single block as testing set. This is another method considered by Bergmeir and Benítez (2012).

delivering robust error estimates. The implemented procedure is outlined in detail in the appendix 8.6.

Choosing the Cross Validation Parameter: Bias-Variance Trade-off

It is widely known that, when applying k-fold CV, there exists a trade-off between bias and variance as a small k gives upward biased error estimates possessing a low variance, whereas a large k reduces bias at the cost of a higher variance (c.f. Hastie, Tibshirani, and J. Friedman 2009 or James et al. 2013). Leave-one-out CV (LOOCV) with $k = N$ delivers unbiased error estimates but suffers from high variance and thus possibly leads to a poor choice of λ. In addition, LOOCV is computationally intense since the model is fitted $k = N$ times on each of the training sets. By choosing a smaller value for k, the computational burden is reduced proportionally.
The variance of the error estimates increases in k because of the increasing similarity of the training sets: If k is chosen large, less observations are removed for the construction of the training sets, which leads to greater overlapping between any two training sets. Therefore, as k approaches N, the estimated models become very similar and the CV-error is computed as the average over positively correlated quantities and hence possesses a higher variance than the average computed from less-correlated quantities. As pointed out by Hastie, Tibshirani, and J. Friedman (2009), common choices are $k = 5$ and $k = 10$ since they have shown to provide a reasonable balance between bias and variance in empirical applications.
In fact, the number of folds is crucial in my analysis as the tuning parameter is extremely sensitive to the assignment of the folds. I therefore pay serious attention to the selection of k in Section 6.1.

4.3 Model Setup

Forecasts are computed for each trading day from Jan 1, 2015 to Jul 31, 2015, which results in 146 predictions. The model is re-estimated at each prediction date using all observations of the previous 12 months. Depending on the trading days, this gives a database of 250 to 252 observations to estimate the model on. The model horizon of 12 months as well as the prediction horizon are chosen rather arbitrarily. Nevertheless, given the fact that, becuse of the smaller amount of data, it is not feasible to make use of the news before Jan 1, 2014, I argue that this is a reasonable choice. In Section 7, I discuss this choice critically.
Re-estimating the model for each prediction date is computationally very intense but necessary since, due to the instability of the tuning parameter, it is not appropriate to estimate the model just once and apply the same model over the whole prediction horizon.

In this case, one lucky (or unlucky) result for the optimal tuning parameter could heavily bias the analysis. In addition, I argue that the focus of financial markets changes over time such that the predictive power of some features is changing as well. It is thus necessary to continuously re-fit the model.

5 Simulation

To verify that the presented methods can detect a small set of meaningful variables within a huge amount of noise, I conduct a simple simulation which is shown in the appendix (reproduced online). In short, these results show that the proposed methods can indeed detect the majority of true predictors within a vast amouunt of noise. However, the selected models are too large as they also pick some of the noise variables. Still, as most coefficients estimated for these noise variables are small and alternate around zero, the estimated models are expected to have some predictive power.

6 Results

6.1 Choosing the Cross Validation Parameter: Sensitivity of the Tuning Parameter

As mentioned in Section 4.2 the optimal tuning parameter is sensible to the assignment of the folds and therefore to the value chosen for k. Shifts in the fold assignment can lead to very different results. The problem of the instability of the LASSO procedure for $p \gg N$ is assessed by Zhang and Yang (2015) as well as Roberts and Nowak (2014). Their recommendations are not applicable in the context of blocked CV but briefly discussed in Section 7. Zhang and Yang (2015) state that, in highly instable cases with $p \gg N$, bias increases severely for small k, while variance decreases monotonically. They argue that choosing $k \in \{5, 10\}$ as a general rule can be misleading and find that, in these cases, $k \leq 10$ can perform significantly worse than LOOCV, i.e. $k = N$.
Another perspective, which should also be taken into account, is the available number of observations. Since I deal with a relatively small dataset (250 to 252 observations), it might not be adequate to choose a small k because this results in training sets which are considerably smaller than the set the model is finally estimated on. As described in Hastie, Tibshirani, and J. Friedman (2009, p. 243), the choice of k depends on the *learning curve* of the model. It would be adequate to choose $k = 5$ if the model estimated on 200 observations performed nearly as good as the model estimated on 250 observations. On the other hand, if the model performed quite

poor for 200 observations, but notably gained performance from the additional 50 observations, it would be appropriate to choose a large k. The drawback is that a larger k comes with a higher variance. To assess this issue for the application at hand, I take a look at the CV-error-curves for different k. Appendix 8.10 shows error curves from the (first stage) standard LASSO procedure for a subset of randomly drawn dates. Theoretically, one would expect the error curves to be lower with increasing k since the bias decreases. In turn, error curves are expected to be instable for high k due to the increasing variance. The Figure partly confirms these expectations. Bias seems to drop for $k > 5$, whereas it is hard to detect any decrease for $k > 20$. Since none of the curves are highly instable, I argue that $k = 20$ provides a reasonable balance between bias and variance at a moderate level of computational costs. This is in line with the findings of Zhang and Yang (2015). In addition – since I deal with a relatively small number of observations – models are estimated on a considerably larger database as for $k \in \{5, 10\}$. Obviously, the presented error curves only constitute a small fraction of the 146 prediction dates. The remaining error curves look similar and allow for the same interpretation.

6.2 Empirical Results

In this section, I present the results obtained when the analysis described in the previous chapters is carried out to forecast the VIX. All results presented are computed under the *mean absolute error* (MAE) loss, the corresponding results for *mean squared error* MSE are shown in the appendix. When comparing results for MAE and MSE, MAE seems to prevail. This might originate from the fact that the MAE weighs small and large deviations equally and is thus less affected by outliers. Such outliers can be observed in case of some event whose market impact dominates all other effects. This scenario is not unlikely for the given application of stock price volatility. Following this argumentation, using the mean absolute error is considered the better choice since it is robust to these situations.
Since the proposed approach uses the *bag-of-words* technique, it is not able to capture any semantics. Take, for example, the word *sanction*, which can have a positive or negative impact depending on its context (whether sanctions are tightened or eased). In any case, *sanction* is expected to cause volatility. 5.1 shows the proportion of correct directions as well as the hypothetical profit achieved, when investing according to the predicted directions at the opening price and evening out the position each day at the closing price. Note that this profit is purely hypothetical since it is not possible to directly invest in the VIX.[11] Still, it helps to evaluate the

[11] The VIX is indirectly investible via VIX-futures or via buying/selling options on the S&P 500. Both strategies do not produce a perfect correlation with the VIX. In addition, VIX futures possess a negative roll yield which causes additional costs. Constructing an option-based strategy is also non-trivial and beyond the scope of this paper.

Table 5.1: Results under MAE

	Proportion of Correct Directions				Hypothetical Profit in %			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	63%	62%	63%	61%	134.5	114.6	137.5	97.7
LASSO-OLS	62%	61%	62%	62%	122.7	90.3	128.5	133.4
rel. LASSO	63%	62%	60%	62%	123.9	98.1	111.3	123.0
aLASSO-O	64%	61%	62%	60%	136.5	90.3	131.4	112.5
aLASSO-L	62%	60%	61%	62%	113.7	88.1	122.0	97.1

prediction system since in contrast to the proportion of correct directions it is not purely binomial.
The system predicts the correct direction in at least 58% percent of the cases. The hypothetical profit is positive but, as explained, can never be achieved in practice. Theoretically, a long-term investment in the VIX generates a performance of −31.76% over the whole horizon. Buying the VIX each morning and selling it in the evening yields −142.04%. A naive trader, short-selling the VIX, could therefore generate a profit of 31.76% from a long-term investment and 142.04% from investing repeatedly each morning. She would be correct in the sense of directions in 64% of the cases. As it can be seen in 5.1, the proposed system can outperform the long-term investment but hardly beats the 64% achieved by the naive trader.
The estimated model sizes are presented in 5.2. Introducing a threshold or investing only if a non-degenerate model is estimated does not improve performance. Nevertheless, I take a closer look at the performance of the non-trivial models[12]. The

[12]Tables illustrating the performance for investing with a threshold of 0.8% are presented in Appendix 8.11.

Table 5.2: Model Size (MAE)

	5 Folds	10 Folds	20 Folds	40 Folds
std. LASSO	3.56 - 55/146	3.29 - 59/146	3.32 - 59/146	3.11 - 59/146
LASSO-OLS	3.56 - 55/146	3.29 - 59/146	3.32 - 59/146	3.11 - 59/146
rel. LASSO	3.54 - 55/146	3.20 - 54/146	3.32 - 59/146	3.11 - 59/146
aLASSO-O	3.03 - 53/146	2.98 - 52/146	3.03 - 58/146	2.86 - 57/146
aLASSO-L	3.16 - 55/146	3.02 - 53/146	3.06 - 58/146	2.87 - 57/146

This table summarizes the estimated model size for different k. The first value corresponds to the average number of non-zero coefficients (including the intercept). The value after the minus sign shows the number of times a non-trivial model (with at least one additional predictor) is estimated. 146 is the length of the prediction horizon.

Table 5.3: Results under MAE for Non-Degenerate Models

	Proportion of Correct Directions				Hypothetical Profit in %			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	64%	63%	63%	56%	46.3	30.3	57.4	17.4
LASSO-OLS	62%	59%	59%	58%	34.5	6.0	48.4	53.1
rel. LASSO	64%	59%	56%	58%	35.6	9.0	31.2	42.7
aLASSO-O	66%	60%	59%	53%	60.6	15.3	47.4	37.7
aLASSO-L	62%	57%	57%	58%	25.5	0.2	37.9	22.3

inferior performance is not surprising since the non-degenerate model invests in less than half of the trading days. In addition, the degenerate model always recommends a short position[13], which is correct in the majority of cases. In terms of correct directions, the non-degenerate models perform worse than the 64% achieved by the naive trader for most specifications. Still, it can beat the benchmark of a long-term investment in some cases. For further analysis, I focus on one of the presented specifications. Although much randomness is involved, I argue that the aLASSO-O method performs well over all considered k. In addition, aLASSO-O yields a good performance during the simulation.
As pointed out in Section 6.1, $k = 20$ provides a reasonable balance between bias and variance. Although the choice of $k = 5$ gives better results for the VIX, I stick to $k = 20$ as this choice is better founded and expected to suffer from less instability. The good results for $k = 5$ are suspected to be coincidental. The choice is in line with the findings of Zhang and Yang (2015), who investigate CV in the context of the LASSO for the case of $p \gg N$.

