

Studies on Competition and Behavioral Economics

Dissertation

zur Erlangung des akademischen Grades

doctor rerum politicarum

(Doktor der Wirtschaftswissenschaft)

eingereicht an der

Wirtschaftswissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

von

Dipl.-Vw. Hans Christian Müller

geb. am 6. Juni 1984 in Detmold

Rektor der Heinrich-Heine-Universität Düsseldorf:

Prof. Dr. Dr. H. Michael Piper

Dekan der Wirtschaftswissenschaftlichen Fakultät:

Prof. Dr. Bernd Günter

Gutachter:

1. Prof. Dr. Justus Haucap

2. Prof. Dr. Hans-Theo Normann

Acknowledgements

I thank my supervisors, Justus Haucap and Hans-Theo Normann, furthermore my colleagues at the Duesseldorf Institute for Competition Economics (DICE), especially Ulrich Heimeshoff and Ralf Dewenter. Furthermore, my co-authors Stefan Haigner, Stefan Jenewein and Florian Wakolbinger; and Matthias Sutter, the supervisor of my diploma thesis.

Moreover, I am deeply grateful to Katharina Dröge and Micheline Prüter-Müller, who always supported me while I was working on this thesis.

In memory of Wolfgang Müller.

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Düsseldorf, im März 2013,

Hans Christian Müller

Contents

1	Introduction	1
2	The First Shall be Last: Serial Position Effects in the Case Contestants Evaluate Each Other	5
2.1	Introduction	6
2.2	The Contest	7
2.3	Data and Hypotheses	8
2.4	Results	9
2.5	Conclusion	12
2.6	References	13
3	The Effects of Gasoline Price Regulations: Experimental Evidence	14
3.1	Introduction	15
3.2	Related Literature	18
3.3	Experimental Setting	20
3.4	Results	27
3.4.1	Competition Dynamics	27
3.4.2	Consumer Welfare	30
3.5	Conclusion	35
3.6	References	37
3.7	Appendix 1: Figures	39
3.8	Appendix 2: Instructions	40
3.8.1	Phase 1 (<i>B-A Treatment</i>)	40
3.8.2	Phase 2 (<i>B-A Treatment</i>)	44

4	Forecast Errors in Undisclosed Management Sales Forecasts: The Disappearance of the Overoptimism Bias	45
4.1	Introduction	46
4.2	Conceptual Framework	50
4.3	Related Literature	51
4.4	Data	54
4.5	Hypotheses and Empirical Analysis	56
4.6	Conclusion	65
4.7	References	67
5	Counter- or Pro-Cyclical Public Spending? The Impact of Political Competition and Government Partisanship	70
5.1	Introduction	71
5.2	Related Literature	75
5.2.1	Opportunism School	75
5.2.2	Partisan School	77
5.2.3	Activist Fiscal Policy	78
5.3	Data	80
5.4	Estimations	84
5.4.1	Cyclicalities	84
5.4.2	Estimation Strategy	88
5.4.3	Partisan Hypothesis	90
5.4.4	Opportunism Hypothesis	91
5.4.5	Partisan versus Opportunism Hypothesis	92
5.5	Conclusion	93
5.6	References	96
6	Final Remarks	100

List of Figures

2.1	Average Number of Assigned and Received Points on Each Day of the Week	9
2.2	Average Rank per Day	10
2.3	Fraction of Victories on Each Day of the Week	10
3.1	10x10-Hotelling-City with Four Shops (Gas Stations)	21
3.2	Representative Group in Treatment B-B	28
3.3	Mark-ups and Wholesale Prices	28
3.4	Average Mark-ups Over the Day	29
3.5	Weighted Average Profits in Phase 1	31
3.6	Weighted Average Mark-ups in Phase 1	31
3.7	Mean Daily Profits and Mark-ups in Phase 1	32
3.8	Weighted Average Profits in Phase 2	33
3.9	Weighted Average Mark-ups in Phase 2	33
3.10	Mean Daily Profits	35
3.11	Average Profits and Mark-ups	39
5.1	Summary Statistics: Coalitions	84

List of Tables

2.1	Regression Results	11
3.1	Treatments	25
3.2	Random Effects Panel Regressions for Phase 1	32
3.3	Random Effects Panel Regressions for Phase 2	36
4.1	Summary Statistics I	54
4.2	Summary Statistics II	55
4.3	Forecast Errors	57
4.4	Regression Results	63
5.1	Summary Statistics: Macroeconomic Values (Means)	82
5.2	Summary Statistics: Election Dates	82
5.3	Correlations	85
5.4	Regression Results I	89
5.5	Regression Results II	94

Chapter 1

Introduction

This thesis covers several aspects of competition economics which are analyzed empirically or experimentally. It can be divided into two parts. Chapters 2 and 3 study how competition outcomes (in contests, respectively consumer markets) are affected by the design of the competitive environment. Chapters 4 and 5 reverse the question and ask how the existence of a competitive environment changes the behavior (of firms, respectively governments). Finally, chapter 6 provides a conclusion.

In chapter 2 (“The First Shall be Last: Serial Position Effects in the Case Contestants Evaluate Each Other”)¹, we analyze competitions where the participants are contestants and jurors at the same time. Searching for possible order effects, we use data of the German TV-cooking contest “Das Perfekte Dinner”, in which five amateur-cooks in turn prepare a dinner for the other four contestants and are afterwards evaluated by them. We find that the starting order indeed has an influence on competition outcomes, namely that the first contestant is significantly disadvantaged. We suspect that this is due to, first, information diffusion, second, Bayesian belief updating taking place over the course of the contest and, third, initial uncertainty about a contestant’s relative quality.

¹This study is based on work with Stefan Haigner, Stefan Jenewein and Florian Wakolbinger (Gesellschaft für Angewandte Wirtschaftsforschung Innsbruck and University of Innsbruck).

Regarding these results, we recommend that organizers of contests should at least try to determine the starting order as randomly as possible and hence avoid using alphabetical orders. Although this would not rule out biases caused by order effects, it would ensure that the contestants' chances are equal at least prior to the determination of the starting order.

Chapter 3 ("The Effects of Gasoline Price Regulations: Experimental Evidence")² seizes the recent debate among German politicians whether price regulations could strengthen competition in the gasoline market. This debate arose after a sector inquiry in Germany had backed suspicions of tacit collusion and had suggested to adopt regulatory pricing rules for gas stations similar to those implemented in Austria, parts of Australia, Luxembourg or parts of Canada. In order to increase consumer welfare, these rules either restrict the number of price changes per day or cap the mark-up for gasoline retail prices.

While economic theory clearly suggests that gasoline retail markets are relatively prone to collusive behavior (since oligopolistic market structures tend to prevail, market interactions occur frequently, rivals' prices are highly transparent and demand is rather inelastic), theoretical predictions about the impact of the proposed measures are less clear. To fill the research gap, a controlled laboratory experiment is conducted in order to evaluate whether the proposals achieve the intended goals.

Our results reveal that two of the suggested rules rather decrease consumer welfare: The Austrian rule which only allows one price increase per day (while price cuts are always possible) and the Luxembourg rule which introduces a maximum mark-up for retailers. While no rule tends to induce lower retail prices, the Western Australian regime which allows at most one daily price change (no matter whether up or down) does at least not harm consumers. Taken together, our results do not answer the hopes which were set in the proposed regulatory changes.

²This study is joint work with Justus Haucap, University of Duesseldorf.

In chapter 4 (“Forecast Errors in Undisclosed Management Sales Forecasts: The Disappearance of the Overoptimism Bias”), I turn to a hitherto unsolved question in behavioral economics, namely, why companies’ forecasts of their own future revenues or profits are (on average) too optimistic, which has been a consistent finding in the literature. Two competing hypotheses have evolved: On the one hand, that a cognitive bias (the “planning fallacy”) is to blame for the distortions (Kahneman and Tversky, 2003), and, on the other hand, that “strategic deception” is main the reason (Flyvbjerg et al., 2002). The latter assumes that forecasts are a strategic tool in highly competitive market environments and that firms hence have incentives to deliberately state too optimistic values. Previous empirical studies were unable to contribute to the solution of this debate.

Unlike all previous works, which evaluated the accuracy of forecasts from public disclosures, this chapter analyzes 6,234 undisclosed, company-internal sales forecasts, which German firms provided anonymously to the IAB Establishment Panel. Since internal projections cannot serve as a tool to mislead the market, the hypothesis that cognitive biases are to blame would be supported, if these forecasts were found to be overoptimistic on average, too.

However, the results reveal the average internal sales forecast to be significantly *overpessimistic*, which, if anything, supports the “strategic deception” hypothesis. Hence, I propose that the non-existence of a general bias towards overoptimism is due to the lack of incentives to consciously overgloss future prospects in internal forecasts and, furthermore, that *overpessimism* may be a consequence of loss aversion.

Regression results further show that joint-stock companies are more overoptimistic, while a greater share of women among the workforce tends to reduce overoptimism. The latter is in line with previous evidence in behavioral economics which revealed womens’ self-assessment of their own capabilities to be less upward distorted.

Chapter 5 (“Counter- or Pro-Cyclical Public Spending? The Impact of Political Competition and Government Partisanship”) covers a well-known debate in political economy, albeit from a different perspective. For several decades now, researchers have been trying to establish the determinants which make governments choose certain types of economic policies. While the “partisan approach” (Hibbs, 1977) assumes economic policy to be mainly driven by the basic convictions of the ruling parties or coalitions (predicting that left-wing governments prefer more “big government” than right-wing coalitions), the “opportunism school” (Nordhaus, 1975) states that the desire to be reelected is the crucial parameter. Hence, it is predicted that governments choose popular policies when facing political competition in election years and unpopular ones afterwards. Although some evidence was found in favor of both hypotheses with regard to monetary and fiscal policy, the overall impression is rather mixed.

While almost all previous works in the recent three decades focused on the influence of opportunism or partisanship on changes in spending or debt *levels*, Seitz (2000) proposed to analyze the effects on the *cyclicality* of fiscal policy (thus, the governments’ propensity to counterbalance economic fluctuations by adversely adjusting public spending). Using data from the German federal states, he tested the partisan hypothesis (that left-wing governments conduct more “Keynesian-style” fiscal policies), but had to reject it.

With more recent data of the German states, I repeat Seitz’ analysis and also test the opportunism hypothesis (that governments conduct more counter-cyclical spending policies in election years in order to signal that they care for those suffering from economic hardships). While the partisan hypothesis cannot be supported again, some evidence can be provided that state governments indeed conduct a significantly more counter-cyclical spending policy when campaigning for reelection.

Chapter 2

The First Shall be Last: Serial Position Effects in the Case Contestants Evaluate Each Other

Abstract

We analyze competitions where the contestants evaluate each other and find the first contestant to be disadvantaged. We suspect that this is due to information diffusion, Bayesian belief updating taking place in course of the contest and initial uncertainty about a contestant's relative quality.

Key Words: *Serial Position Effects, Ordering Effects*

JEL-Classification: *C12, D81*

This study is joint work with Stefan Haigner, Stefan Jenewein and Florian Wakolbinger (Gesellschaft für Angewandte Wirtschaftsforschung Innsbruck and University of Innsbruck).

2.1 Introduction

Many economic decisions involve the evaluation of alternatives which are presented in sequence. For example, job applicants are interviewed one after the other, contributions to architecture competitions are presented in a row, or decisions on the upcoming host of Olympic Games are made this way.

While from an efficiency point of view, best alternatives should be ranked first, the second-best ranked second, and so on, such evaluations have been found to be subject to a serial position effect. For instance, data on synchronized swimming (Wilson, 1977), the Eurovision Song Contest (Haan et al., 2005; Bruine de Bruin, 2005), the Queen Elisabeth piano competition (Ginsburgh and van Ours, 2003; Glejser and Heyndels, 2001) or figure skating (Bruine de Bruin, 2006) have been employed to show that contestants who perform later are more likely to win, i.e. the last shall be first, although random assignment to the starting position in the series assured that the expected quality of all contestants is equal.