6.3 Predictors and Estimated Coefficients

In this section, I take a closer look at the estimation results of the aLASSO-O method with $k = 20$ under the mean absolute error loss. 5.1 summarizes all predictions made by this system. It also illustrates whether predictions are based on a degenerate model and whether the direction is predicted correctly. Interestingly, at each prediction date before May 1, 2015 a degenerated model is estimated. Keeping in mind that estimation is always carried out on the last 12 months prior to the prediction date, this cannot be led back to an increasing database. Instead, it implies that during the first four months the system is not able to detect any meaningful predictors from the given database. Looking at the performance of the VIX, which is presented in Appendix 8.12, does not show any peculiarities in the dependent variable that could have dropped out or joined the database around May 1, 2014 or May 1, 2015,

[13] It always predicts a value in $[-0.8, -0.1]$.

respectively. Instead, this has to be interpreted as the result of a process. Possibly, at that point, enough information about some topic(s) joined the database such that a pattern is recognized and a non-degenerated model is estimated.
It is also possible that the effects of topics change over time, which lowers the predictive power of the corresponding feature and makes it harder to detect a pattern. Take, for example, the stem *ukrain*: Surprisingly, a negative impact of this feature on volatility is estimated if it is included in the model. Nevertheless, the impact of the corresponding news obviously heavily depends on the context and for sure has not always been negative over the last year. It is likely that at the beginning of the Ukraine crisis the stem *ukrain* was a driver of volatility and that it adopted a calming affect in the recent past as the crisis passed its climax such that news were reducing, rather than causing, uncertainty.
According to this argumentation, a feature's impact can depend on the context such that it cannot easily be detected by the given approach. This is a drawback and discussed in Section 7. Another issue is the twelve-month calibration interval, which is a long horizon in fast-moving financial markets. I again refer to Section 7, where I discuss this issue in detail.
So far, I have only investigated one feature, namely *ukrain*. 5.4 shows the number of times each feature is included in the model along with some more detailed information. Additionally, 5.2 illustrates the estimated coefficients for the eleven most frequent features over time. The feature which is most often included in the model, is *obama*

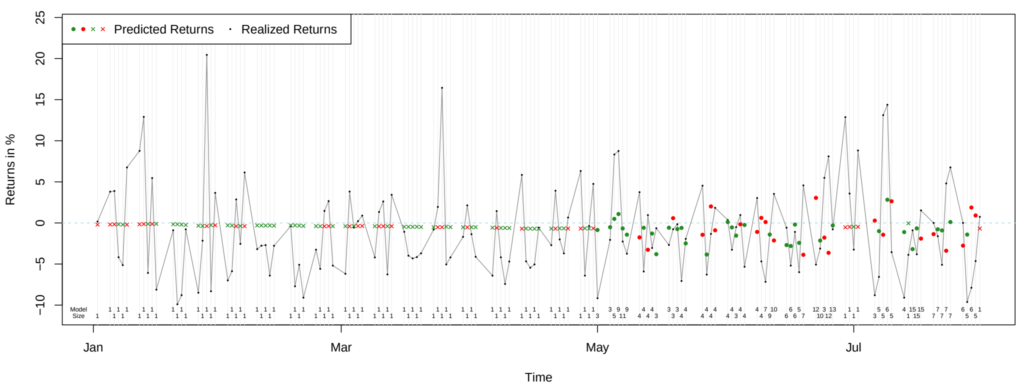


Figure 5.1: This figure shows the true realized returns (black) of the VIX along with the predictions (red/green) created by the aLASSO-O procedure for $k = 20$ under the absolute error loss. Colors indicate whether the direction is predicted correctly. Predictions are marked with dots if they come from a non-trivial model. An X is drawn if the underlying model is degenerated. In addition, at the bottom of the figure, the size of the estimated model is presented. Vertical lines are drawn at each date a model is estimated.

Table 5.4: Frequency of Features and Summary of Corresponding Coefficients

Feature	Freq.	Pos.	Neg.	Min.	Max.	Avg.	Corresponding Words
(Intercept)	146	9	137	−3.51	0.50	−0.84	
obama	58	0	58	−0.25	−0.13	−0.19	obama, obamas
sanction	49	49	0	0.25	0.47	0.4	sanctions, sanction, sanctioned, sanctioning
report	36	36	0	0.11	0.28	0.23	report, reports, reported, reporting, reporter, reporters
leader	25	25	0	0.31	0.45	0.38	leader, leaders
ukrain	18	0	18	−0.18	−0.06	−0.13	ukraine, ukraines
china	17	17	0	0.04	0.18	0.13	china, chinas
iran	17	0	17	−0.16	−0.01	−0.05	iran, irans
crash	14	14	0	0.05	0.12	0.09	crash, crashes, crashing, crashed
gaza	14	14	0	0.08	0.19	0.12	gaza, gazas
iraq	12	0	12	−0.24	−0.08	−0.18	iraq, iraqs
greek	10	0	10	−0.17	−0.10	−0.15	greek, greeks
death	5	0	5	−0.19	−0.15	−0.17	deaths, death
mai	5	0	5	−0.32	−0.29	−0.31	may, mais, mays
polic	5	0	5	−0.10	−0.07	−0.09	police, policing, polices
deal	3	0	3	−0.07	−0.07	−0.07	deal, deals, dealings, dealing
japan	3	0	3	−0.22	−0.21	−0.21	japan, japans
vote	3	0	3	−0.13	−0.12	−0.12	vote, votes, voting, voted
cuba	1	0	1	−0.17	−0.17	−0.17	cuba, cubas
take	1	1	0	0.40	0.40	0.40	takes, take, taking
u	1	0	1	−0.03	−0.03	−0.03	us, u

This table shows the number of times (stemmed) words are included in the model. It also shows the number of times the estimated coefficients are positive or negative. In addition, it shows the maximum, minimum and average of all coefficients for each word.

and has a negative impact on volatility. This is not immediately intuitive as a U.S. president's actions or wording could have severe effects on financial markets. On the other hand, Obama as well as the American government is certainly not interested in highly volatile markets, especially not when facing a period of fragile economic growth. It is therefore reasonable that he might have chosen his actions and wording to reduce uncertainty, enforcing markets to stay calm. According to the presented results this has been successful to some extent.

Other features with negative coefficients are *ukrain* and *greek*. As pointed out before, this has not been expected beforehand. The negative coefficients suggest that news including the corresponding words were reducing uncertainty rather than containing

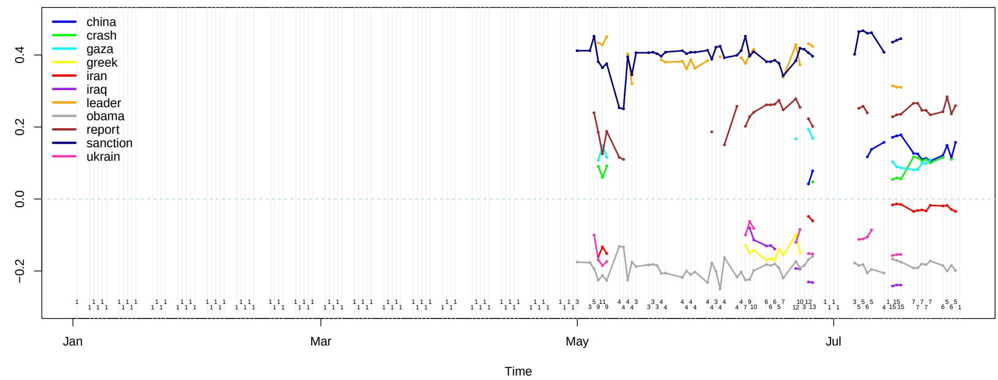


Figure 5.2: This figure illustrates the coefficients for the eleven most frequent features estimated by the aLASSO-O procedure for $k = 20$ under the absolute error loss. At the bottom of the figure the size of the estimated model is presented. Vertical lines are drawn at each date a model is estimated.

new, unanticipated information. However, both features are not persistently included in the model, such that one should not give too much credit to this interpretation. As explained above, the impact of those features on volatility is probably very context dependent. This might contribute to the poor prediction performance.

Among others, positive coefficients are estimated for the features *sanction*, *report* and *leader*. The positive coefficients of *sanction* are plausible and most likely connected to the Ukraine crisis and the sanctions which were introduced by various countries against the Russian government. Sanctions – no matter if they are introduced, tightened or eased – should cause volatility as they raise uncertainty about their economic impact on all involved countries. They should therefore have a positive impact on volatility.

The other features are harder to interpret. The feature *report* corresponds to a greater number of words. It includes forms of the verb *to report* as well as the nouns *reporter*, *reporters* and *report*. Here, another drawback of the approach arises. The words *reporter* and *reporters* do not fit to the other words that can be linked to the release of a financial report or some economic key figures, causing the positive coefficient. The words *reporter* and *reporters* are expected to be mentioned in a different context such that noise is caused.

At first sight, the feature *leader* is puzzling. Looking at some corresponding headlines shows that these words are often used for the leader of some country or organization like for instance *Iran Leader*, *E.U. Leaders* and *Greek Leader*. Thus, *leader* typically refers to an influential person or groups of persons, who can affect financial markets by their actions and statements and therefore cause volatility. Interestingly, *leader* in general seems to cause volatility, whereas *obama* reduces it. Some reasoning for this has been given above when assessing the intention of the American government. Since the remaining features are less persistent and/or possess smaller coefficients, I cease from a detailed interpretation. In general, a positive coefficient indicates that the topic has been a driver of volatility and that the corresponding news raised uncertainty. In turn, a negative coefficient implies a calming effect such that news containing the corresponding words reduced uncertainty and did not contain much unexpected information. As mentioned above, one should keep in mind that the coefficients are estimated on data from the past 12 months, such that recent shifts in the impact of features can hardly be reflected by the coefficients.

7 Criticism and possible Extensions

At various points I have been referring to this section, in which I want to discuss the approach critically. Due to the approach's interdisciplinarity there exists a wide range of possibilities for modifications and improvements.

First and foremost, there are two factors which are key in my opinion: The prediction interval and the horizon the model is calibrated on. Both characteristics are obviously connected and cannot be chosen independently as the length of the prediction interval affects the number of observations in the calibration horizon.

In my opinion, the poor prediction performance likely originates from the long prediction interval and the large model horizon. Financial markets are fast-moving and twelve months seem to be a long time for patterns to persist. Also, markets, especially those as liquid as the S&P 500, are fast in processing new information such that a prediction interval from market opening to market closing is probably not adequate. The same holds for the interval that allocates the headlines to the prediction dates. 24 hours seem to be a lot if markets are fast-moving. It is likely that news released in the beginning of the interval are already fully incorporated in the opening price.