Beyond that, there is mixed evidence about the question whether this advantage rises linearly with the starting number (Bruine de Bruin, 2005) or whether there is a J-shaped relationship with the very first starter having a slight advantage compared to those with mid-level numbers, and greater advantages for later players (Page and Page, 2010; Haan et al., 2005). Contradicting both results, we find in analyzing the TV cooking contest “The Perfect Dinner” that the first, and only the first, shall be last.

This cooking contest notably differs from other competitions in that competitors are both performers as well as jury members. Thus, each performance is judged by the rivals. Such scenarios might frequently be encountered if groups have to choose a leader from their midst, for example if party executives meet to find a new chairman after a sudden resignation of the predecessor.

We find the contestant who performs first to be disadvantaged. She has a lower probability of winning, although assignment to the starting position in the series is random. Thus, contestants should beware of performing first, while

importantly, it makes no difference whether the contestant is second, third, fourth and so on in the row.

The chapter is organized as follows. Section 2.2 introduces the contest and Section 2.3 describes the data and formulates some hypotheses. Section 2.4 presents the results and Section 2.5 summarizes.

2.2 The Contest

The TV show “Das Perfekte Dinner” (The Perfect Dinner)¹ features five contestants who compete for the “best dinner of the week” award. From Monday to Friday, one of the contestants cooks a three-course-meal which is evaluated by the four other contestants. It is possible to assign up to 10 points for a meal, such that the maximum total number of points to receive is 40.² Consumption and evaluation takes place immediately after cooking. However, the performing contestants are not informed about the number of points the others assigned. Importantly, the contestants’ starting order is randomly assigned.

On Fridays, after each of the five meals has been evaluated, the winner of the week, who receives a price of 1,500 Euros, is announced. In addition, everyone receives a lump-sum allowance of 600 Euros to cover the expenses. In case of a tie, the prize is shared equally among the winning contestants.

Within this setting, each contestant is both a performer (on one day) as well as an evaluator (on four days). From the point of view of contestant i , clearly the probability of winning decreases in the number of points she assigns to her competitors $-i$. For an egoistic and rational contestant it is thus a dominant strategy to assign zero points to each rival, and in the unique Nash-equilibrium each contestant then assigns and receives zero points.

¹The show is broadcasted by VOX on evenings from Monday to Friday. Channel 4 (UK) previously broadcasted it under the name “Come Dine with Me”, TLC (USA) called it “Dinner Takes All”.

²For 22 weeks in 2009, the maximum total number of points is 80. We have normalized the results of these weeks to be consistent with the remaining data.

2.3 Data and Hypotheses

We use data from 186 competitions which took place between March 2006 and February 2010. For each contestant, we observe the day on which she has performed and how many points she assigned to, and received from, each of her competitors. This allows calculating her rank. Moreover, we observe the contestants' gender and age, but unfortunately not their tastes (e.g. vegetarianism). The data are available online at the webpage of the producer (www.xov.de).

Although it might be optimal for every contestant to assign zero points to each rival, there seem to be moral constraints or social norms which prevent them from doing so. Moreover, although those performing sooner and later do on average not differ in quality due to random assignment, it might be better not to perform first for three reasons. First, later contestants have more information on their opponents since there is some informal chatting about the meal provided by the previous contestants. Chatting might transmit information on the opponents' preferences and thus makes later contestants able to respond to those preferences, i.e. by providing a special vegetarian meal for vegetarian opponents.

Second, assigning at most a medium number of points to the very first meal might be a sign of uncertainty about its relative quality compared to the following (Bruine de Bruin, 2005). Such a rating might serve as a benchmark for the later starters which must leave space for possibly better cooks. Hence, the very first starters would face the disadvantage that the jurors could not estimate the overall quality at the beginning.

Third, some Bayesian belief updating might take place. While the first evaluation mirrors the consistency of the jury's prior beliefs with the performance, later evaluations might be based on updated beliefs. If, say, contestants first expect an extraordinarily delicious meal, they will rate a mediocre meal worse than if they expect a mediocre meal.

2.4 Results

Points

Figure 2.1: Average Number of Assigned and Received Points on Each Day of the Week

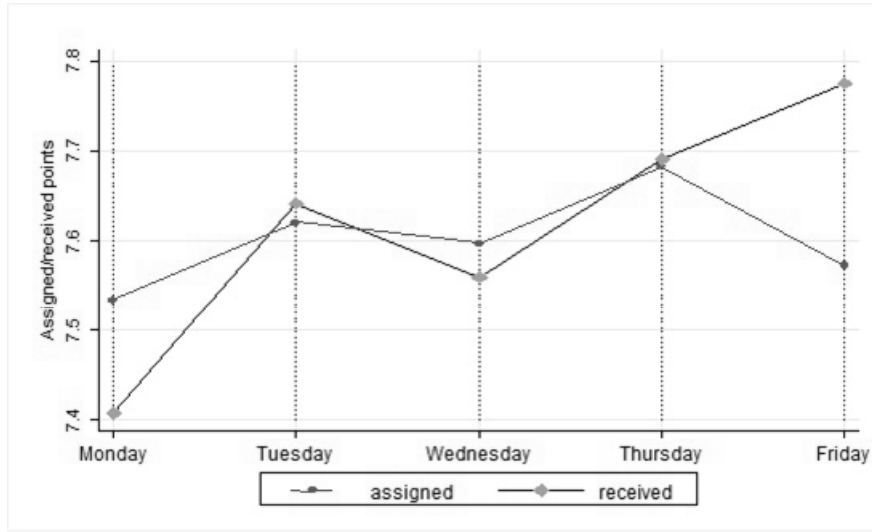


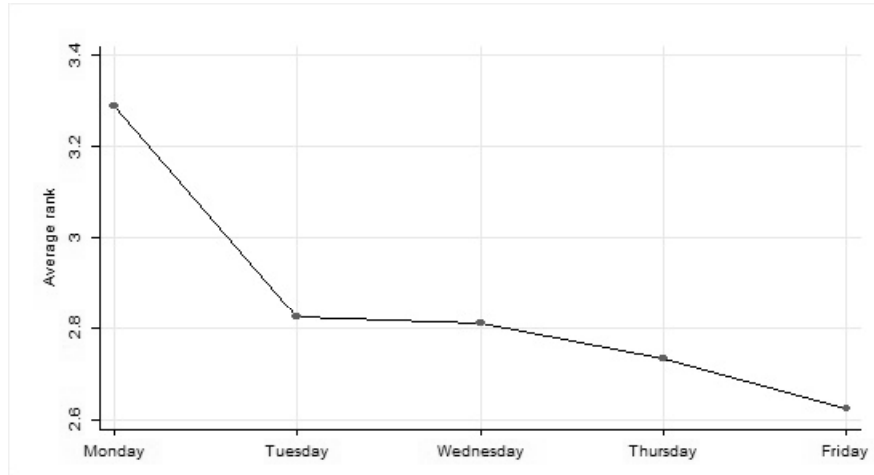
Figure 2.1 shows how many points (out of ten) contestants cooking on a specific day of the week on average received from and assigned to their competitors. Unlike standard economic theory predicts, average received and assigned points are far away from the lower bound of zero. Moreover, Monday-performances are on average rated significantly worse than performances on later days of the week.³

Ranks

Figure 2.2 shows the average rank which the contestants of each of the five days of the week achieved. The average rank of Monday-contestants is higher by more than 0.4 ranks than that of Tuesday to Friday-contestants, which differ in average ranks by less than 0.2 at most. Thus, we find a serial position effect where only the first performer seems to be disadvantaged.

³Mann-Whitney U-test p-values for a comparison of Monday's points with the points of other days of the week are <0.01 (Tuesday, Thursday and Friday) and 0.057 (Wednesday).

Figure 2.2: Average Rank per Day



Winning

Figure 2.3: Fraction of Victories on Each Day of the Week

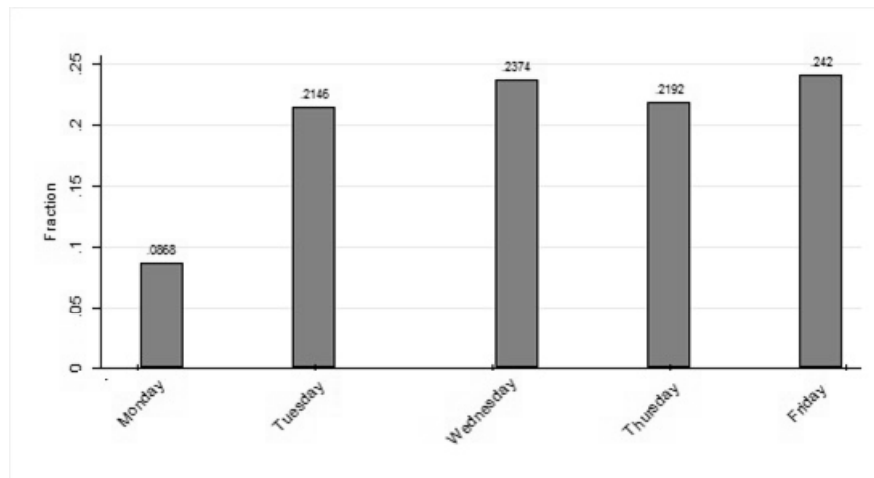


Figure 2.3 shows that contestants cooking on Mondays have won in only 8.6 percent of the series, while contestants performing on Tuesdays to Fridays have won in more than 20 percent of competitions. Moreover, there is surprisingly little variation in the probability of winning among Tuesdays' to Fridays' contestants. Figure 2.3 presents further evidence for the serial position effect with only the first contestant to be disadvantaged.

Determination of Ranks and Probabilities of Winning

Table 2.1 presents the results from ordered probit as well as probit regressions explaining the rank and probability of winning.⁴ For the regressions given in columns (1) and (2), we use dummies for the day of the week (Monday is the basis) as well as indicators for gender and age as explanatory variables. The coefficient estimates show that the probability of winning when performing on Tuesdays to Fridays is by 21 to 26 percentage points (and significantly) higher than when performing on Mondays. However, the coefficients for the day-dummies do not significantly differ from one another.

Table 2.1: Regression Results

Method Variable	Ordered Probit Rank	Probit Winner	Probit Winner
Monday			-0.16*** (0.039)
Tuesday	-0.327*** (0.115)	0.217*** (0.062)	
Wednesday	-0.347*** (0.118)	0.248*** (0.063)	
Thursday	-0.398*** (0.111)	0.226*** (0.061)	
Friday	-0.495*** (0.118)	0.261*** (0.064)	
Time			0.009 (0.015)
Female	0.144* (0.077)	-0.049 (0.030)	-0.049 (0.030)
Age	0.000 (0.004)	0.000 (0.001)	0.000 (0.001)
Observations	902	902	902
Clusters	181	181	181

robust standard errors in brackets;
*: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$

Moreover, we test for the influence of a linear time trend and the first-performer disadvantage in an additional regression given in column (3). The

⁴We allow for correlation among those performers who compete against each other by clustering the observations by weeks.

latter disadvantage persists, while the time trend is insignificant, which indicates that apparently all relevant information diffuses at discussion of the first performance and that the competitors update their beliefs after the first performance while later on there is no essential updating anymore.

2.5 Conclusion

Previous research has found that in contests with independent juries, competitors performing later have advantages over those performing sooner, i.e. the last shall be first. This holds for both end-of-sequence (all contestants are evaluated at the end of the competition) as well as step-by-step (each contestant is evaluated right after her performance) judgements (Bruine de Bruin, 2006).

For the TV-cooking contest “The Perfect Dinner” in which the competitors evaluate each other we find, however, that only the very first shall be last. The first contestant has a significantly higher rank and a significantly lower probability of winning than her competitors performing afterwards. However, the following four cooks do not have any disadvantages relative to each other.

Why is the first performer, and only the first, last when competitors evaluate each other? We suspect that this is due to the following three reasons. First, the first performer might set a reference point for further performers, and beliefs on the performances to be expected might be updated after the first performance. Consistency (leading to higher evaluations) of updated beliefs and performance might be easier to achieve than consistency of prior beliefs and performance. Second, jurors might assign only a medium number of points to the first performer out of uncertainty about the relative quality of her meal compared to the upcoming ones. As such, the first rating might be a benchmark which must leave space for possibly better meals. Third, in course of the contest, information on the contestant’s preferences could be revealed, which disadvantages earlier performers, in particular the first contestant.

2.6 References

- Bruine de Bruin, W. (2005) Save the Last Dance for me: Unwanted Serial Position Effects in Jury Evaluations, *Acta Psychologica* 118, 245–260.
- Bruine de Bruin, W. (2006), Save the Last Dance II: Unwanted Serial Position Effects in Figure Skating Judgments, *Acta Psychologica* 123, 299–311.
- Ginsburgh, V. and van Ours, J. (2003), Expert Opinion and Compensation: Evidence from a Musical Competition, *American Economic Review* 93, 289–296.
- Glejser, H. and Heyndels, B. (2001), Efficiency and Inefficiency in the Ranking in Competitions: The Case of the Queen Elisabeth Music Contest, *Journal of Cultural Economics* 25, 109–129.
- Haan, M., Dijkstra, S. and Dijkstra, P. (2005), Expert Judgment Versus Public Opinion Evidence from the Eurovision Song Contest, *Journal of Cultural Economics*, 29, 59–78.
- Page, L. and Page, K. (2010), Last Shall be First: A Field Study of Biases in Sequential Performance Evaluation on the Idol Series, *Journal of Economic Behavior and Organization* 73, 186–198.
- Wilson, V. (1977), Objectivity and Effect of Order of Appearance in Judging of Synchronized Swimming Meets, *Perceptual and Motor Skills* 44, 295–298.

Chapter 3

The Effects of Gasoline Price Regulations: Experimental Evidence

Abstract

Economic theory suggests that gasoline retail markets are prone to collusive behavior. Oligopolistic market structures prevail, market interactions occur frequently, prices are highly transparent, and demand is rather inelastic. A recent sector inquiry in Germany backed suspicions of tacit collusion and suggested to adopt regulatory pricing rules for gas stations similar to those implemented in Austria, parts of Australia, Luxembourg or parts of Canada. In order to increase consumer welfare these rules either restrict the number of price changes per day or they limit the mark-up for gasoline retail prices. As theoretical predictions about the impact of these measures are mixed and empirical studies rare, we analyze the effects, using an experimental gasoline market in the lab. Our results reveal that two of the suggested rules rather decrease consumer welfare: The Austrian rule which only allows one price increase per day (while price cuts are always possible) and the Luxembourg rule which introduces a maximum markup for retailers. While no rule tends to induce lower retail prices, the Western-Australian rule which allows at most one daily price change (no matter whether up or down) does at least not harm consumers.

Key Words: *Gasoline Prices, Fuel Prices, Experimental Gasoline Market, Fuel Price Regulation, Retail Price Regulation, Gas Stations*

JEL-Classification: *L13, L71, L81, L88, K23, C90*

This study is joined work with Justus Haucap, University of Duesseldorf.

3.1 Introduction

Few prices in the Western world are regularly fueling so much interest and debate as gasoline prices. In many jurisdictions around the world the up and down of fuel prices is a subject of regular discussion in the press and among politicians. In many countries, public uproars have put pressure on governments to limit gas stations' power to exploit their perceived market power. To name but a few, price regulations have recently been discussed in Canada, Germany and the Netherlands, and new regulations have been introduced in Austria as of January 1, 2011. Many competition authorities around the world (such as those in Australia¹, Austria², Bulgaria³, Germany⁴, Ireland⁵, Portugal⁶, to name just a few) have also conducted in-depths investigations of the gasoline market, as gasoline market are under notorious suspicion to be cartelized. Motoring organizations around Europe regularly call for investigations into market manipulation⁷, and in June 2011 the US Federal Trade Commission (FTC) said that it would begin another investigation to determine whether rising retail gasoline prices are the result of market manipulation or other anticompetitive behavior⁸, following several studies by the US Energy Information Administration (EIA). In fact, economic theory suggests that gasoline retail markets are indeed prone to collusive behavior. Oligopoly market structures prevail, market interactions occur frequently, prices are highly transparent, and demand is rather inelastic.

A recent sector inquiry by Germany's competition authority, the Federal Cartel Office, backed suspicions of tacit collusion (Bundeskartellamt, 2011) and recommended to consider regulatory pricing rules for gas stations similar to those implemented in Austria, parts of Australia, Luxembourg or parts of Canada.

¹See Australian Competition and Consumer Commission (2007).

²See http://www.bwb.gv.at/BWB/English/annual_report_2006.htm.

³See http://ec.europa.eu/competition/ecn/brief/04_2011/bg.si.pdf.

⁴See Bundeskartellamt (2011).

⁵See National Consumer Agency (2008).

⁶See <http://www.concorrenca.pt/en/publications/competition.asp>.

⁷See <http://www.guardian.co.uk/money/2012/jan/19/aa-inquiry-cost-diesel>.

⁸See <http://www.cspnet.com/news/fuels/articles/ftc-launches-gasoline-price-investigation>.

With the aim to bring down gas prices these rules either restrict the number of price changes per day or they limit the mark-up for gasoline retail prices.

The in-depth sector inquiry into the German gasoline market, which the Federal Cartel Office published in 2011, accused the five major gas station operators of having established a “competition-free” oligopoly that was characterized by tacit collusion. The sector inquiry revealed that the number of price changes has increased quite dramatically over the last years and that the two largest chains act as price leaders within the oligopoly. The leaders’ price changes are typically followed by all other gas stations within three hours, leading to almost perfectly parallel price developments and only small price differences. Furthermore, gas stations tend to uniformly raise gas prices during rush hours. The Federal Cartel Office thus concluded that the market lacks substantial competition and that the Government should consider establishing new pricing rules by law (Bundeskartellamt, 2011).

While the interpretation of the data by the Federal Cartel Office is debatable, as for example more frequent price changes among gas station typically tend to be a sign of more rather than less intense competition (Noel, 2007a), we do not wish to discuss the findings of the sector inquiry and its interpretation at length here. The focus of our study is rather on the proposed remedies, i.e. the different pricing rules suggested by the Federal Cartel Office.

In order to make collusion more difficult, the following pricing rules pricing regulations for gas stations have been proposed:

- The Austrian rule (in force in Austria since January 2011) which allows price increases only at 12 noon, while price cuts are always possible.
- The Western-Australian rule (also known as the Fuel-Watch-Concept) which was established by the state government of Western Australia in 2001. The regime constrains the gas stations to change their prices at most once a day (no matter whether up or down). These price changes must be announced at 2 pm the day before.

- Price ceilings (in force in Luxembourg and several Canadian states) which set a maximum mark-up for gasoline retail prices based on wholesale market prices.

Although all three regulations were established in their particular jurisdictions in order to protect consumers from exploitative price setting behavior by gas stations, a solid theoretical underpinning is missing at least for the Austrian and the Western-Australian rule while empirical or experimental support for the rules' effectiveness is even missing in all three cases.

To help filling this gap, we use a controlled lab experiment to analyze how the three pricing rules affect market outcomes. The setup of our experimental gasoline market is similar to the one studied by Deck and Wilson (2008), albeit with an entirely different focus. As Deck and Wilson, we use a spatially differentiated oligopoly model where robot consumers face certain transport costs. For a fictional product, four players can set prices within varying time intervals, depending on the implemented pricing rule. Both demand and input prices vary in order to closely reflect the realities of most gasoline markets. While demand fluctuations follow the same pattern over every four-round period (i.e., a full day in our game) and are communicated to all players at the beginning of the experiment, the periodical input price variations are random from the four players' perspective.

The results of our experiment are as follows: The Austrian rule tends to lead to higher prices and firms' profits compared to an unregulated market. The increase in both the mark-up and the profit level is statistically and economically significant. Similar results emerge for the Luxembourg scenario, even though we were only able to conduct the experiment for one particular price ceiling while one could test, at least in theory, an infinite number of price ceilings. The Western-Australian rule did not result in any price or profit level that have been different in terms of statistical significance when compared to an unregulated baseline market.

We conclude that governments should rather refrain from introducing the Austrian pricing rule, as it may lead to increased prices and thereby reduce consumer welfare. The Luxembourg rule is also unlikely to provide consumer benefits if the price ceiling is not binding, while we cannot explore the dynamic effects of the Luxembourg rule (including entry and exit) if the regulatory price ceiling is binding. Although the Western-Australian rule does not lead to significant improvements for consumers, it does not do much harm either. If there is a strong consumer preference for price stability over the day, our experiments suggest that the Western-Australian rule may at least be implemented without inducing price increases.

The remainder of our study now is organized as follows: Section 3.2 provides an overview over the related literature, while section 3.3 describes the experimental setting and derives hypotheses. Section 3.4 presents the results, before section 3.5 summarizes and concludes.

3.2 Related Literature

Apart from the many sector inquiries that have already been quoted above there is also a limited academic literature on pricing strategies in retail gasoline markets. A particular expert in this field is Michael Noel whose research has focused on the theory and practice of so-called Edgeworth price cycles. The theory of Edgeworth price cycles was first formalized in a seminal contribution by Maskin and Tirole (1988), who consider two identical firms that produce an homogeneous good and compete by setting prices in an alternating fashion. As Maskin and Tirole have shown two types of equilibria can possibly emerge. Either prices remain sticky or, alternatively, an asymmetric price cycle emerges, where prices rise fast and then fall slowly.

In an extension of this theory, Noel (2008) has shown that Edgeworth Cycles are a robust equilibrium even if there are cost shocks, as is typical in retail gasoline markets. Maybe even more importantly, Noel (2008) has also shown

that the inclusion of a third firm is compatible with the emergence of Edgeworth cycles. However, now situations can emerge where a firm may first increase its price, but then revert to a lower price if the other firms continue to compete at the bottom of the cycle for too long without following.

In Noel (2007a) weekly data on average prices is studied for 19 Canadian cities over eleven years, accounting for variation in concentration ratios across cities and over time. As Noel shows Edgeworth cycles are more prevalent where small and price aggressive independent gas stations have a larger presence. In addition, Noel finds that the presence of aggressive firms results in faster and taller cycles. The empirical findings support the argument that the cycles are generated by an Edgeworth Cycles process and that these cycles are associated with more competitive markets. Further support for these results is provided by a detailed study of the Toronto retail gasoline market in Noel (2007b).

Finally, while Noel (2009) has shown that Edgeworth cycles can be a cause of asymmetric cost pass-through in gasoline retail markets, Lewis and Noel (2011) find that pass-through rates are much faster in Edgeworth Cycles markets than in non-cycling markets. Hence, there is a welfare benefit to consumers in markets where there are Edgeworth Cycles.

There is also very limited experimental work on retail gasoline markets. The main contribution on experimental gasoline markets is the paper by Deck and Wilson (2008) which analyzes the effects of so-called zone pricing rules in wholesale gasoline markets, using a controlled laboratory experiment. Deck and Wilson investigate how zone pricing affects consumers, retail stations, and refiners in comparison to uniform wholesale pricing to retailers. In a nutshell, they find, firstly, that uniform wholesale pricing rules benefit retailers at the expense of consumers, and, secondly, that the retail prices are not as well related to underlying costs.

The experimental work by Engelmann and Normann (2009) and by Engelmann and Müller (2011) is also related to our Luxembourg rule, as their papers

show that price ceilings do not have any effect on prices or on the stability of collusion, as long as the price ceiling is not binding. Hence, their experiments cannot support the idea that price ceilings may serve as focal points that help to facilitate collusion.