Therefore, it would be interesting to repeat the analysis with smaller intervals, such as 1 hour or even 30 minutes, while taking the news released during the prior interval as independent variables. This modification yields up to 13 observations per trading day. In turn, it drastically decreases the number of news falling into each of the intervals such that it would be necessary to find another source which is providing more frequent news releases[14]. This increase in observations per day allows to shorten the prediction horizon drastically. In this paper, the model was calibrated on a 12-month-horizon with 250 to 252 observations. With 13 observations per day, twice the amount can be reached by using data from the prior two months. This larger number of observations could be beneficial for the estimation of the

[14]A candidate could be the news-ticker provided by Bloomberg terminals as it is customizable and delivers real-time news from various sources.

tuning parameter and improve stability. In addition, the shorter calibration horizon is expected to better reflect the fast-moving character of financial markets, where focus can move rapidly and topics' impacts might change over time. Whether such a system is able to yield a better prediction performance is still to be investigated.
Next, I turn to the instability of the CV procedure. Some research has already addressed this topic. For one, there exists the one-standard error rule (Hastie, Tibshirani, and J. Friedman 2009, p. 244), according to which one should choose the largest λ whose CV-error lies within one standard deviation of its minimum. The intention is to choose the simplest model possessing an accuracy that is comparable to the best model. Unfortunately, for the given approach, this procedure is unrewarding as it always leads to a degenerate model. This implies that the training-set performance of the chosen model is never significantly better than the performance of a degenerate model. From that finding the predictive power of the given system can be questioned. As pointed out before, this can hopefully be solved by adjusting prediction and calibration horizons.
Another workaround has been proposed by Roberts and Nowak (2014). Since they do not assume any time-series context, their approach relies on standard CV where folds are assigned randomly. They introduce a procedure called *percentile-LASSO*, which repeatedly performs CV to get a distribution of the optimal λ and takes a particular percentile from that distribution for LASSO estimation. Similar to the one-standard error rule, they state that the models chosen by CV tend to be too large and show that choosing values grater or equal to the 0.75-percentile can improve performance. They suggest to use the 0.95-percentile, which is supported by their simulations. According to Roberts and Nowak (2014), this approach can also be implemented together with the one-standard error rule.
Zhang and Yang (2015) investigate CV for model selection. They state that choosing k according to a general rule can be misleading and find that in highly instable cases with $p \gg N$, choosing $k \in \{5, 10\}$ can lead to a significantly larger CV-error than LOOCV. Overall, they state that LOOCV and repeated k-fold CV[15] with $k = 20$ or $k = 50$ perform best in this context. They conclude that since k-fold CV can be instable, repeated k-fold CV seems most promising for prediction.
The just described procedures by Roberts and Nowak (2014) and Zhang and Yang (2015) cannot easily be implemented in this paper since the blocked form of CV, which is used to reflect the time-series character of the data, does not involve randomness. As pointed out by Bergmeir and Benítez (2012), standard CV – although theoretically less adequate – also works well in time-series contexts such that one could cease from using the blocked form and instead implement standard CV. Then, one of the just proposed modifications could be implemented. This may improve stability and lead to better prediction performance at the cost of a theoretically less accurate estimation of λ^*. Nevertheless, if gains obtained from improved stability

[15] Repeated k-fold CV chooses the optimal λ such that it minimizes the CV-error over all repetitions.

prevail, this can lead to an estimate of λ^* which performs better in practice. Whether this holds obviously depends on the application at hand and cannot be answered in general. An appropriate simulation reflecting the particular application could help to give a recommendation.

Another starting point for modifications is the processing of news. As discussed in Section 6.3, a major drawback is the approach's inability to detect semantics. Even the incorporation of simple semantics could severely improve performance since the impact of words can depend heavily on their context. There exist some simple methods that can provide this improvement. In this paper, single words (*1-grams*) are taken as separate features. Correspondingly, possible alternatives are *2-grams* or *3-grams*, which take two or three subsequent words as single features such that features correspond to short expressions. The drawback of these alternatives is that they heavily increase the number of potential predictors as there exist much more three-word combinations than single words. In addition, a particular three-word combination is generally observed less often than a single word such that the majority of features will contain mostly zeros. Other techniques that can be used are *two-word combinations* and *noun phrases*. They are described in more detail, e.g., in Hagenau, Liebmann, and Neumann (2013). Also, one could consider the use of a dictionary to identify features or to capture news-sentiment. Especially the latter is interesting as it approaches the prediction problem from a different perspective.

Lastly, one could not only process the headlines. Instead, one could also use the news-body and group news according to their topics. This can be done by using a topic model such as *latent dirichlet allocation* (LDA), which was proposed by Blei, Ng, and Jordan (2003). Topic models estimate the probabilities of a particular text to belong to a number of prespecified topics[16]. The estimated frequency of news in each topic can then be used to predict some financial time series. In addition, some topics can be manually discarded by the researcher if they are considered to be irrelevant for forecasting. This is especially helpful if the database contains general news rather than news already focusing on financial topics. When dealing with general news, one could apply LDA two times. Once, with a small number of topics, to identify relevant news and a second time to identify different subjects within the group of relevant news. I find it promising to investigate whether this approach is better suited to explain stock market volatility.

Using a different feature representation can also be considered. A basic overview over possible feature representations is given in Nassirtoussi et al. (2014). In particular, using the multiplier of *term frequency* and *inverse document frequency* TF×IDF for measurement seems interesting as it takes into account that some words are more common than others.

To summarize, there exist a lot of starting points for potential improvements. Further research is needed to investigate whether some of the suggested modifications can im-

[16]By adjusting the total number of topics one can implicitly set the scope for each topic.

prove the system's prediction performance. It remains unclear whether LASSO-based methods are suited to be applied in this context.

8 Conclusion

In this paper I have applied various LASSO-based methods to forecast stock market volatility from news headlines. It is found that the system in its proposed form cannot achieve a forecasting performance which is better than chance. Nevertheless, it yields some insights about the topics that have recently been driving financial markets. The system comes with a whole lot of possibilities for potential improvements, which have been discussed extensively in Section 7. Due to its interdisciplinarity, the approach remains both interesting and challenging at the same time.
The increasing amount of data and computational powers provide great opportunities for future text-based analyses. Due to the growing digitalization, this development is bound to accelerate such that this area will likely become even more important. As pointed out by Varian (2014), collaborations between computer scientists and econometricians are promising in this context. Especially machine learning can provide helpful tools to researchers working in this area.
Although the approach proposed in this paper does not yield any remarkable prediction performance, I do not wish to discard it. Instead, further research is needed to determine whether the suggested modifications are able to improve performance. The approach is still considered to be promising and worth investing further effort.

Bibliography

Antweiler, Werner and Murray Z Frank (2004). "Is all that talk just noise? The information content of internet stock message boards". In: *The Journal of Finance* 59.3, pp. 1259–1294.

Arlot, Sylvain and Alain Celisse (2010). "Survey of cross-validation procedures for model selection". In: *Statistics Surveys* 4, pp. 40–79.

Bergmeir, Christoph and José M. Benítez (2012). "On the use of cross-validation for time series predictor evaluation". In: *Information Sciences* 191, pp. 192–213.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent dirichlet allocation". In: *The Journal of machine Learning research* 3, pp. 993–1022.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011). "Twitter mood predicts the stock market". In: *Journal of Computational Science* 2.1, pp. 1–8.

Bouchet-Valat, Milan (2014). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library*. R package version 0.5.1. http://CRAN.R-project.org/package=SnowballC.

Burman, Prabir, Edmond Chow, and Deborah Nolan (1994). "A Cross-Validatory Method for Dependent Data". In: *Biometrika* 81.2, pp. 351–358.

Butler, Matthew and Vlado Kešelj (2009). "Financial forecasting using character n-gram analysis and readability scores of annual reports". In: *Advances in artificial intelligence*. Springer, pp. 39–51.

Chatterjee, A and S. N. Lahiri (2013). "Rates of convergence of the Adaptive LASSO estimators to the Oracle distribution and higher order refinements by the bootstrap". In: *The Annals of Statistics* 41.3, pp. 1232–1259.

Das, Sanjiv R. and Mike Y. Chen (2007). "Yahoo! for Amazon: Sentiment extraction from small talk on the web". In: *Management Science* 53.9, pp. 1375–1388.

Fama, Eugene F. (1965). "Random Walks in Stock Market Prices". In: *Financial Analysts Journal* 21.5, pp. 55–59.

— (1970). "Efficient capital markets: A review of theory and empirical work". In: *The Journal of Finance* 25.2, pp. 383–417.

Feinerer, Ingo and Kurt Hornik (2015). *tm: Text Mining Package*. R package version 0.6-2. http://CRAN.R-project.org/package=tm.

Frank, Ildiko E. and Jerome H. Friedman (1993). "A Statistical View of Some Chemometrics Regression Tools". In: *Technometrics* 35.2, pp. 109–135.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1. http://www.jstatsoft.org/v33/i01/, pp. 1–22.

Groth, Sven S and Jan Muntermann (2011). "An intraday market risk management approach based on textual analysis". In: *Decision Support Systems* 50.4, pp. 680–691.

Hagenau, Michael, Michael Liebmann, and Dirk Neumann (2013). "Automated news reading: Stock price prediction based on financial news using context-capturing features". In: *Decision Support Systems* 55.3, pp. 685–697.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Texts in Statistics.

Huang, Chenn-Jung et al. (2010). "Realization of a news dissemination agent based on weighted association rules and text mining techniques". In: *Expert Systems with Applications* 37.9, pp. 6409–6413.

James, Gareth et al. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics.

Jin, Fang et al. (2013). "Forex-foreteller: Currency trend modeling using news articles". In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 1470–1473.

Knight, Keith and Wenjiang Fu (2000). "Asymptotics for Lasso-Type Estimators". In: *The Annals of Statistics* 28.5, pp. 1356–1378.

Kraemer, Nicole, Juliane Schaefer, and Anne-Laure Boulesteix (2009). "Regularized Estimation of Large-scale Gene Association Networks Using Graphical Gaussian Models". In: *BMC Bioinformatics* 10.1.

Meinshausen, Nicolai (2007). "Relaxed Lasso". In: *Computational Statistics and Data Analysis* 52.1, pp. 374–393.

Mittermayer, Marc-André (2004). "Forecasting intraday stock price trends with text mining techniques". In: *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. IEEE.

Nassirtoussi, Arman Khadjeh et al. (2014). "Text mining for market prediction: A systematic review". In: *Expert Systems with Applications* 41.16, pp. 7653–7670.

— (2015). "Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment". In: *Expert Systems with Applications* 42.1, pp. 306–324.

Nikfarjam, Azadeh, Ehsan Emadzadeh, and Saravanan Muthaiyah (2010). "Text mining approaches for stock market prediction". In: *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*. Vol. 4. IEEE, pp. 256–260.

Osborne, Michael R., Brett Presnell, and Berwin A. Turlach (2000). "On the lasso and its dual". In: *Journal of Computational and Graphical statistics* 9.2, pp. 319–337.

Peramunetilleke, Desh and Raymond K. Wong (2002). "Currency exchange rate forecasting from news headlines". In: *Australian Computer Science Communications* 24.2, pp. 131–139.

Porter, M.F (1980). "An Algorithm for Suffix Stripping". In: *Program* 14.3, pp. 130–137.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* http://www.R-project.org/. R Foundation for Statistical Computing. Vienna, Austria.

Racine, Jeff (2000). "Consistent cross-validatory model-selection for dependent data: HV-block cross-validation". In: *Journal of Econometrics* 99.1, pp. 39–61.

Riloff, Ellen and Janyce Wiebe (2003). "Learning extraction patterns for subjective expressions". In: *Proceedings of the 2003 conference on Empirical methods in natural language processing.* Association for Computational Linguistics, pp. 105–112.