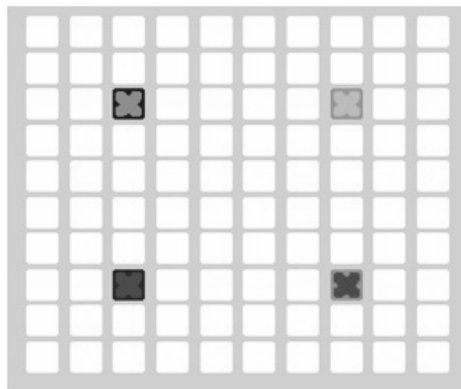
Finally, a recent paper by Berninghaus et al. (2012) provides a very simple experimental analysis of the Austrian rule, using a two player game where the two players can choose between four different prices. They show that the percentage of two-player pairs that play the collusive equilibrium (which yields twice the payoff of the competitive equilibrium) increases from 66 to 78 percent when the Austrian rule is implemented. However, their game lacks many of the features of most retail gasoline markets. First of all, they study only duopoly situations which are known to be much more prone to collusion than games with three or four players (see Huck et al., 2004; Isaac and Reynolds, 2002). Secondly, players can only choose between four different prices. Thirdly, there are no cost variations which, however, are a typical feature of retail gasoline markets. And fourthly, there are no demand fluctuations in their market.

3.3 Experimental Setting

We have developed an experimental gasoline market which reflects many key features of a real retail gasoline market. Firstly, it models the spatial oligopoly structure of many retail gasoline markets where customers are located at different points so that they face different transportation costs. Secondly, we have incorporated different demand levels at different times, as the majority of car owners tend to buy gasoline at rush hours rather than at noon or at night (Bundeskartellamt 2011). Thirdly, interactions are regular and occur repeatedly. And fourthly, wholesale prices (retail costs) fluctuate over time, as do wholesale gasoline prices in reality.

To implement these features into our experimental gasoline market, we use a modified version of the experimental setting of Deck and Wilson (2008). Our

Figure 3.1: 10x10-Hotelling-City with Four Shops (Gas Stations)



participants were randomly allocated into fixed groups of four, playing the role of a supplier of a fictional good which they sold at a chosen price to robot buyers. In order to prevent framing effects, the fictional good was not specified as gasoline. The shops were located symmetrically over a two-dimensional, quadratic 10x10-Hotelling-type city at blocks the (3,3), (3,8), (8,3) and (8,8), respectively, as depicted in Figure 3.1.

Buyers

A number of 10,000 robot buyers was uniformly distributed over our two-dimensional city. Each block was inhabited by 100 buyers, so that the entire city had 10,000 buyers in total. Each of the buyers purchased the fictional good once per day (where a day consist of four rounds, which reflect different times of day). However, buyers purchased the good at different daytimes. 35 percent of the daily demand was known to materialize in the morning and another 35 percent in the afternoon, while 20 percent of the buyers purchased at noon and the remaining 10 percent at night. Thus, 3,500 customers bought in the first and third round of each day, 2,000 in the second round and the remaining 1,000 in the last round, before a new day started. This rule held for the entire city.

The robot buyers automatically purchased at the shop where their total costs of purchase were minimal. The total costs of purchase at shop i consisted of

travel costs and the shop i 's retail price. The travel costs were known to be one "Taler" for every side of every block travelled. Buyers were assumed to only travel horizontally and vertically, but not transversely. Hence, the maximum distance that a buyer could travel was 14 blocks, for example if the buyers at block (10,10) would purchase at the shop at (3,3). The minimum distance was zero, for example if the buyers at block (3,3) bought at shop (3,3). If two or more shops offered prices such that the same minimum total costs of purchases resulted for a group of buyers at a given block, the buyers located in the respective block were automatically split up equally between these shops. All buyers were known to have a maximum willingness to pay of 40 Talers less their respective transport costs. Hence, buyers at block (10,10), e.g., only purchased at shop (8,8) if the price of that shop was not higher than 36 Talers and the other three shops did not offer prices which led to a smaller total costs of purchase. In general, buyer j only purchases the good if at least for one shop $p_i + t_{ji} \leq 40$ holds, where t_{ji} are the transport costs buyer j faces when travelling to shop i .

Sellers

Each participant in each market ran exactly one single shop in a fixed-matching setting against the same three rivals throughout the whole experiment, since repeated strategic interactions are a common feature of gasoline markets. All sellers had to choose their retail price simultaneously. When asked to choose their selling price, sellers were also informed about the current wholesale price and the market results of the previous round (daytime), i.e. the number of buyers at their shop, their profit, their market share and their rivals' selling prices, as rival price monitoring is a key characteristic of gasoline markets (see, e.g., Bundeskartellamt 2011). Sellers were not allowed to charge retail prices below the wholesale price.

The profit seller i made during a round was calculated as the number of buyers at shop i multiplied by the mark-up, which consists of the difference between i 's retail price p_i and the uniform wholesale price where the latter was always the same for all participants.

The uniform wholesale price which all sellers equally had to pay (as long as they made any retail sales) changed once a day (always before the first round of the day). As wholesale prices were always known before retail prices had to be chosen, sellers did not have to speculate or to form beliefs about future wholesale price developments in order to manage any inventories, as inventories were ruled out to avoid artificial effects induced by risk aversion and other characteristics of the participants.⁹

The next day's wholesale price was random from the participants' perspective and only known to fluctuate between 15 and 25 Taler. However, following Deck and Wilson (2008), the same price path, with a mean of 20 Taler, was used throughout all sessions in order to keep the results comparable.

Treatments

The experiment comprised two phases and each phase consisted of eight “days” which in turn were split into four “daytimes” (to reflect varying demands at morning, noon, afternoon and night). We used an identical wholesale price path in the first phase (days 1 to 8) and the second phase (days 9 to 16) of the game in order to facilitate an unbiased comparison of the two phases of the game.

In the baseline scenario, the suppliers were allowed to change their retail prices every round, that is four times per day. Thus, sellers could make 32 retail price choices over an 8-day phase.

Apart from the baseline scenario (B), in which participants were able to change their price in every single round and which represents the current unregulated situation in many countries, we have tested three different price setting rules:

- The Austrian rule (A), which allows prices to increase only once a day (in the morning), while price cuts are allowed at all four rounds of given day.¹⁰

⁹This contrasts with Deck and Wilson (2008) who have inventories in their game.

¹⁰This slightly differs from the true Austrian regime where any price increase must be

- The Western-Australian rule (W), which only allows one price change (whether up or down) per day at all. In order to keep the timing of the experiment constant, participants had to click through the four rounds of any day like in the baseline scenario, but were only able to change their retail price in the morning.
- The Luxembourg rule (L), which sets a maximum mark-up, by which the retail price may at most exceed the wholesale price. We have set a maximum mark-up of 8 Talers, which is above the two single stage Nash equilibrium mark-ups (which come to 5 and 6 Talers), for two main reasons: Firstly, a relatively high mark-up allows us to test the hypothesis that the maximum mark-up can serve as a focal point which helps to facilitate collusion. To test this hypothesis, a mark-up that exceeds the competitive benchmark is necessary. And secondly, we have chosen a mark-up of 8 Talers, as it is half of the cartel price, if we take the average purchasing price of 20 Talers as a basis.

To analyze the effects of the different pricing rules, we have conducted seven different treatments. In the first treatment, participants were playing the baseline game in both the first phase and the second phase. In the second, third and fourth treatments, the experiment started with one of the three pricing rules which was then repealed after phase one (days 1 to 8) for the second phase (days 9 to 16). In treatments 5, 6 and 7, this timing was reversed. The experiment started with the baseline scenario, followed by one of the three pricing rules which were implemented in the second stage.

This approach allows us to compare the results between treatments as well as within treatments. First, we are able to compare the behavior of groups which operate under different pricing rules from the beginning by looking at the first-phase results of treatments 2, 3 and 4 and comparing them to the other four implemented at 12 noon. We have chosen the first round of a day (morning) instead, so that participants did not need to form any expectations about the following morning's wholesale price.

treatments which all begin with the baseline game. Since the participants were only informed that the rules of the game may or may not change in the second phase of the game before they actually entered the second phase, but not how the rules may change, we can sum up the first-phase results of treatments 1, 5, 6 and 7 for this purpose. This facilitates a comparison between treatments.

Secondly, we are able to analyze the effects of introducing different pricing rules in the second phase by comparing the second-phase behavior of treatments 1, 5, 6 and 7 which all started with the baseline game in phase one. Table 3.1 summarizes the different treatments and gives the number of groups which played the corresponding treatments.

Table 3.1: Treatments

Treatment	Name	First phase	Second phase	# Groups
1	B-B	Baseline	Baseline	10
2	A-B	Austria	Baseline	12
3	W-B	Western Australia	Baseline	9
4	L-B	Luxembourg	Baseline	13
5	B-A	Baseline	Austria	11
6	B-W	Baseline	Western Australia	10
7	B-L	Baseline	Luxembourg	9

The experiment took place in the laboratory of the Duesseldorf Institute for Competition Economics (DICE) between November 2011 and January 2012. 296 subjects participated in 15 sessions which all lasted about two hours. 145 participants were females (49 percent), and the mean age was about 24 years. Participants received a show-up fee of 4 Euros and the sum of their total profits, converted into Euros at an exchange rate of 10,000 Talers to 1 Euro. The total amount was rounded up to full multiples of 50 Eurocents. Participants earned an average total amount of 19.32 Euros (median: 19 Euros).

Hypotheses

Our experimental gasoline market has three theoretical reference points: First of all, there is the cartel case with a price of 36 Talers which maximizes the suppliers' profits. At a price of 36, the market is completely covered, since no buyer has to drive more than four lines to reach the closest shop (and 40 Talers is the buyers' maximal willingness to pay). However, if all players charge a price of 36, every single player has an incentive to deviate downwards, as a price of 35 generates a higher profit if the three other firms charge a price of 36.

The second and third reference points are the two symmetric Nash equilibria of the stage game with mark-ups of 5 and 6 Talers, respectively (with the latter being the payoff-dominant equilibrium). If one's three rivals charge this mark-up, no additional profit can be gained by deviating upwards or downwards. In contrast, if one's three rivals charge a mark-up of less than 5 Talers, it is beneficial to raise the price, while it is profitable to lower one's price, if the rivals' uniform mark-up is larger than 6 Talers.

From a purely theoretical perspective, one might expect one of the two symmetric Nash equilibria to emerge (also, since the game ends with certainty after 64 rounds). However, as is well known from experimental economics, other outcomes regularly emerge.

Focusing on the effect of our pricing rules, it is clear that they do not change the symmetric Nash equilibria nor the collusive outcome. With respect to the Western-Australian rule, where only one price change per day is allowed, the effect on the likelihood of a collusive outcome is ambivalent at the outset. On the one hand, deviation becomes more profitable as one's rivals can only react with some delay. On the other hand, any punishment is also more severe, as one cannot directly react if being undercut. Prices also become easier to monitor as they change less frequently. In this context it is also useful to recollect the theoretical result of Noel (2008) that in retail gasoline markets with more than two players situations can emerge where one firm may first increase its price, but

then revert to a lower price if the other firms continue to compete at the bottom of the cycle for too long without following. The incentive to “lead the pack” with a price increase may be reduced if one cannot easily revert to a lower price.

This last point would not concern the Austrian rule where one can easily increase the price and then reduce it again once the rivals do not follow. Hence, a price increase that is not followed by the rivals is less costly and, therefore, a price increase less risky under the Austrian pricing rule. Hence, we would expect prices to be higher on average under the Austrian pricing rule than under the Western-Australian rule.

In addition, recent empirical studies on gas station cartels (Andreoli-Versbach, 2011) as well as on other cartels (von Blanckenburg et al., 2012) have found that, in order to make monitoring easier, price changes occur less often when markets are cartelized than in competitive markets (Noel, 2007a). In line with this research, we expect the Austrian and the Western-Australian rule rather to facilitate than to prevent collusion.

Regarding the Luxembourg rule, previous experimental research by Engelmann and Normann (2009) and Engelmann and Müller (2011) has not found any focal point effect. Hence, we would expect the non-binding price ceiling of an 8 Taler mark-up (which exceeds the two symmetric Nash mark-ups of 5 and 6 Talers) not to have any effect.

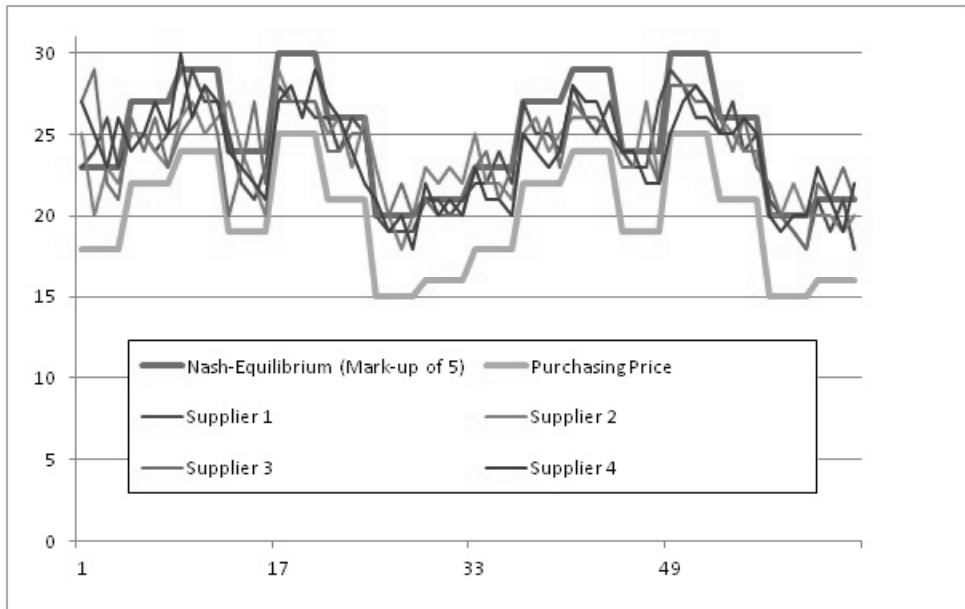
3.4 Results

3.4.1 Competition Dynamics

Representative Group

Figure 3.2 shows the retail prices chosen by a representative group of the first treatment (B-B). Mean profits and the dispersion of retail prices are most close to the mean values of all groups of treatment 1, so this one was chosen as a representative group. The thick blue curve, depicting the mean selling prices per

Figure 3.2: Representative Group in Treatment B-B



daytime, clearly reveals that prices were almost always chosen below the stage game Nash equilibria, a result that holds for a vast majority of groups within our experiment.

Figure 3.3: Mark-ups and Wholesale Prices

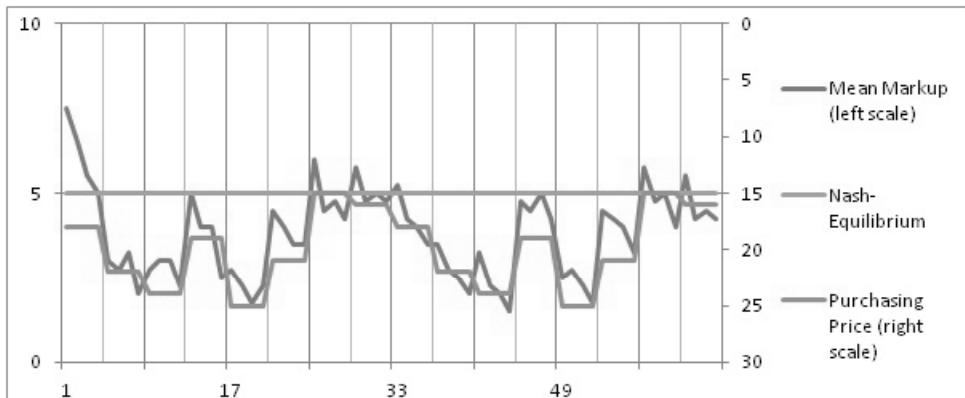


Figure 3.3 shows the representative group's mean price mark-ups (the difference between retail and wholesale prices). The orange curve depicts the respective absolute wholesale prices on an upside-down scale (right scale). The picture emerges that retail prices (blue line) evolved negatively correlated to the whole-

sale prices, which is also a common result for other groups. Thus, wholesale price changes were not fully passed through to the buyers.

Mark-ups Depending on Daytime

Figure 3.4: Average Mark-ups Over the Day

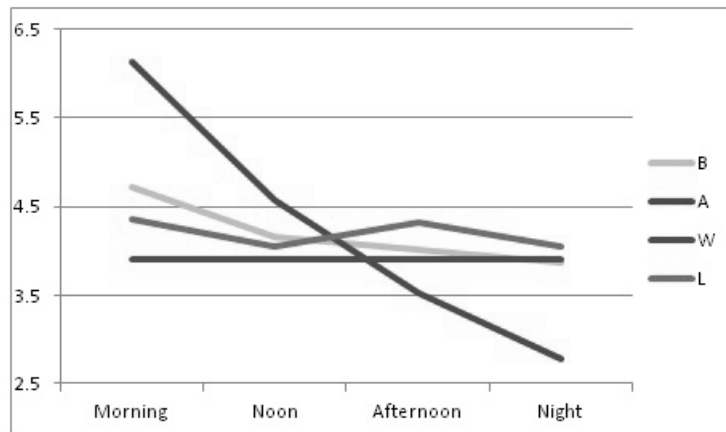


Figure 3.4 depicts the mean mark-ups, depending on the implemented pricing rule and the time of day. The red curve, showing the results under the Austrian regime, is by definition downward sloping, since participants were not allowed to increase prices, but had to keep them constant or even cut them during the day. Nevertheless, the decline is quite large, as the mark-up more than halves from morning to night. The purple curve (Western Australian regime) is by definition flat, since participants were not allowed to change their retail prices during the day.

According to the Bundeskartellamt (2011), it is a common feature of the (German) gasoline market that prices peak during rush-hours, that is in the mornings' and the afternoons' of our setting. Figure 3.4 therefore illustrates the experimental results under the Luxembourg rule and the baseline scenario (where price changes are not regulated). As it turns out, the groups under the Luxembourg regime on average followed the path described by the Bundeskartellamt (2011) quite accurately. Mark-ups are high in the morning and in the afternoon and low during low-demand times of day. In the baseline scenario, the picture however

is less clear. Average mark-ups on average decline constantly after the morning, although the difference between noon and afternoon is quite small.

Focal Point

In treatments B-L and L-B which both included the Luxembourg rule (either in the first or second phase), the maximum mark-up of 8 Taler was hardly ever chosen. In the first phase of the L-B treatment, only 2.5 percent of the chosen mark-ups were equal to 8, while no group ever completely agreed on this maximum mark-up. The picture slightly changes once the second-phase results of the B-L treatment are taken into account. Here, 18.7 percent of the mark-ups equaled the maximum markup of 8, and there were 29 occasions where entire groups “agreed” on this maximum level. However, this was almost exclusively due to the behavior of one single group which managed to “agree” on the maximum mark-up 20 times in a row. Among all other groups of treatment B-L, only one percent of the chosen mark-ups equaled 8 Taler.

3.4.2 Consumer Welfare

A pricing rule may increase consumer welfare if it results in lower retail prices and, therefore, lower supplier profits. In fact, this expectation was the main impetus from a political perspective to implement the rules. This subsection tests the hypotheses that the three price regulation rules meet this requirement. Put differently, the question is whether the rules achieve their political objective.

First-Phase Results

We now turn to the comparison between the different treatments and start by comparing the first-phase results. Figure 3.5 shows the groups’ mean daily profits in phase one in treatments A-B, W-B, L-B and B-X. The latter embraces the treatments B-B, B-A, B-W and B-L, since these groups started with the baseline scenario game. Figure 3.6 depicts the groups’ weighted average daily mark-up.

Figure 3.5: Weighted Average Profits in Phase 1

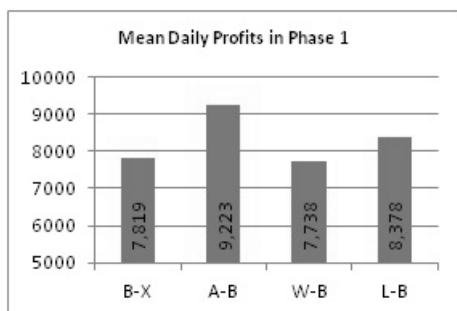
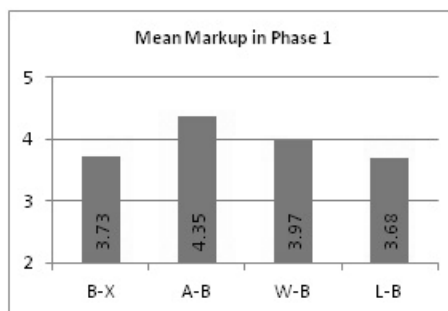


Figure 3.6: Weighted Average Mark-ups in Phase 1



When computing these average daily mark-ups, the values of the four rounds per days (the daytimes) were weighted by the differing demand levels.

The two figures show that none of the three pricing rules leads to noteworthy improvements for consumers. Suppliers in treatment B-X gained the second smallest average profit (only the groups in Western-Australian regime earned slightly less), and they charged the second lowest mark-ups (only the groups in Luxembourg regime charged slightly less).

According to a non-parametric Mann-Whitney test which uses the groups' mean values of phase one, the groups in treatment A-B made higher profits than those in treatments B-X (weakly significant), while the difference is statistically significant at the five percent level when a two-group mean-comparison T-test is used. Regarding the mean mark-ups, both tests reveal that A-B groups charged higher mark-ups than B-X groups (statistically significant at the five percent level). Comparing the B-X groups to the W-B and L-B groups, respectively, shows no significant differences regarding the groups' mean daily profits or daytime mark-ups.

Figure 3.7, depicting the groups' mean daily profits (solid lines, left scale) and the weighted average daily mark-ups (spotted lines, right scale), confirms this impression: The red line (Austrian price regime) is almost always above the orange line (B-X), while the others do not seem to differ qualitatively.

Figure 3.7: Mean Daily Profits and Mark-ups in Phase 1

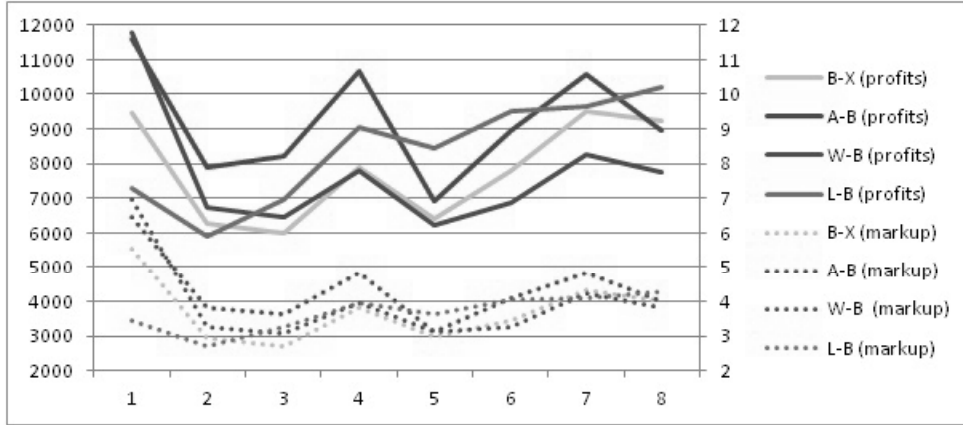


Table 3.2: Random Effects Panel Regressions for Phase 1

Variable	Daily Profits	Daily Averaged Weighted mark-ups
Method	RE	RE
Regimes:		
Baseline	(base)	(base)
Austrian	1403.1561** (-0.0414)	0.6281** (-0.0329)
W.- Australian	-81.3370 (-0.8938)	0.2394 (-0.5008)
Luxembourg	558.2365 (-0.4150)	-0.0467 (-0.8659)
Wholesale Price	-345.1227*** (<0.0001)	-0.2158*** (<0.0001)
Day	-38.8273 (-0.6441)	-0.1651*** (<0.0001)
Constant	14896.6502*** (<0.0001)	8.7843*** (<0.0001)
Observations	2368	2368
Subjects	296	296
Groups	74	74
R2 (overall)	0.1129	0.1263
R2 (within)	0.1431	0.1701
R2 (between)	0.0567	0.0415
p-values in brackets; *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$		

Our regression analysis further substantiates the impression. Table 3.2 presents the results of a random-effects panel model (with group-clustered standard er-

rors) which we chose to use in order to account for the panel structure of the data. We use the subjects' mean daily profits of the first phase or their weighted average price mark-up, respectively, as dependent variable.