Roberts, S. and G. Nowak (2014). "Stabilizing the lasso against cross-validation variability". In: *Computational Statistics & Data Analysis* 70, pp. 198–211.

Schumaker, Robert P et al. (2012). "Evaluating sentiment in financial news articles". In: *Decision Support Systems* 53.3, pp. 458–464.

Taşcı, Şerafettin and Tunga Güngör (2013). "Comparison of text feature selection policies and using an adaptive framework". In: *Expert Systems with Applications* 40.12, pp. 4871–4886.

Tetlock, Paul C. (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More than words: Quantifying language to measure firms' fundamentals". In: *The Journal of Finance* 63.3, pp. 1437–1467.

The New York Times (2015). *New York Times Article Search API v2.* http://developer.nytimes.com/docs/read/article_search_api_v2.

Tibshirani, Robert (1996). "Regression Shrinkage and Selection Via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1, pp. 267–288.

Trapletti, Adrian and Kurt Hornik (2015). *tseries: Time Series Analysis and Computational Finance.* R package version 0.10-34. http://CRAN.R-project.org/package=tseries.

Varian, Hal R. (2014). "Big Data: New Tricks for Econometrics". In: *The Journal of Economic Perspectives* 28.2, pp. 3–27.

Wiebe, Janyce and Ellen Riloff (2005). "Creating subjective and objective sentence classifiers from unannotated texts". In: *Computational Linguistics and Intelligent Text Processing.* Springer, pp. 486–497.

Wuthrich, B. et al. (1998). "Daily stock market forecast from textual web data". In: *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics.* Vol. 3. IEEE, pp. 2720–2725.

Zhang, Yongli and Yuhong Yang (2015). "Cross-validation for selecting a model selection procedure". In: *Journal of Econometrics* 187.1, pp. 95–112.

Zou, Hui (2006). "The Adaptive Lasso and Its Oracle Properties". In: *Journal of the American Statistical Association* 101.476, pp. 1418–1429.

Zou, Hui, Trevor Hastie, and Robert Tibshirani (2007). "On the "degrees of freedom" of the lasso". In: *The Annals of Statistics* 35.5, pp. 2173–2192.

Online Appendix to: LASSO-Based Forecasting of Financial Time Series on the Basis of News Headlines by Adrian Waltenrath

8.1 Related Literature

In this section I want to give a short overview on some literature related to this paper. More extensive reviews are provided by Nassirtoussi et al. (2014), Hagenau, Liebmann, and Neumann (2013) as well as Nikfarjam, Emadzadeh, and Muthaiyah (2010). As Nassirtoussi et al. (2014), I start with briefly addressing the Efficient Market Hypothesis (Fama 1965 and Fama 1970), which states that markets reflect all relevant information at all time. According to this hypothesis, market behavior is not predictable at all. In practice, however, the Efficient Market Hypothesis does not hold since market participants seldom possess knowledge about all available information. Also, new incoming information is not processed immediately by markets but over time. Therefore, prediction is theoretically possible during that process.

The most common approach to predict movements from textual data involves the usage of classifiers like *support vector machines* (SVM) or *naive Bayes* (c.f. Hastie, Tibshirani, and J. Friedman 2009) to assign news to categories according to their market impact. Such a procedure is applied e.g. by Mittermayer (2004) who uses three categories (*Buy*, *Sell* and *Neutral*) to predict single stock movements. Usually, data is divided into a training and a validation set, whereas the elements of the training set have to be assigned to the previously defined categories. Mittermayer (2004) does this according to the market performance within 15 minutes after the news release. The classifier – a SVM in case of Mittermayer (2004) – is then calibrated on the training set and predictions are done by using the classifier on the validation set to assign the news to the class they most likely belong to. This concept is followed by the majority of research related to this topic.

Another indirect, behavioral-economic argumentation states that news do not affect stock prices directly but influence the sentiment of investors, which in turn affects demand and therefore prices. Estimating the sentiment of textual data is often referred to as *sentiment analysis* or *opinion mining.* An example for this approach is the work of Schumaker et al. (2012) who use financial news to estimate market-sentiment and try to predict future movements from an estimated sentiment score. For computing the sentiment, they use a prespecified dictionary, which assigns scores for implied (positive or negative) sentiment to each word. Making use of such a dictionary is convenient as it allows to put different weights on different features[17]. As a drawback, one relies on a decent dictionary. Schumaker et al. (2012) and others use *OpinionFinder* (Riloff and Wiebe 2003 and Wiebe and Riloff 2005), a tool to evaluate the sentiment of whole sentences. Other examples for dictionaries that can be used for sentiment analysis are the Harvard IV-4 psychosocial dictionary, used e.g. by Tetlock, Saar-Tsechansky, and Macskassy (2008), or the Google-Profile of Mood States (GPOMS), used by Bollen, Mao, and Zeng (2011). Wuthrich et al. (1998), who were one of the first to use news articles to predict financial data, make use of a dictionary of word sequences, which was especially designed by an expert. As pointed out before, prediction is mostly performed by using some kind of classification. However, when sentiment is computed as an intermediate step, there have been some regression-based approaches (Tetlock, Saar-Tsechansky, and Macskassy 2008, Bollen, Mao, and Zeng 2011 and Jin et al. 2013).

Various news sources have been used by different authors. They range from print news (e.g. Tetlock 2007 and Tetlock, Saar-Tsechansky, and Macskassy 2008) over digital news (e.g. Schumaker et al. 2012, Wuthrich et al. 1998 and Nassirtoussi et al. 2015), ad-hoc announcements (e.g. Groth and Muntermann 2011) to internet stock messages boards (e.g. Antweiler and M. Z. Frank 2004 and Das and Chen 2007) and even cover social media platforms such as twitter (e.g. Bollen, Mao, and Zeng 2011).

Previous work also differs substantially in the prediction horizon, the predicted quantity as well as the definition of a feature. The considered prediction horizon ranges from several minutes (e.g. Mittermayer 2004) to one year (Butler and Kešelj 2009), while most of the research focuses on short term predictions of less than 24 hours. The predicted quantities are typically single stocks or stock indices. Some authors also focus on volatility or currency exchange rates.

In this work, a feature corresponds to a single word-stem. This can and has been generalized to expressions containing two or more words to capture some of the syntax. However, the technique of treating each word as a separate feature is applied by the vast majority of researchers and called *bag-of-words* or *1-grams.* Some alternatives, like *noun-phrases* or *n-grams*, are briefly discussed in Section 7. At this point, I again refer to the review by Nassirtoussi et al. (2014), who provide detailed tables about the characteristics of most of the references discussed here.

Authors typically evaluate their approach by assessing the proportion of correctly predicted directions. Success ranges between 50 and 70 percent, while everything

[17]Feature is the machine-learning equivalent to explanatory variable. The term feature is more general as it can apply to regressions as well as classifiers and other techniques.

above 55 is considered to be report-worthy (Nassirtoussi et al. 2014). Often, a trading strategy is derived, allowing to evaluate the approach by the hypothetical profit achieved by this strategy. In most cases and if there is no intermediate step, such as the estimation of sentiment, feature selection and therefore dimension reduction is done using the *term frequency* (TF). This means selecting features according to their occurrences, i.e. dropping all features which occur less than a certain threshold. When using a dictionary, selection is done implicitly by assigning zero scores to all words not contained in the dictionary.

TF can also be used as feature representation like it is done in this work. The term feature representation refers to the measurement of observations, i.e. the unit variables are represented in, which in this work is the number of occurrences per day. Another method is called *inverse document frequency* (IDF), which is defined as the logarithm of the total number of documents over the number of documents containing the term. This implies putting additional weight on rare expressions by assuming that they contain more essential information. Often, as a combination, their multiplier TF$\times$IDF is used. This measure increases in the number of times a feature appears in the document and is offset by the feature's overall frequency. It therefore takes into account that some words are more common than others (c.f. Taşcı and Güngör 2013 and Mittermayer 2004). A basic overview over possible representations is given, e.g., in Nassirtoussi et al. (2014).

In short, most of the research done on this topic tackles the problem from a different perspective than I do. To the best of my knowledge, LASSO-based procedures have not yet been applied in this context, although they are expected to provide reasonable performance when facing the problem of feature selection from a large number of possible predictors.

8.2 List of Stopwords

The following table shows the list of stopwords applied in my analysis. It is taken from the R-package *tm* (Feinerer and Hornik 2015):

i	me	my	myself	we	our	ours
ourselves	you	your	yours	yourself	yourselves	he
him	his	himself	she	her	hers	herself
it	its	itself	they	them	their	theirs
themselves	what	which	who	whom	this	that
these	those	am	is	are	was	were
be	been	being	have	has	had	having
do	does	did	doing	would	should	could
ought	i'm	you're	he's	she's	it's	we're
they're	i've	you've	we've	they've	i'd	you'd
he'd	she'd	we'd	they'd	i'll	you'll	he'll
she'll	we'll	they'll	isn't	aren't	wasn't	weren't
hasn't	haven't	hadn't	doesn't	don't	didn't	won't
wouldn't	shan't	shouldn't	can't	cannot	couldn't	mustn't
let's	that's	who's	what's	here's	there's	when's
where's	why's	how's	a	an	the	and
but	if	or	because	as	until	while
of	at	by	for	with	about	against
between	into	through	during	before	after	above
below	to	from	up	down	in	out
on	off	over	under	again	further	then
once	here	there	when	where	why	how
all	any	both	each	few	more	most
other	some	such	no	nor	not	only
own	same	so	than	too	very	

8.3 Stemming

5.5 and 5.6 show some examples of the outcome.

Table 5.5: Example of Stemming

orig. word	**stem**	**orig. word**	**stem**	**orig. word**	**stem**
walk		sensitiveness		controlling	
walks	walk	sensitivity	sensit	control	control
walked		sensitization		controller	

Table 5.6: Empirical Example of Headline Stemming

Timestamp (UTC)	Original Headline
2014-05-28T01:04:10Z	China Sacks Senior Energy Official Amid Corruption Crackdown: Xinhua
2015-01-26T18:54:53Z	China's Li Says to Create 10 Million Jobs in 2015: China Daily
2015-07-20T16:54:38Z	Rates Rise at Weekly US Treasury Auction

Date	Local Time	Stemmed Headline
2014-05-27	21:04:10	china, sack, senior, energi, offici, amid, corrupt, crackdown, xinhua
2015-01-26	13:54:53	china, li, sai, creat, million, job, china, daili
2015-07-20	12:54:38	rate, rise, weekli, u, treasuri, auction

8.4 Most Frequently Occurring Words and Empirical Quantiles

5.7 shows the ten most frequently occurring words, their number of occurrences as well as the corresponding unstemmed words. Note that these are not all possible words leading to a particular stem but the words empirically found in the news headlines over the whole horizon. These words are already converted to lower case. Also special characters like apostrophes are already dropped. In addition, 5.8 shows empirical quantiles of feature frequencies. As expected, the distribution is heavily left-skewed.