As regressors we include a time trend (since Figure 3.7 suggests that profits and mark-ups change over time), the corresponding wholesale price (since profits and mark-ups seem to vary inversely proportional to the wholesale price) and dummy variables representing the three pricing regimes.

The regressions reveal that profits and mark-ups are significantly higher in the Austrian regime compared to the baseline scenario: All other things equal, groups in the Austrian regime gain an additional 1,403 Talers per day (plus 17.9 percent) and charge additional 0.63 Talers (plus 16.9 percent). The coefficients further show that, compared to the baseline scenario, profits are higher under the Luxembourg regime, but prices are lower, while the opposite appears to hold for the Western-Australian regime. Note, however, that here all coefficients are insignificant. As expected, the wholesale price has a negative influence on profits (highly significant). There is also a negative effect of time on mark-ups.

Second-Phase Results

Figure 3.8: Weighted Average Profits in Phase 2

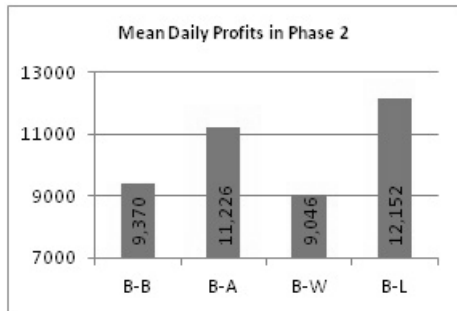
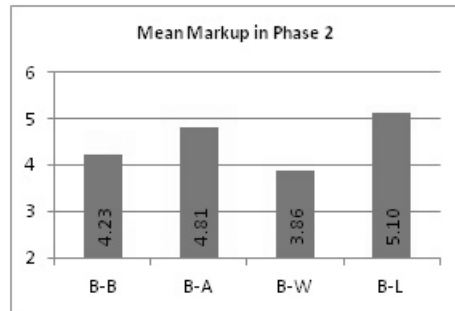


Figure 3.9: Weighted Average Mark-ups in Phase 2



We now compare the second-phase results of those treatments which started with the baseline scenario in order to check the robustness of our results. Figure 3.8 shows the mean daily profits per treatment, while figure 3.9 depicts the mean

mark-ups. A picture emerges that groups under the B-W regime made the smallest profits and charged the smallest mark-ups, followed by the B-B treatment groups. Both the B-A and the B-L treatment groups earn and charge considerably more compared to the two other treatments. Profits and mark-ups were highest in the Luxembourg regime.

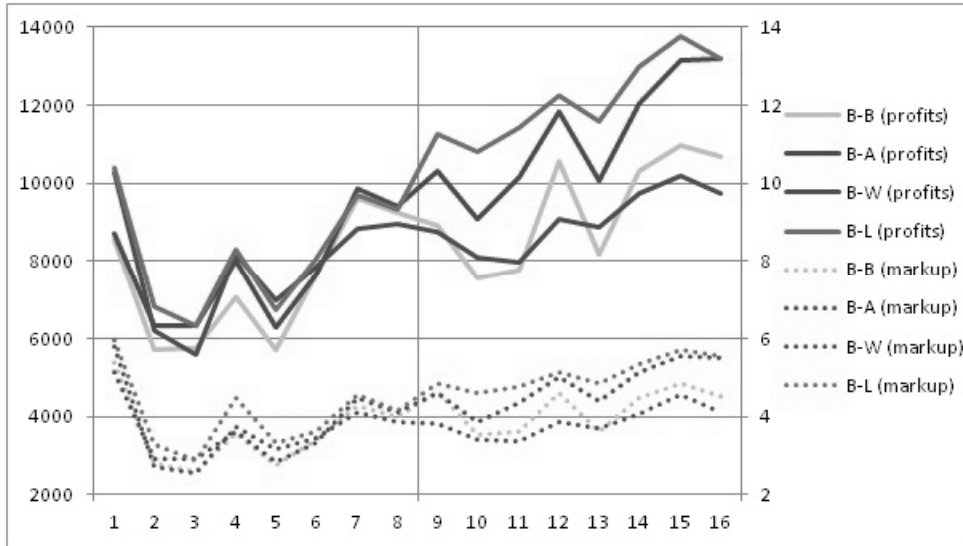
Using a non-parametric Mann-Whitney test as well as a two-group mean-comparison T-test reveals that the mean values did not significantly differ between the groups in the first phase. This was expectable, since all these groups started with the baseline game and were not previously told which price-setting regime would later follow in phase two. Hence, we are largely confident that observed differences are caused by the different pricing rules and not by other characteristics of the groups.

Comparing the second-phase results of the different treatments (group averages), using a non-parametric Mann-Whitney test, reveals that profits were higher in the B-A treatment compared to the B-B treatment (weakly significant), a result which is confirmed by a two-group mean-comparison T-test. According to both tests, mark-ups did not differ significantly. A T-test further finds profits (five percent level) and mark-ups (ten percent level) to be significantly higher in the B-L treatment compared to the B-B treatment, while the Mann-Whitney test reveals no significant differences. Furthermore, both tests reveal no significant differences between the B-B and the B-W treatment.

Figure 3.10, again showing mean daily profits (solid lines, left scale) and the weighted average mark-ups (spotted lines, right scale), further reveals how the treatment-results diverged after the beginning of phase two. While they were fairly similar in phase one (during which the groups of all four treatments played the baseline game), the red (B-A) and the green (B-L) lines are always located at higher levels throughout phase two.

Random effects panel regressions with data of the second phase, whose results are presented in Table 3.3, support the previous impression: Compared to the

Figure 3.10: Mean Daily Profits



baseline scenario, higher daily profits can be reported for the Luxembourg regime (2782 Talers, or 29.7 percent higher, all other things equal) and the Austrian (1856 Talers or 19.8 percent) regimes (weakly significant in both cases). Also the mark-ups differ systematically, albeit not statistically significant. In contrast, the Western-Australian regime does not appear to make a difference.

3.5 Conclusion

Having compared the results both phases, the following picture emerges: While it can be reported that the Austrian regime (in both phases) and the Luxembourg rule (only in the second phase) tend to increase mark-ups (and profit levels) compared to the baseline scenario, no significant differences have been found between the Western-Australian regime and the unregulated baseline scenario.

Policy makers thus should be rather skeptical about the merits of the Austrian or Luxembourg pricing rule. The Western-Australian regime does not appear to lead to lower average prices and profit level either, but, at least, it does not harm consumers.

Table 3.3: Random Effects Panel Regressions for Phase 2

Variable	Daily Profits	Daily Averaged Weighted mark-ups
Method	RE	RE
Regimes:		
Baseline	(base)	(base)
Austrian	1856.2074* (-0.0945)	0.5788 (-0.1718)
W.- Australian	-324.2212 (-0.7409)	-0.3725 (-0.3422)
Luxemb.	2781.8435* (-0.0519)	0.8673 (-0.1116)
Purchasing Price	-166.3161*** (<0.0001)	-0.0856*** (<0.0001)
Day	297.5162*** (<0.0001)	0.0846** (-0.0161)
Constant	8977.3734*** (<0.0001)	4.8856*** (<0.0001)
Observations	1280	1280
Subjects	160	160
Groups	40	40
R2 (overall)	0.2608	0.1362
R2 (within)	0.1916	0.1256
R2 (between)	0.1619	0.1188
p-values in brackets; *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$		

Regarding the focal point theory, our experimental results are somewhat surprising. Although summary results in subsection 3.4.1 suggested that the maximum mark-up in the Luxembourg regime does not serve as a focal point, since the maximum mark-up was hardly ever chosen, it turned out that, in the second phase, profits were (at least weakly significantly) higher compared to the baseline scenario. Thus, one might conclude that although a maximum mark-up does not make suppliers to choose exactly this maximum mark-up, it may still induce suppliers not to undercut the focal point by too much, thereby leading to higher profits than in the non-regulated market.

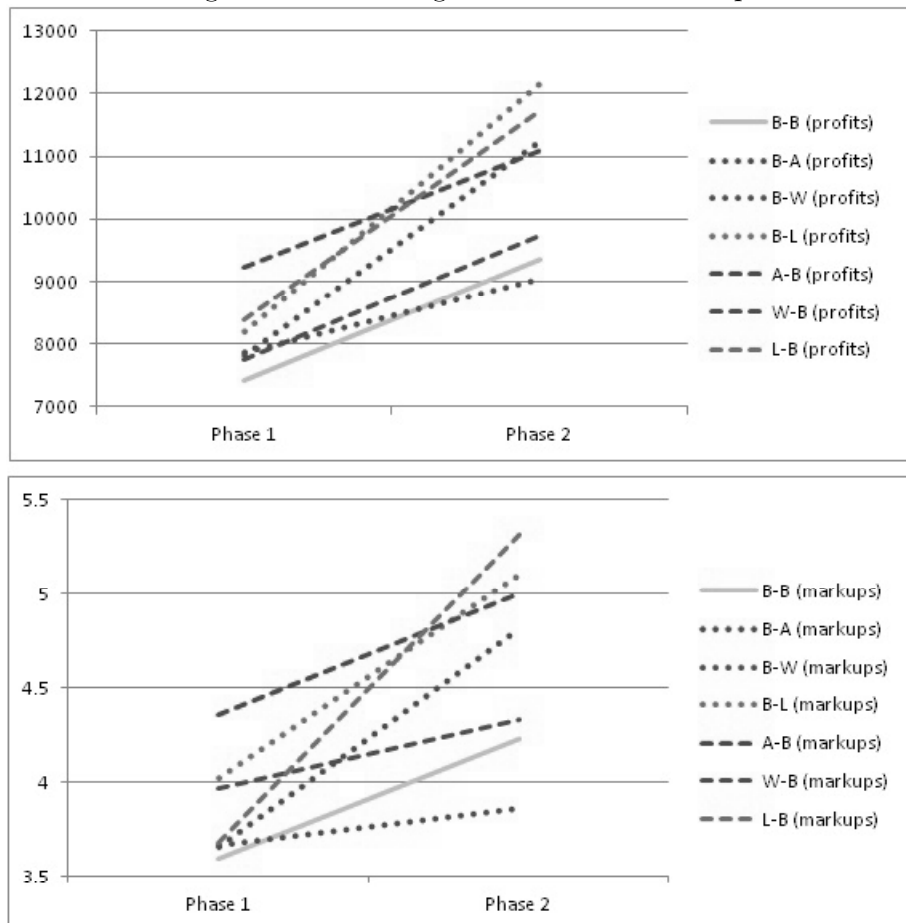
3.6 References

- Andreoli-Versbach, P. (2011), Collusion Through Delayed Pricing: Evidence from the Italian Petrol Market, Working Paper, University of Munich.
- Australian Competition and Consumer Commission (2007), Petrol Prices and Australian Consumers: Report of the ACCC Inquiry into the Price of Unleaded Petrol, ACCC: Canberra.
- Berninghaus, S., Hesch, M. and Hildenbrand, W. (2012), Zur Wirkung Regulatorischer Preiseingriffe auf dem Tankstellenmarkt, *Wirtschaftsdienst* 92, 46–50.
- von Blanckenburg, K., Geist, A. and Kholodilin, K. (2012), The Influence of Collusion on Price Changes: New Evidence from Major Cartel Cases, *German Economic Review*, 13(3), 245–256.
- Bundeskartellamt (2011), Fuel Sector Inquiry: Final Report May 2011, Bundeskartellamt, Bonn.
- Deck, C. and Wilson, B. (2008), Experimental Gasoline Markets, *Journal of Economic Behavior and Organization* 67, 134–149.
- Engelmann, D. and Normann, H. (2009), Price Ceilings as Focal Points? An Experimental Test, in: J. Hinloopen and H.-T. Normann (Eds.): *Experiments for Antitrust Policies*, Cambridge University Press: Cambridge, 61–80.
- Engelmann, D. and Müller, W. (2011), Collusion Through Price Ceilings? In Search of a Focal-Point Effect, *Journal of Economic Behavior and Organization* 79, 291–302.
- Huck, S., Normann, H. and Oechssler, J. (2004), Two are Few and Four are Many - On Number Effects in Cournot Oligopoly, *Journal of Economic Behavior and Organization*, 53, 435–446.