Table 5.7: Most Frequently Occurring Words

Stemmed Word	Occurrences	Corresponding Words
u	17198	us[18], u
sai	12935	says, say, saying
new	9083	new, news
kill	6766	kill, killed, killing, kills, killings
china	5847	china, chinas
polic	5351	police, policing, polices
bank	4617	banks, bank, banking, bankings, banked
man	4521	man, mans, manning, mannings
ukrain	4521	ukraine, ukraines
court	4421	court, courts, courting, courted
state	4305	state, states, stated, stately
obama	4285	obama, obamas
deal	4218	deal, deals, dealings, dealing
case	3699	case, cases, casings, casing
year	3636	year, years
russia	3540	russias, russia
charg	3533	charges, charge, charged, charging, charg
plan	3519	plans, planned, plan, planning
talk	3396	talks, talk, talking, talkative, talked
eu	3257	eu, eus

[18] Since special characters are already removed *us*, represents the personal pronoun *us* as well as the abbreviation for the United Stated *U.S.*

Table 5.8: Empirical Quantiles of Frequencies

50%	60%	70%	80%	90%	92.5%	95%	97.5%	100%
1	2	3	7	28	47	91	232	17198

8.5 LASSO Procedures

Standard LASSO

The LASSO (in its standard form hereafter abbreviated as std. LASSO) is a form of penalized regression and was originally proposed by Tibshirani (1996). It is described fairly well in James et al. (2013) and Hastie, Tibshirani, and J. Friedman (2009), whereas the latter is being more precise in handling the topic than the first one. LASSO is closely related to ridge regression as both procedures belong to the group of bridge estimators. In fact, bridge estimators were originally developed by I. E. Frank and J. H. Friedman (1993) as a generalization of ridge regression and generally apply a penalty of order $q > 0$. Ridge regression and LASSO are special cases and apply a L_2 and L_1–penalty, respectively. The L_1–penalty of the LASSO results in the fact that, in contrast to ridge regression, LASSO-coefficients are not only shrunken towards but exactly to zero[19], meaning that some sort of variable selection is done implicitly. Formally, the LASSO solves the following minimization problem:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 \right\} \quad \text{s.t.} \quad \sum_{j=1}^{p} |\beta_j|^q \leq t. \tag{1}$$

Or equivalently in Lagrangian form:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q \right\}, \tag{2}$$

with $q = 1$ to represent the L_1–penalty, $\boldsymbol{X} = [\boldsymbol{x_{\cdot 1}}, \dots, \boldsymbol{x_{\cdot p}}]$ being the predictor matrix with $\boldsymbol{x_{\cdot k}} = (x_{1k}, \dots, x_{nk})^T \ \forall\ k \in \{1, \dots, p\}$, $\boldsymbol{y} = (y_1, \dots, y_n)^T$ being the dependent variable, N being the number of observations and p being the number of independent variables, i.e. the number of potential predictors. The parameters t and λ are often referred to as tuning or shrinkage parameters since they control the amount of shrinkage and implicitly determine the number of variables included in the estimated model. Therefore, the selection of these parameters is crucial. I deal with this in detail in Section 4.2.

The given minimization problem as well as all subsequent ones is solved using the R-package *glmnet*[20] (J. Friedman, Hastie, and Tibshirani 2010), which possesses

[19] This can be visualized nicely by a graph showing the contours of the error and constraint functions for both procedures. Such a graph is presented in all three references mentioned above: Tibshirani (1996), James et al. (2013) and Hastie, Tibshirani, and J. Friedman (2009)

[20] Note that the *glmnet* package needs an input matrix with at least two columns (excluding the intercept). Therefore, in my analysis if a model of size two (the intercept and one predictor) is

a fast performing algorithm that uses cyclical coordinate descent to successively optimize the objective function over each parameter while keeping the others fixed[21]. One drawback of the standard LASSO is that not only the coefficients of predictors which are not included in the model are shrunken (to zero), but that the remaining nonzero coefficients are also biased towards zero. This leads to the fact that the estimated coefficients in general are not consistent (Hastie, Tibshirani, and J. Friedman 2009, p. 91). Since I focus on prediction performance rather than on obtaining the correct estimates, this is not a deal-breaker. Nevertheless, it is inconvenient. An obvious workaround, which is also suggested in Hastie, Tibshirani, and J. Friedman (2009, p. 91), is to perform LASSO to identify a subset of non-zero predictors and in a second step estimate OLS on that subset. I abbreviate this procedure as LASSO-OLS. As pointed out by Hastie, Tibshirani, and J. Friedman (2009, p. 91), this not feasible if the selected subset is large. Another common approach is the so-called *relaxed LASSO*, which aims to reduce bias and therefore mitigates the problem of inconsistency. This procedure is described next.

Relaxed LASSO

The relaxed LASSO (rel. LASSO) was proposed by Meinshausen (2007) and is a two-step-procedure. First, the standard LASSO is performed to determine a subset of non-zero predictors. Then, to estimate the model, LASSO is applied again to the subset of non-zero predictors identified during the first step. The idea is that the optimal amount of shrinkage applied in the second step should be smaller due to the smaller number of noise variables. Therefore, the estimated coefficients should suffer from less bias compared to the first step solution, which is equal to the standard LASSO. Sticking to the Lagrangian notation of equation (2), the problem solved by the relaxed LASSO estimator can be written in the following way:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} \left(x_{ij}\beta_j \cdot \mathbf{1}_{\{\beta_j \neq 0\}} \right) \right)^2 + \phi\lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{3}$$

with $\mathbf{1}_{\{\beta_j \neq 0\}}$ being an indicator-function which takes the value 1 if β_j is non-zero in the first stage estimation and 0 otherwise. The second stage LASSO-parameter ϕ controls the amount of shrinkage applied during the second step. It is defined on $(0, 1]$, while $\phi = 1$ corresponds to the first stage solution, i.e. the standard LASSO. For $\phi \to 0$ the coefficients are estimated as an unconstrained solution equal to

estimated, it is not possible to carry out estimation for the two-step procedures described below. In these situations, as a workaround, I drop the predictor and only estimate an intercept for all procedures. These situations are rare such that this practice – if at all – should have a minor impact on the results.

[21]The intuition is described, e.g., in Hastie, Tibshirani, and J. Friedman (2009, p. 92)

performing OLS on the subset of non-zero predictors (LASSO-OLS). As mentioned before, the relaxed LASSO procedure reduces bias by applying less (or at most equal) shrinkage than the standard LASSO, nevertheless (since $\phi > 0$) it yields inconsistent estimators. Another variation of the LASSO, which can deliver consistent estimates, is the *adaptive LASSO* outlined next.

Adaptive LASSO

A technique that implicitly performs variable selection like the standard LASSO and can deliver consistent estimates is the adaptive LASSO proposed by Zou (2006). To obtain consistency, it allows for different shrinkage factors by assigning individual weights w_j to the amount of shrinkage applied to each of the coefficients. It solves the following problem:

$$\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} |w_j \beta_j| \right\}. \tag{4}$$

Zou (2006) shows that adaptive LASSO estimators are consistent if $\boldsymbol{w} = \frac{1}{|\hat{\boldsymbol{\beta}}|^{\gamma}}$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$ is a root-n consistent estimator[22]. He suggests to use OLS estimates for $\hat{\boldsymbol{\beta}}$ and to determine $\gamma > 0$ together with λ by two-dimensional cross-validation[23]. In the present case, since $p > n$, it is not feasible to use OLS estimates as weights. This limitation for high-dimensional problems is also pointed out by Zou (2006). As a solution he suggests the use of ridge regression estimators.[24] Unfortunately, due to the huge amount of noise caused by the 47 023 predictors, ridge regression estimates can suffer from instability, which makes me cease from this idea.

Instead, two other variations are performed. For one, I use the standard LASSO estimates to plug in for $\hat{\boldsymbol{\beta}}$ (aLASSO-L). Predictors whose standard LASSO coefficients are equal to zero are excluded in the adaptive LASSO step. This implies the application of infinite shrinkage to those predictors. Using the standard LASSO estimates for $\hat{\boldsymbol{\beta}}$ is a common workaround in literature in case of $p > n$ (c.f. Chatterjee and Lahiri 2013 and Kraemer, Schaefer, and Boulesteix 2009).

As a second variation, I use the OLS coefficients, which have been estimated on the subset of non-zero predictors (LASSO-OLS). Again, predictors whose standard (first stage) LASSO coefficients are equal to zero are excluded. Both variations yield

[22] Note that root-n consistency of $\hat{\boldsymbol{\beta}}$ is not necessarily required since this condition can be weakened. See Zou (2006) for details.

[23] Cross-validation is discussed in detail in Section 4.2

[24] Note that this modification raises the need to estimate an additional tuning parameter λ^{ridge} to determine the best ridge regression fit.

consistent estimates[25] and, in a way, combine the advantages of relaxed and adaptive LASSO.

[25] OLS is root-n consistent. Root-n consistency of the LASSO estimates is shown by Knight and Fu (2000). The relevant Lemma is also stated in Zou (2006).

8.6 Blocked k-fold cross-validation

Due to the time-series character of the data, observations can be numbered consecutively from 1 to N. To create training and testing sets, the data is divided into k equally large, non-overlapping sets of subsequent observations (blocks).[26] Each block constitutes a fold and serves as testing set once, while the other sets are preserved as the corresponding training set.
To eliminate dependence between folds, h observations at the inner borders[27] of all k training sets are dropped. The advantage of using blocks rather than assigning the folds randomly is illustrated by the fact that this procedure minimizes the loss of data due to dropping h observations around the members of the testing set.[28]
Note that h should at least be set to one as a 24-hour-interval prior to market opening is used to predict daily returns. This setup implies that, in general, news released during the prior trading period can be used to forecast the next day's returns such that some information in the t-th row of $\boldsymbol{X}$ could have had influence on y_{t-1} and thus contradicts independence between folds for $h = 0$. In addition, as pointed out by Burman, Chow, and Nolan (1994), h should be set according to the order of autocorrelation of the data. They argue that, to ensure independence, h should be large and suggest $h = \frac{N}{6}$ as a rule of thumb. This results in removing one third of the data, which can be problematic, especially in small samples. Therefore, a trade-off between sample size and the degree of dependence between folds arises. Racine (2000) gives the same arguments as Burman, Chow, and Nolan (1994) and shows that even small values of h can significantly improve cross-validation performance.[29]
The 5.3, 5.4 and 5.5 show empirical autocorrelations for the VIX-returns, the S&P 500-returns, the ten most frequent features as well as ten randomly drawn features. Note that the randomly drawn features are the same as those that have been tested for stationarity. After observing these autocorrelations I set $h = 20$ for the entire analysis. I argue that this is more than adequate for the vast majority of predictors, while not being too costly in the sense of dropping too much information. It is acknowledged that this does not eliminate autocorrelation for some predictors since this would require h to be much greater. Nevertheless, in the given application it is necessary to find a compromise that serves the prediction performance.
The optimal tuning parameter λ (or the optimal vector (γ, λ) in case of two-dimensional cross-validation) is chosen from a prespecified grid[30] such that it mini-

[26] Note that in contrast to standard cross-validation there is no randomness involved. For certain modifications this can be a drawback as discussed later in Section 7.

[27] *Inner* means that only observations lying *between* two folds are dropped. The first and the last h observations are never dropped since there are no other folds to which dependence could occur.

[28] When using blocked cross-validation at most $2h$ observations are dropped. For standard cross-validation this can be up to $2\lceil \frac{N}{k} \rceil h$, which is much larger, as long as k is not close to N.

[29] Note that Burman, Chow, and Nolan (1994) introduce a correction term to take into account the loss of data. This correction is not applied by Racine (2000), nor is it in this paper.

[30] The grids λ and γ are chosen from, are described in Appendix 8.7.

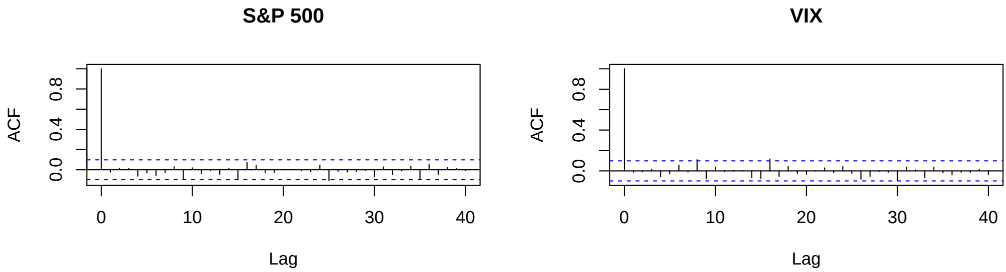


Figure 5.3: Empirical autocorrelations for the returns of S&P 500 and VIX from Jan 1, 2014 to Jul 31, 2015.

mizes the average cross-validation error, given a particular Loss-function L. To write this down mathematically, I introduce the following notation: Let K_1, K_2,...,K_k denote the sets of observations, which represent the folds. Let n_1, n_2,...,n_k be the corresponding numbers of observations in each of the folds. In addition, set $n_0 = 0$. Therefore, $\underline{n_j} = \sum_{i=0}^{j-1}(n_i)+1$ is the number of the first observation in K_j and, equivalently, $\overline{n_j} = \sum_{i=0}^{j} n_i$ corresponds to the number of the last observation in K_j. Also, let $\boldsymbol{x_{i\cdot}}$ denote the i-th row of the regressor matrix $\boldsymbol{X}$ and $X_{(\underline{n_j}:\overline{n_j})} = \{\boldsymbol{x_{i\cdot}} : \underline{n_j} \le i \le \overline{n_j}\}$ denote the set containing all observations from $\underline{n_j}$ to $\overline{n_j}$ such that $K_j = X_{(\underline{n_j}:\overline{n_j})}$. Correspondingly, let $K_j^c = X_{-(\underline{n_j}:\overline{n_j})} = \{\boldsymbol{x_{i\cdot}} : 1 \le i < \underline{n_j}\} \cup \{\boldsymbol{x_{i\cdot}} : \overline{n_j} < i \le N\}$ be the complement, i.e. all observations except those in fold j. Using this notation, the

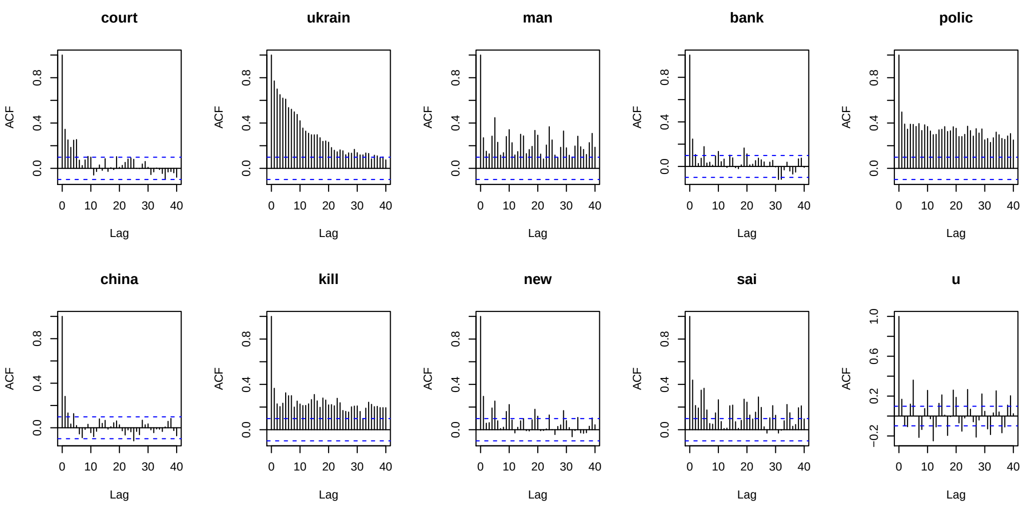


Figure 5.4: Empirical autocorrelations for the ten most frequent features from Jan 1, 2014 to Jul 31, 2015.

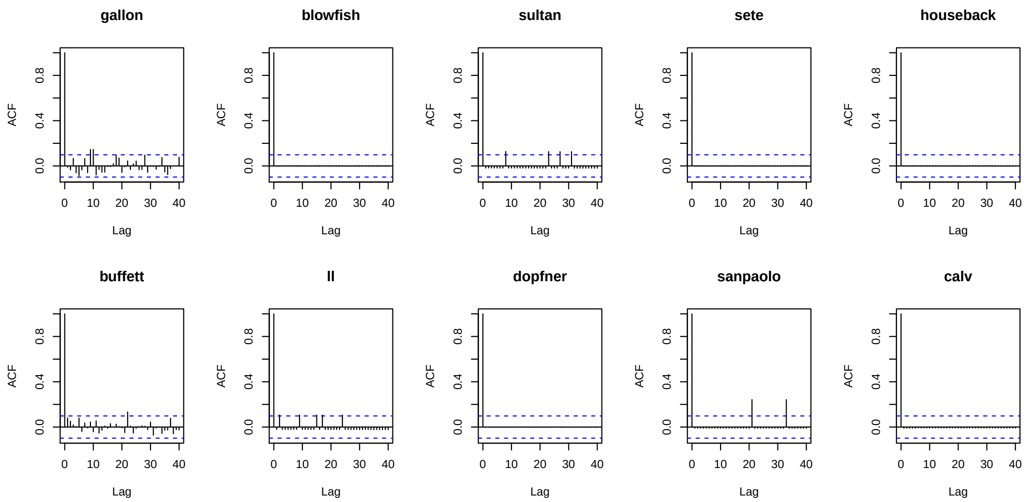


Figure 5.5: Empirical autocorrelations for ten randomly drawn features from Jan 1, 2014 to Jul 31, 2015.

implemented cross-validation procedure can be written down as follows:

$$\min_{\lambda} \frac{1}{N} \sum_{j=1}^{k} \sum_{i=\underline{n_j}}^{\overline{n_j}} L\big(y_i, \hat{f}^{\mathcal{X}_j}(\boldsymbol{x_{i\cdot}}, \lambda)\big), \tag{5}$$

with $\mathcal{X}_j = X_{-(\underline{n_j}-h:\overline{n_j}+h)}$ representing the set $\hat{f}$ is estimated on.[31] This can easily be extended to the two-dimensional case giving

$$\min_{\gamma} \min_{\lambda} \frac{1}{N} \sum_{j=1}^{k} \sum_{i=\underline{n_j}}^{\overline{n_j}} L\big(y_i, \hat{f}^{\mathcal{X}_j}(\boldsymbol{x_{i\cdot}}, \lambda, \gamma)\big). \tag{6}$$

Note that the described method is closely related to the h-block cross-validation procedure proposed by Burman, Chow, and Nolan (1994) and the hv-blocked cross-validation by Racine (2000). It can be seen as a k-fold-version of h-block cross-validation or an incomplete version of hv-blocked cross-validation.[32]

Error measure

As pointed out by Bergmeir and Benítez (2012) as well as Hastie, Tibshirani, and J. Friedman (2009, p. 219), different error measures can be used. The most common

[31] Being even more accurate it should be written as $\mathcal{X}_j = X_{-(\max(1,\underline{n_j}-h):min(\overline{n_j}+h),N)}$ since for the first and the last observation it does not make sense to subtract or add some $h > 0$.

[32] Both procedures, h-block cross-validation as well as hv-blocked cross-validation, are not suitable for my analysis as they are computationally too intense.

choices are the squared error loss, which gives the mean squared error (MSE), and the absolute error loss, yielding the mean absolute error (MAE):

$$L(Y, \hat{Y}) = \begin{cases} (Y - \hat{Y})^2 & \text{MSE} \\ |Y - \hat{Y}| & \text{MAE.} \end{cases} \tag{7}$$

The latter does not put additional weight on large deviations and is therefore less affected by outliers.

Standard Errors

As pointed out by Tibshirani (1996) as well as Knight and Fu (2000), it is non-trivial to compute standard errors for LASSO-type estimators. Tibshirani (1996) and Osborne, Presnell, and Turlach (2000) provide some approximations which are considered unsatisfactory by Knight and Fu (2000). Another suitable approach, which is also suggested by Tibshirani (1996) as well as Knight and Fu (2000), is to use the bootstrap to obtain valid standard errors. Since I focus on prediction, the computation of standard errors is not pursued here.

8.7 Grids for λ and γ

The Parameter λ

The Parameter λ is determined on a grid of length 100 which is chosen by the default option of the *glmnet*-package (J. Friedman, Hastie, and Tibshirani 2010) and depends on N, p as well as $\boldsymbol{w}$. For more details I refer to the corresponding help files in R or equivalently to the *glmnet*-reference manual, which is available under https://cran.r-project.org/web/packages/glmnet/glmnet.pdf.

The Parameter γ

The parameter γ is is determined on a grid from 0.025 to 10 with increasing distances between the single values. From 0.025 to 2 the next value increases by 0.025. From 2 to 4 the next value increases by 0.05. From 4 to 6 the next value increases by 0.25. From 6 to 10 the next value increases by 0.5. This results in the following grid of length 136:

$\gamma \in$ {0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.275, 0.3, 0.325, 0.35, 0.375, 0.4, 0.425, 0.45, 0.475, 0.5, 0.525, 0.55, 0.575, 0.6, 0.625, 0.65, 0.675, 0.7, 0.725, 0.75, 0.775, 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 1, 1.025, 1.05, 1.075, 1.1, 1.125, 1.15, 1.175, 1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.35, 1.375, 1.4, 1.425, 1.45, 1.475, 1.5, 1.525, 1.55, 1.575, 1.6, 1.625, 1.65, 1.675, 1.7, 1.725, 1.75, 1.775, 1.8, 1.825, 1.85, 1.875, 1.9, 1.925, 1.95, 1.975, 2, 2.05, 2.1, 2.15, 2.2, 2.25, 2.3, 2.35, 2.4, 2.45, 2.5, 2.55, 2.6, 2.65, 2.7, 2.75, 2.8, 2.85, 2.9, 2.95, 3, 3.05, 3.1, 3.15, 3.2, 3.25, 3.3, 3.35, 3.4, 3.45, 3.5, 3.55, 3.6, 3.65, 3.7, 3.75, 3.8, 3.85, 3.9, 3.95, 4, 4.25, 4.5, 4.75, 5, 5.25, 5.5, 5.75, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10}.