- Isaac, R. and Reynolds, S. (2002), Two or Four Firms: Does It Matter?, in Charles A. Holt and R. Mark Isaac (Eds.), *Research in Experimental Economics*, Vol. 9: Experiments Investigating Market Power, Elsevier Science Ltd.
- Lewis, M. and Noel, M. (2011), The Speed of Gasoline Price Response in Markets with and without Edgeworth Cycles, *Review of Economics and Statistics* 92(3), 672–682.
- Maskin, E. and Tirole, J. (1988), A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles, *Econometrica*, 56(3), 571–599.
- National Consumer Agency (2008), *Investigation into Petrol and Diesel Price Movements*, National Consumer Agency: Dublin.
- Noel, M. (2007a), Edgeworth Price Cycles, Cost-based Pricing and Sticky Pricing in Retail Gasoline Markets, *Review of Economics and Statistics* 89, 324–334.
- Noel, M. (2007b), Edgeworth Price Cycles: Evidence from the Toronto Retail Gasoline Market, *Journal of Industrial Economics* 55, 69–92.
- Noel, M. (2008), Edgeworth Price Cycles and Focal Prices: Computational Dynamic Markov Equilibria, *Journal of Economics & Management Strategy*, 17(2), 345–377.
- Noel, M. (2009), Do Gasoline Prices Respond Asymmetrically to Cost Shocks? The Effect of Edgeworth Cycles, *RAND Journal of Economics* 40, 582–595.

3.7 Appendix 1: Figures

Figure 3.11: Average Profits and Mark-ups



3.8 Appendix 2: Instructions

3.8.1 Phase 1 (*B-A Treatment*)

Welcome to this experiment! Thank you for your participation!

This is an experiment about decision making. During the experiment, please do not talk to the other participants in the room. If you have any questions, please raise your hand, then we will come to your place. After the experiment you will be rewarded according to your performance in the experiment. You will receive the respective payment in cash.

Structure of the experiment:

The experiment consists of two phases. The instructions you are currently reading only pertain in the first phase. After the end of the first phase, we will stop the experiment and distribute the instructions for the second phase. After the end of the second phase, we will kindly ask you to fill in a questionnaire. Afterwards, we will pay you the amount of money that you won.

The first phase consists of eight game-“days”, which in turn consist of four “day-times” (morning, noon, afternoon and night). Altogether, you have to make 32 decisions in the first phase.

Your task:

Your role in this experiment is that of a seller. You buy a fictional product at the purchase price and sell it at your chosen sale price on to the buyer. Within your group, there are three other sellers. The game currency in this experiment is called Talers.

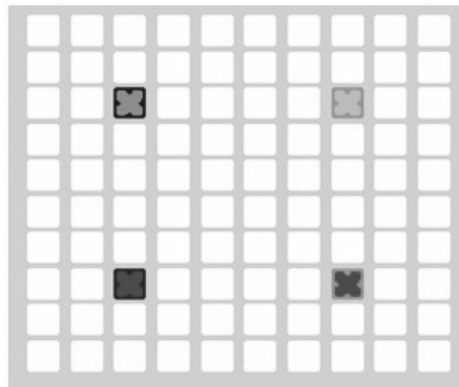
Any buyer who buys from you earns you a profit which is equal to the difference between the sale and the purchase price. You may not choose any sale price that is smaller than the purchase price.

The purchase price you have to pay does not differ from the price that has to be paid

by the other sellers in your group. However, the purchase price will change on every game-day. On top of the screen, the currently valid purchase price will be displayed. The three other sellers in your group were chosen randomly and are also located somewhere in this room, but you do not know them. In both phases of the experiment, you are playing against the same three other sellers.

The buyers:

The buyers will be played by the computer. They live on a quadratic field, which consists of 100 points (see below). 100 buyers live on every point, so there are a total of 10,000 buyers. The colored dots mark the shops where the sellers sell their products. At the top of the screen, you can see which shop belongs to you (i.e., what color your shop has). The color and the location of your shop does not change during the entire experiment.

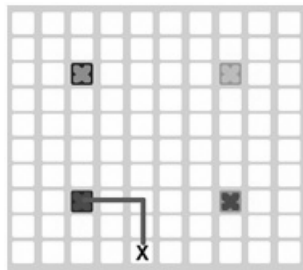


At every game-day, each buyer wants to buy exactly one time and one good. However, not all buyers go shopping at the same time of day: In the mornings and in the afternoons, the number of buyers is greater than at noons or in the nights. The exact distribution is as follows: 35 percent of the buyers go shopping in the morning, 20 percent at noon, 35 percent in the afternoon and 10 percent at night. Thus, the number of buyers is 3,500 in the mornings and in the afternoon, 2,000 at noons and 1,000 at night.

Each buyer buys only if her total buying costs are not greater than the amount of money that is available per day. This amount is called “maximum willingness to

pay” and equals 40 Talers. The total buying costs consists of the selling price and the individual transportation costs. These transportation costs are the higher the further a buyer has to travel to a shop. Each step from one point to the next one will cost the buyer one Taler of transportation costs. A buyer can only travel horizontally or vertically, but not diagonally.

Example: A buyer living on the point marked with “X” wants to buy at the red store. She has to run four fields to get there. The transportation costs in this case are four Talers.



A buyer will always automatically go to the shop which holds the best offer for her ready. That means she will choose the shop where the sum of selling price and individual transportation costs are the smallest.

Once a buyer would have to pay more than 40 Talers even at the shop with the best offer, she will not go shopping at all. When the buyers living at one point would have to pay the same total buying costs at two or more shops, the buyers will split up equally between the respective shops.

Procedure of Phase 1: You and the other three sellers of your group always have to choose a selling price simultaneously. Therefore, the input screen pops up where you please type in your chosen selling price. Afterwards, please confirm with a click on “OK”. By the time all sellers have chosen a price, the computer calculates how the buyers are distributed to the four shops.

Afterwards, you will see the results screen. Here, the prices of all four sellers are listed, further the number of buyers who bought at your shop and how much profit you have made.

Your profit at a particular daytime is calculated as follows:

$$\text{Profit} = (\text{selling price} - \text{purchase price}) \times \text{number of buyers}$$

In addition, your total profits are listed. On the right side of the results screen, you also see a figure that depicts how the buyers are distributed to the individual stores. The graph therefore represents the respective market shares. A point on the field always represents exactly one percent market share. The color of a field indicates to which shop the buyers who are living on this particular point went. A gray point means that the buyers have split up between two or more shops, since their total buying costs were exactly identical. A white point means that buyers did not go shopping at all, since their total buying costs were higher than the maximum willingness to pay of 40 Talers. After the results screen is displayed, the next daytime begins and you must select a selling price again.

End of the experiment:

After the end of the second phase, we ask you to fill in the receipt we passed out to you as described on the screen. Afterwards, we kindly ask you to answer the questions of the questionnaire. Following this, we will pay out your total profit in cash. The total profit is converted into Euros at an exchange rate of 10,000 Talers = 1 euro.

The euro amount is rounded up to the next higher 50-cent amount. In addition to your total profits, you will receive a show-up fee of four Euros for participating in the experiment.

If you still have questions, please hold your arm out of the cabin!

3.8.2 Phase 2 (*B-A Treatment*)

Procedure of Phase 2: The second phase of the game is constructed identically to the first, with one exception: You can only increase your selling price in the mornings. Afterwards, you are only allowed to keep your selling price constant or to reduce it.

You may therefore continue to freely choose a selling price at each morning. But on the following three daytimes of a day (noon, afternoon and night), you must not choose a selling price which is greater than the one of the previous daytime.

If you still try to choose a forbidden selling price, a message appears telling you that this is not possible. However, you can then choose a correct price.

The second phase again lasts eight game-days, which in turn consist of four daytimes each. Thus, you have to choose a selling price 32 times again.

If you still have questions, please hold your arm out of the cabin!

Chapter 4

Forecast Errors in Undisclosed Management Sales Forecasts: The Disappearance of the Overoptimism Bias

Abstract

Previous empirical evidence which evaluated the accuracy of management earnings or sales forecasts consistently revealed these forecasts to be on average significantly overoptimistic. However, all studies analyzed forecasts from public disclosures, which are an important signal to investors and analysts and thus possibly biased by strategic considerations. To disentangle whether and to which extent strategic deception or cognitive biases are responsible for this overoptimism, the present study analyzes the accuracy of 6,234 undisclosed, company-internal sales forecasts, which German firms provided anonymously to the IAB Establishment Panel. Quite surprisingly, the study reveals the average forecast to be significantly overpessimistic. I propose that the non-existence of a general bias towards overoptimism is due to the lack of incentives to consciously overgloss future prospects in undisclosed forecasts, and that *overpessimism* may be a consequence of loss aversion.

Key Words: *Management forecasts, Overoptimism, Overpessimism, Germany*

JEL-Classification: *D22, D84, L21, M41*

The dataset used in this study is based on the IAB Establishment Panel (Waves: 1993-1997). Controlled remote data access was kindly provided by the Research Data Centre (FDZ) of the Institute for Employment Research (IAB).

4.1 Introduction

Being able to accurately forecast your firm’s future is a key to success and survival in hard-fought markets. If managers, for example, have to decide about their future production capacity, over- and underestimating future demand or costs can be dangerous. In forecasting research, quadratic loss functions are commonly used to account for this danger, assuming that the damage of bad forecasts increases exponentially.

In recent years, however, empirical researchers in the fields of behavioural economics, industrial organization or accounting provided overwhelming evidence that managers’ assessments of their ventures’ future prospects are on average too optimistic. Mergers and acquisitions, for example, fail to achieve their intended goals in far more than every second case¹; public infrastructure projects face cost overruns in almost nine of ten cases (Flyvbjerg et al., 2002); and also firms’ disclosed sales or earnings forecasts turned out to be too positive on average, whenever they were analyzed.

This clear evidence raises the question about the main reasons for this overoptimism bias. Kahneman and Lovallo (2003a) argue that overoptimism on the one hand occurs due to an unconscious cognitive bias, which is mostly called the “planning fallacy” in the behavioral literature. If a firm is successful, the managers may wrongfully trace this development back to their own skills and decisions in the past, rather than to luck or other factors that cannot be influenced by the firm itself.

This “misattribution of cause” (Camerer and Malmendier, 2007) may lead to too much optimism about future outcomes, if the lucky streak ends and external conditions worsen. Furthermore, managers may underestimate the probability of expensive or time-consuming problems because they oversee that, although each thinkable single risk may occur with a low probability, the chance that *none* of these dangers occurs at all is pretty low.

¹See Straub (2007) for an elaborate overview over the literature.

On the other hand, Kahneman and Lovallo (2003a) argue that overoptimism may further be amplified, if forecasters can benefit from announcing promising prospects. If, for example, a principal is known for explicitly disliking bad news, employees may consciously sugarcoat their estimates about the likely success of a project. Flyvbjerg et al. (2002) argue that, since many projects only get started, if its prospects are good enough, managers may often choose the most overglossed project instead of the one with the objectively best prospects. Thus, if such organizational pressures are present, the forecasters have strong incentives to cheat against their own best knowledge, often referred to as “strategic deception”. This behavior may likely produce forecasts which later turn out to be overly optimistic.

But there may not only be internal pressures, but first and foremost also external ones: Listed joint-stock companies may have incentives to publish too optimistic forecasts of their future sales or earnings, since these forecasts are an important, if not the most important, signal to analysts and investors (Pedwell et al., 1994). Using Japanese data, Ota (2010), for example, provides evidence how closely analysts follow managements’ forecasts when providing their own forecasts, so joint-stock companies may benefit from (mis-)leading the market by stating that their prospects are bright.