8.8 Simulation

To verify that the presented methods can indeed detect a small set of meaningful variables within a huge amount of noise, I conduct a simple simulation. As a first step, in order to create a problem similar to the one in my analysis, I generate time-series for 47 020 variables and 252 time points $t \in \{1, \ldots, 252\}$[33]. They are modeled as an autoregressive process of order one with a low coefficient:

$$x_t = 0.2x_{t-1} + \varepsilon_t \quad \forall\, t, \tag{8}$$

while ε_t is standard normally distributed: $\varepsilon_t \sim N(0,1)\ \forall\, t$. To add meaning to some variables I construct the dependent variable $\boldsymbol{y} = (y_1, \ldots, y_{252})^T$ from the first 20 of the just generated variables and add a great amount of noise $\varepsilon_t^* \sim N(0,10)\ \forall\, t$. With $\boldsymbol{x}^i$ denoting the i-th generated variable, $\boldsymbol{y}$ is obtained from

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*, \tag{9}$$

with $\boldsymbol{X} = [\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{20}]$ and $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_{252}^*)^T$. The variance of the error terms is chosen large to reflect a situation where $\boldsymbol{y}$ is affected by a lot of unmodeled factors. This situation is expected to be given in the application below. Coefficients are chosen as follows:

$$\begin{aligned} \boldsymbol{\beta} = (&-0.2, -0.4, \quad 0.6, -0.8, \quad 1.0, \quad 1.2, \quad 1.4, -1.6, \quad 1.8, -2.0, \\ &2.2, -2.4, \quad 2.6, -2.8, \quad 3.0, -3.2, \quad 3.4, -3.6, \quad 3.8, -4.0)^T. \end{aligned} \tag{10}$$

Note that the elements of $\boldsymbol{\beta}$ sum up to 0. 5.9 and 5.10 summarize the performance of all considered methods on the generated data. The results are presented for different k and computed under the mean absolute error with $h = 20$. Results for the squared error loss are presented in Appendix 8.9. The first value displayed in 5.9 corresponds to the number of non-zero coefficients among the first 20 variables. These variables, which were used to generate $\boldsymbol{y}$, are from now on referred to as the true predictors. The second value presented in 5.9 is the total number of non-zero coefficients. In addition, the true predictors that are not detected by the model, i.e. the variables whose coefficients are incorrectly estimated as zero, are presented. One can see that for given k the adaptive LASSO with OLS weights performs best in the sense of including the smallest number of noise variables. Thus, this method estimates the smallest models. As a consequence, the adaptive LASSO with OLS weights runs a higher risk of dropping one of the true predictors, as seen, e.g., for $k = 40$ and $k = 20$. An exception to this is the case of $k = 5$, where aLASSO-O estimates the smallest model while including 17 of the first 20 variables.[34]

[33] Since the majority of models is estimated on 252 observations.

[34] The statement of a higher probability of dropping additional true predictors is additionally supported by the results under the MSE presented in Appendix 8.9.

Table 5.9: Simulation Results MAE - Model Size

	5 Folds	10 Folds
std. LASSO	17/189 - V1, V2, V3	16/120 - V1, V2, V3, V4
LASSO-OLS	17/189 - V1, V2, V3	16/120 - V1, V2, V3, V4
rel. LASSO	16/180 - V1, V2, V3, V4	15/112 - V1, V2, V3, V4, V5
aLASSO-O	17/151 - V1, V2, V3	15/ 88 - V1, V2, V3, V4, V5
aLASSO-L	16/175 - V1, V2, V3, V4	16/110 - V1, V2, V3, V4

	20 Folds	40 Folds
std. LASSO	16/134 - V1, V2, V3, V4	16/115 - V1, V2, V3, V4
LASSO-OLS	16/134 - V1, V2, V3, V4	16/115 - V1, V2, V3, V4
rel. LASSO	16/130 - V1, V2, V3, V4	16/110 - V1, V2, V3, V4
aLASSO-O	15/111 - V1, V2, V3, V4, V5	15/ 89 - V1, V2, V3, V4, V5
aLASSO-L	16/128 - V1, V2, V3, V4	16/100 - V1, V2, V3, V4

The first values correspond to the number of non-zero coefficients for the first 20 variables. The second number corresponds to the total number of non-zero coefficients i.e. the size of the estimated model. The variables, whose coefficients are incorrectly estimated to be zero, are displayed after the minus sign.

5.10 summarizes the quality of the estimated coefficients. It shows the sum of absolute deviations of the estimated coefficients to the true coefficients, i.e. the coefficients used to generate the data. This sum is presented separately for the true predictors and the noise variables. Note that all estimated models fit an intercept, which is not included in the calculation of deviations.

When considering the coefficients for the first 20 variables, the adaptive LASSO procedures clearly perform best, while OLS weights seem to be superior to those obtained from the standard LASSO. The picture is different for the noise variables, since the standard LASSO performs much better than all two-step procedures. This originates from the greater amount of shrinkage applied to the coefficients and illustrates this exact property which was the reason for introducing the relaxed LASSO as well as the other two-step procedures in the first place. For the noise-variables, whose true coefficients are zero, the greater shrinkage leads to better estimates. In turn – again following the argumentation for introducing the relaxed LASSO – the standard LASSO performs badly for the first 20 variables, as these coefficients are also shrunken towards zero. Therefore, although this method provides the smallest total deviation, it might not be the best choice for forecasting. Considering only the two-step methods the adaptive LASSO procedures again yield the most promising results. In addition, using OLS weights seems to prevail slightly.

To summarize, these results show that the proposed methods can indeed detect

Table 5.10: Simulation Results MAE - Deviations of Coefficients

***k* = 5**	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.35	16.52	16.32	**15.82**	15.92
others	**18.45**	27.10	24.77	24.87	24.96
total	**34.80**	43.62	41.08	40.69	40.87

***k* = 10**	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.93	14.63	14.84	**12.95**	14.21
others	**10.47**	22.73	20.74	19.09	20.07
total	**27.40**	37.36	35.58	32.04	34.28

***k* = 20**	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.73	15.97	16.01	**15.23**	15.73
others	**12.09**	25.05	23.62	22.85	23.11
total	**28.82**	41.03	39.64	38.08	38.84

***k* = 40**	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	17.03	14.70	14.85	**13.39**	13.93
others	**9.96**	21.62	20.73	19.23	19.99
total	**26.99**	36.32	35.58	32.62	33.92

This table shows the sums of absolute deviations of the estimated coefficients to the true coefficients i.e. the coefficients used for generating the data. The first row displays the sum of deviations for the first 20 variables. The second row corresponds to the remaining 47000 variables. The third row shows the sum over all variables.

the majority of true predictors, while having difficulties detecting those with low coefficients. However, all selected models are too large as they also pick some of the noise variables. Still, as most coefficients estimated for these noise variables are small and alternate around zero, the estimated models are expected to have some predictive power.
It is acknowledged that the proposed methods are far from detecting the correct model. Nevertheless, reducing the predictors from 47 020 to less than 200 while retaining most of the true predictors can be regarded as (partial) success.

8.9 Simulation Results – MSE

The following tables show the simulation results computed under the squared error loss. The simulation was performed using the same data as for the tables presented in the text.

Table 5.11: Simulation Results MSE - Model Size

	5 Folds	10 Folds
std. LASSO	17/209 - V1, V2, V3	16/120 - V1, V2, V3, V4
LASSO-OLS	17/209 - V1, V2, V3	16/120 - V1, V2, V3, V4
rel. LASSO	16/198 - V1, V2, V3, V4	15/112 - V1, V2, V3, V4, V5
aLASSO-O	15/142 - V1, V2, V3, V4, V5	15/ 89 - V1, V2, V3, V4, V5
aLASSO-L	16/158 - V1, V2, V3, V4	15/111 - V1, V2, V3, V4, V5

	20 Folds	40 Folds
std. LASSO	16/129 - V1, V2, V3, V4	16/140 - V1, V2, V3, V4
LASSO-OLS	16/129 - V1, V2, V3, V4	16/140 - V1, V2, V3, V4
rel. LASSO	15/121 - V1, V2, V3, V4, V5	16/133 - V1, V2, V3, V4
aLASSO-O	15/ 92 - V1, V2, V3, V4, V5	15/115 - V1, V2, V3, V4, V5
aLASSO-L	15/118 - V1, V2, V3, V4, V5	16/124 - V1, V2, V3, V4

The first values correspond to the number of non-zero coefficients for the first 20 variables. The second number corresponds to the total number of non-zero coefficients i.e. the size of the estimated model. The variables, whose coefficients are incorrectly estimated to be zero, are displayed after the minus sign.

Table 5.12: Simulation Results MSE - Deviations of Coefficients

$\boldsymbol{k = 5}$	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.36	16.01	16.19	**14.94**	15.71
others	**20.25**	26.85	24.21	24.22	24.70
total	**36.61**	42.86	40.40	39.16	40.40
$\boldsymbol{k = 10}$	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.93	14.63	14.84	**13.11**	14.39
others	**10.47**	22.73	20.74	19.32	20.37
total	**27.40**	37.36	35.58	32.43	34.76
$\boldsymbol{k = 20}$	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.84	15.62	15.54	**13.58**	14.75
others	**11.00**	24.61	22.35	20.14	21.43
total	**27.84**	40.23	37.89	33.73	36.18
$\boldsymbol{k = 40}$	std. LASSO	LASSO-OLS	rel. LASSO	aLASSO-O	aLASSO-L
V1–V20	16.69	16.03	16.05	15.17	**15.14**
others	**12.62**	25.07	23.44	22.97	22.54
total	**29.32**	41.10	39.48	38.14	37.67

This table shows the sums of absolute deviations of the estimated coefficients to the true coefficients i.e. the coefficients used for generating the data. The first row displays the sum of deviations for the first 20 variables. The second row corresponds to the remaining 47000 variables. The third row shows the sum over all variables.

8.10 VIX – Cross-Validation Error Curves

The figure shows the cross-validation error curves using the VIX returns as dependent variable for nine randomly drawn dates.
Curves correspond to the (first stage) standard LASSO procedure. The optimal parameter chosen for the standard LASSO is most crucial as it affects all considered models. Note that curves for the other LASSO procedures are not easily comparable since the models estimated in the first stage differ in k such that estimation is carried out on different data. The *glmnet* package does rescale λ according to the number of variables. One could instead plot the cross-validation error against the degrees of freedom implied by each λ. This is not expected to yield further insights. Additionally note that, since not only the variables which are included in the final model were used for LASSO-estimation, the degrees of freedom are not adequately quantifying the complexity of the model. For this reason, the concept of effective degrees of freedom has been proposed by Zou, Hastie, and Tibshirani (2007).

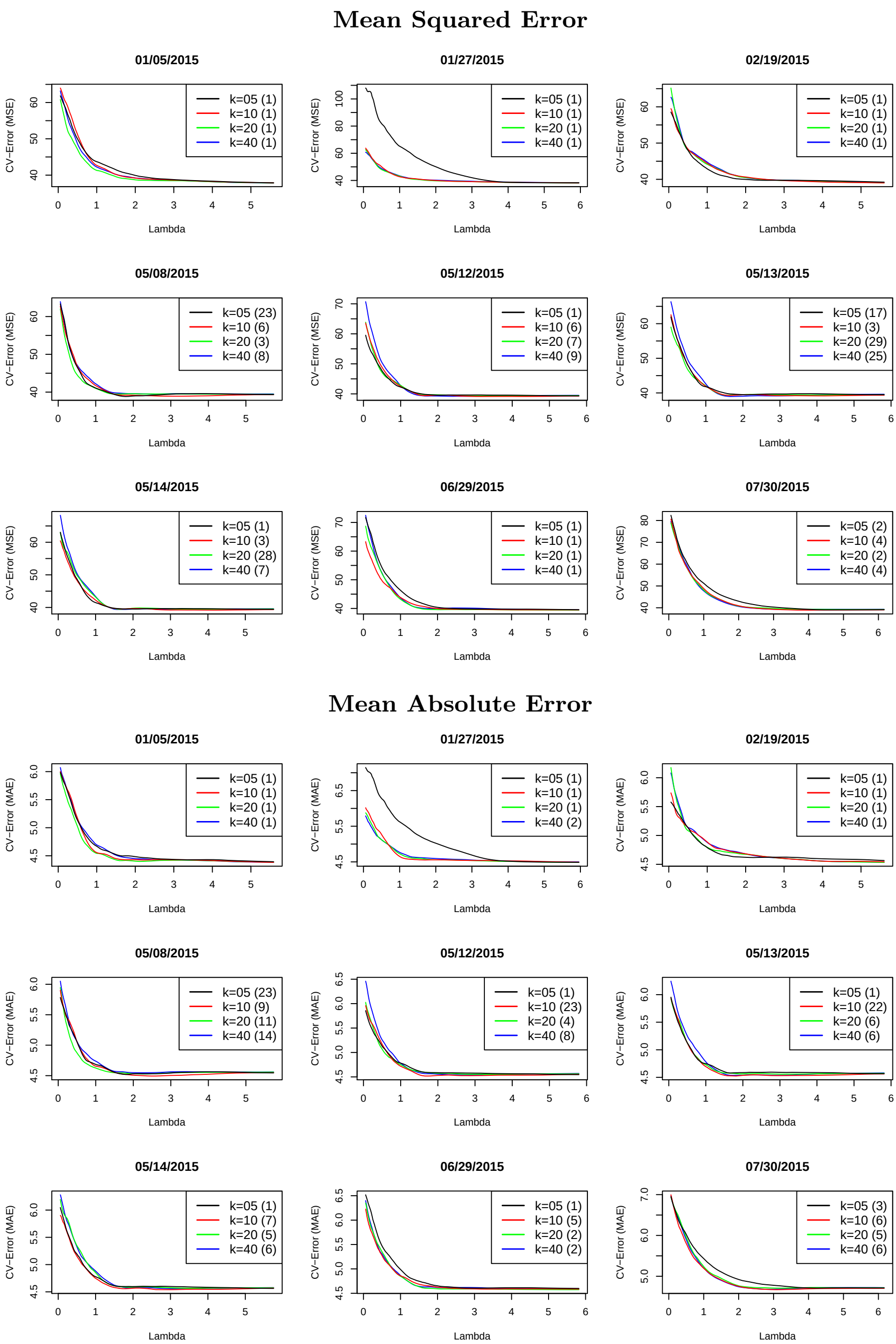


Figure 5.6: Cross-Validation Error Curves for the two Error measures and nine randomly drawn dates with VIX returns as dependent variable. For each k the number in brackets corresponds to the size of the estimated model implied by the optimal parameter. For (1) only an intercept is estimated.

8.11 Additional Performance Tables

Results under MSE

Table 5.13: Results under MSE

	Proportion of Correct Directions				Hypothetical Profit in %			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	63%	63%	62%	64%	129.7	121.8	115.4	126.7
LASSO-OLS	66%	61%	60%	59%	156.5	69.9	65.2	55.9
rel. LASSO	66%	61%	60%	59%	156.5	69.9	65.2	55.9
aLASSO-O	66%	61%	60%	59%	156.5	72.4	80.3	68.9
aLASSO-L	66%	60%	60%	58%	149.6	66.3	78.1	34.7

Table 5.14: Results under MSE for Non-Degenerate Models

	Proportion of Correct Directions				Hypothetical Profit in %			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	59%	63%	59%	65%	23.7	35.7	28.4	53.6
LASSO-OLS	71%	57%	53%	51%	50.5	−16.2	−21.9	−17.2
rel. LASSO	71%	57%	53%	51%	50.5	−16.2	−21.9	−17.2
aLASSO-O	71%	57%	51%	51%	50.5	−13.7	−6.8	−4.2
aLASSO-L	68%	54%	53%	47%	43.6	−19.8	−8.9	−38.4

Model Size (MSE)

Table 5.15: Model Size (MSE)

	5 Folds	10 Folds	20 Folds	40 Folds
std. LASSO	2.72 - 41/146	2.81 - 46/146	3.63 - 49/146	3.71 - 51/146
LASSO-OLS	2.72 - 41/146	2.81 - 46/146	3.63 - 49/146	3.71 - 51/146
rel. LASSO	2.70 - 41/146	2.81 - 46/146	3.62 - 49/146	3.70 - 51/146
aLASSO-O	2.52 - 41/146	2.60 - 46/146	3.18 - 49/146	3.34 - 51/146
aLASSO-L	2.51 - 41/146	2.66 - 46/146	3.32 - 49/146	3.49 - 51/146

This table summarizes the estimated model size for different k. The first value corresponds to the average number of non-zero coefficients (including the intercept). The value after the minus sign shows the number of times a non-trivial model (with at least one additional predictor) is estimated. 146 is the length of the prediction horizon.

Investing with Threshold

In the following I present tables for the proportion of correct directions and the hypothetical profit with an investing-threshold. As threshold 0.8% is chosen since the degenerate models always predict in $[-0.1, -0.8]$. Therefore it is only invested if a decrease of more than 0.8% or an increase of more than 0.8% is predicted by the system. The first table shows the number of investments taken. Note that the prediction horizon consists of 146 trading days. Introducing this threshold decreases the number of investments severely. Basically, it reduces the investments by all predictions of degenerate models plus those that are close to zero.

Table 5.16: Number of Investments with Threshold 0.8%

	MAE				MSE			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	17	19	25	21	12	17	16	18
LASSO-OLS	37	35	39	37	20	29	28	32
rel. LASSO	34	30	37	34	21	30	27	32
aLASSO-O	38	35	38	34	21	26	26	31
aLASSO-L	38	35	39	36	21	28	28	31

Table 5.17: Proportion of Correct Directions with Threshold 0.8%

	MAE				MSE			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	65%	53%	60%	48%	58%	47%	50%	50%
LASSO-OLS	57%	54%	56%	49%	70%	62%	57%	47%
rel. LASSO	56%	53%	57%	53%	67%	60%	56%	47%
aLASSO-O	61%	51%	50%	44%	67%	54%	54%	45%
aLASSO-L	61%	51%	56%	56%	67%	54%	54%	42%

Table 5.18: Hypothetical Profit in % with Threshold 0.8%

	MAE				MSE			
	$k = 5$	$k = 10$	$k = 20$	$k = 40$	$k = 5$	$k = 10$	$k = 20$	$k = 40$
std. LASSO	−8.2	−4.4	20.7	4.0	6.2	−18.1	−11.4	−5.0
LASSO-OLS	12.9	−0.8	30.9	37.6	26.7	4.3	8.2	−8.5
rel. LASSO	1.1	5.0	26.8	40.0	23.0	0.8	−0.5	−8.5
aLASSO-O	29.9	−10.0	18.3	16.4	21.2	−10.5	7.1	−9.5
aLASSO-L	17.7	−13.9	24.7	35.8	21.2	−13.7	1.4	−26.2

8.12 Performance over Whole Horizon

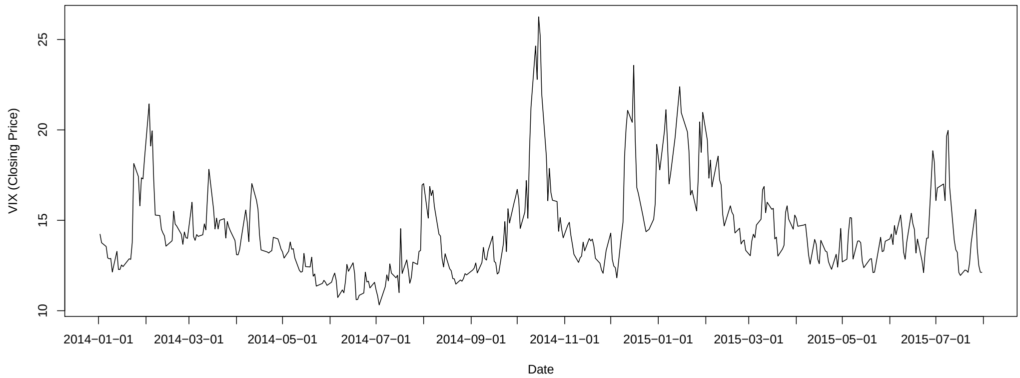


Figure 5.7: This figure shows the VIX in levels over the whole horizon.

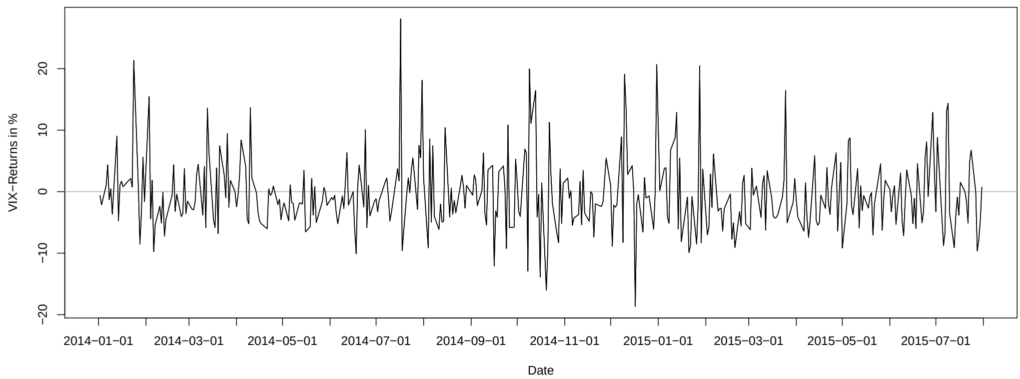


Figure 5.8: This figure shows the VIX-Returns in % over the whole horizon.