Hence, overoptimism may be caused by strategic deception as well as by unconscious cognitive biases. From the point of view of behavioural economics, the question arises, to what extent and - in which kind of situations - these approaches are responsible for the established overoptimism. Flyvbjerg (2003) criticized Kahneman and Lovallo (2003a) for underrating the likely influence of strategic deception in their seminal paper. Kahneman and Lovallo (2003b) retorted by emphasising that the cognitive bias in their eyes is the main reason why the majority of forecasts is biased upwards.

Finding the foundations of overoptimism is surely of great interest: If a firm unconsciously overestimates its future sales, and plans its capacity and workforce according to these forecasts, it will have to pay the price for its wrong forecasts.

If a firm, however, publishes overglossed forecasts in order to mislead the market, their investors have to bear the damage, disregarding possible negative reputation effects.

However, empirical evidence which tries to disentangle the influence of both causes is still scarce. To the best of my knowledge, only Rogers and Stocken (2005) contributed to this research field. They analyze earnings forecasts of almost one thousand US-companies between 1995 and 2000 and find that management forecasts are less biased, if managers are in danger of being sued for intentionally misleading the market or if the market's ability to verify the forecasts is high (the latter is measured by the analysts' agreement on a firm's prospects). Thus, they can reveal some evidence for intentional misrepresentation.

The present study tries to further close the research gap with a new approach. I analyze the biasedness and accuracy of managements' sales forecasts from 2,511 German firms (timeframe: 1993-1997). Unlike all previous studies about management earnings or sales forecasts, I use company-internal forecasts (which were provided secretly to the anonymized IAB Establishment Panel) instead of publicly disclosed ones. Every year in June, a large sample of firms is asked to provide a forecast about their sales for the upcoming business year, so the forecast period starts six and ends 18 months later.

I chose to analyse sales instead of earnings forecasts because this key figure, compared to earnings, is less subject to possible misleadings through creative accounting or earnings management.² Furthermore, I do not restrict my analysis to data from joint-stock companies ("Aktiengesellschaften"), as all previous studies did, and use data from firms with all different legal forms and sizes instead.

Since the forecasts were retrieved under secrecy and the dataset is strictly anonymized, the forecasters could not use their forecasts as a tool to influence the market and hence faced no (external) pressures to cheat intentionally. Thus, I aim at testing the hypothesis that management forecasts, compared to pre-

²See Kasznik (1999) and Matsumoto (2002) on the relationship of forecast accuracy and accounting decisions.

vious evidence, reveal less overoptimism in situations where managers have no incentives (at least no external ones) to state different numbers than they truly believe. The second hypothesis assumes that joint-stock companies display more overoptimism compared to other legal forms (like limiteds). This appears to be rather likely, since their managers may not be able to fully abstract from the pressure to report good news which comes from the financial markets.

My results add some interesting new aspects to the literature: Quite surprisingly, I find the average firm's sales forecast to be rather *overpessimistic* than overoptimistic (and am to my best knowledge the first to do so): The mean (median) forecast error is -5.58 (-0.31) percent (measured as the difference of the forecasted and the actual sales numbers, divided by the forecasted value). While previous literature always reported a majority of too optimistic forecasts, I find a majority (50.90 percent) to be too *pessimistic*. Hence, I am able to provide some support for Flyvbjerg's (2003) hypothesis that overoptimism strongly depends on the forecasters' incentives to cheat: In this case there are no external benefits from strategic deception and no overoptimism can be found on average.

But since the average forecast is found to be even *overpessimistic*, this explanation is not sufficient. I argue that firms may dislike bad surprises to a greater extent than they like good surprises, so their strategic overpessimism may be an expression of loss averting behaviour.

A subsample that is restricted to data from joint-stock companies reveals partly different results: Here, the mean forecast error is -1.57 percent, hence overpessimistic as well, though less than within the whole sample. Furthermore, the majority of forecasts from joint-stock companies is overoptimistic, since the median forecast error is 1.96 percent.

Using probit and logit estimation methods, further evidence can be provided that joint-stock companies appear to be more overoptimistic than other firms, while controlling for a broad range of micro- and macroeconomic values. Furthermore, hints for the existence of an "misattribution of cause"-bias could be

found: Current success (measured by the sales growth-rate) is highly significantly related to the firms' overoptimism, as suggested by a third hypothesis. OLS and between-effects panel regressions further provide some support for the fourth hypothesis that firms with a higher share of women among the workforce are less overoptimistic, which is at least related to previous evidence about gender differences which showed women to be less overconfident and -optimistic. No support can be provided for the hypothesis that younger firms display more overoptimism.

The remainder of this study is organized as follows: Section 4.2 introduces the conceptual framework. Section 4.3 then sums up previous literature, while section 4.4 describes the IAB Establishment Panel. Section 4.5 derives five hypotheses, provides descriptive statistical analyses and sums up results from econometric regressions. Finally, section 4.6 concludes and recommends some steps for future research.

4.2 Conceptual Framework

Camerer and Malmendier (2007) define individual overoptimism as the overestimation of general prospects. The opposite case will be referred to as *overpessimism* throughout this study. Hence, a forecast is considered as overoptimistic, if the (ex-ante) forecasted value exceeds the (ex-post) actual value, and as *overpessimistic* in the opposite case.

Statistically, overoptimism (respectively *-pessimism*) of company i in year t is measured on a percentage base as the difference between the forecasted value F_{it} of X (for example sales, earnings or costs) and the actual value A_{it} , deflated by the forecasted value F_{it} , and multiplied by 100 (McDonald, 1973; Imhoff, 1978; Pedwell et al., 1994). Thus, I define the percental forecast error $PFE_{it}(X)$ as:

$$PFE_{it}(X) = \frac{F_{it}(X) - A_{it}(X)}{|F_{it}(X)|} \cdot 100 \quad (4.1)$$

Collective (or general) overoptimism bias is stated, if a sample's average PFE significantly exceeds zero and if a majority of observations reveals a positive error. Hence, significantly positive mean and median PFE values indicate an overoptimism bias. This holds, if desirable values like sales, earnings or gains from a merger are forecasted. If costs are forecasted, *underestimations* must be considered as overoptimistic.

The collective forecast quality, on the other hand, is measured by the standard deviation of the forecast errors throughout the sample. A higher variation thus represents greater uncertainty or worse forecast techniques.

4.3 Related Literature

Researchers from several disciplines have published analyses about forecast accuracy, providing overwhelming evidence of structural overoptimism: In the field of planning management, Flyvbjerg et al. (2002) provide a seminal evaluation of the accuracy of cost forecasts for public infrastructure projects. Using international data of 258 projects from almost the last 100 years, they state considerable cost overruns for almost nine out of ten projects and an average forecast error of 28 percent.

In Industrial Organization, several studies aimed at evaluating whether mergers and acquisitions on average managed to achieve its initially forecasted financial goals. Straub (2007) provides an exhaustive overview about the relevant studies which on average report a failure-rate of almost two thirds (with no study stating a failure-rate smaller than 40 percent).

Similar overestimations could also be found for initial public offerings (IPOs): Firth and Smith (1992) report 56 percent of 89 earnings forecasts published in IPO-broschures from New Zealand to be too positive, while the rate is 76 percent for 112 Canadian IPO-forecasts in Pedwell et al. (1994). A general overoptimism bias is also found in 168 Australian IPO-broschures (covering dividend as well as earnings forecasts), as Brown et al. (2000) reveal.

To go public with your firm, to start infrastructure projects or to plan mergers are of course extraordinary, non-routine situations. But empirical evidence, the vast majority of it coming from the field of accounting, found considerable amounts of overoptimism also for everyday forecasts, namely for managers' forecasts of their firms' next-years' earnings or sales. These forecasts are mostly published voluntarily as a component of the firms' annual reports, but are also mandatory in some countries (Japan or New Zealand, for example). To the best of my knowledge, all studies about forecast accuracy have analyzed such kind of published disclosures of joint-stock companies and found a general overoptimism bias, although differing broadly in scale, covered countries and time periods.

McDonald (1973), analyzing 201 American one-year-ahead earnings forecasts from the late 1960s, found 64 percent of them to be overoptimistic. Imhoff (1978) repeated McDonald's study with data from four further years and found similar results. Cho et al. (2011), analyzing the accuracy of management earnings forecasts of almost 2,700 Japanese firms between 1988 and 2005, find 53 percent to be too optimistic.

Kato et al. (2009) are the only ones who also analyzed sales instead of earnings forecasts and found only 39 percent of about 30,000 examined forecasts from Japanese firms between 1997 and 2007 to be too pessimistic. Since Japanese companies do not only publish one-year-ahead forecasts, but update them several times throughout the year, the authors can show that managers tend to adjust their initial forecasts towards less overoptimistic predictions over time. Thus, the established amount of overoptimism in six-month-ahead-projections is much smaller than in the case of one-year-ahead-forecasts.

While evidence of overoptimism is overwhelming, explanations for this bias are still fragmental, especially with respect to the question to which extent cognitive biases or strategic deception are to blame in the first instance. Flyvbjerg et al. (2002) conclude that the huge average cost overruns found for infrastructure projects are most likely due to strategic lying by policymakers, rather than to cognitive reasons. They argue that otherwise cost overruns should decrease over

time, since planners have a growing archive of similar projects to learn from. However, this point is disapproved by Kahneman and Lovallo (2003b): They argue that the lack of learning is a component of the cognitive bias and cite evidence about startups' failure rates which cannot be traced back to strategic deception and which did not decrease over time.

Rogers and Stocken (2005) are able to find that managers' overoptimism depends to some extent on the risk that cheating will be detected, and its costs. They show that overoptimism within US-firms' earnings forecasts between 1995 and 2000 is greater, if they run a low risk of being convicted (measured by the consistency of analysts' forecasts of the same value) or to be punished for consciously overestimating their firms' prospects. The latter is operationalized by an index measuring to what extent the firm is active in a high- or low-litigation industry.

Cho et al. (2011) show that a large fraction of the established overoptimism within their sample of Japanese firms can be traced back to the fact that firms avoid to forecast losses: Only less than 20 percent of those firms who later had to reveal a loss in their balance sheets had also forecasted negative earnings in the year before. The authors suggest that the Japanese bank-oriented corporate system might be one reason for this: Japanese managers reporting losses are in danger of being replaced due to pressure of their house bank, so they have strong incentives not to forecast losses which may prolong their tenure.

In order to provide additional evidence about the likely foundations of overoptimism biases, the present study is based on the methods of accounting research about management forecast accuracy, but analyzes undisclosed instead of published forecasts. I hypothesize that such forecasts are on average less overoptimistic than published forecasts due to lacking external pressures and incentives to cheat. In a second step, regression methods will be used to assess which firm characteristics enhance managers' optimism.

4.4 Data

The dataset used in this study is a subsample of the IAB Establishment Panel, a large anonymized German firm panel. Data access is restricted to researchers and not open for commercial market researchers. Between 1993 and 1997, the IAB asked the firms to provide a secret forecast of their sales numbers for the upcoming financial year. The IAB always sends its questionnaires in June, so the forecasted period starts 6 and ends 18 months afterwards (Fischer et al., 2009). Although the data does not reveal who actually filled in the questionnaire, surveys for similar datasets showed that in the large majority of firms a member of the upper management takes over this task.³

Table 4.1: Summary Statistics I

Category	Variable	Description	Unit	Mean	Median	StD	Obs.	Mean (JSC)	Obs.
Size	Sales	Total sales (prices of 1995)	Mill. Euro	774.8745	28.4381	5066.7089	2048	1216.8643	142
	Workforce	Number of employees	#	614.8232	129	1788.4166	6234	1769.8078	281
Situation	Sales Growth	Infl.-adj growth-rate of sales	%	6.9449	0.6543	47.4394	5889	1.4945	252
Key figures	Vacant	Vacant positions (in % of workforce)	%	1.5825	0	6.8043	6217	1.1722	281
	Hired	Newly hired empl. (in % of workforce)	%	5.3351	1.6878	10.0585	6181	2.9135	279
	Fired	Newly fired empl. (in % of workforce)	%	5.7566	2.7027	10.2008	6194	5.352	278
Further characteristics	Women	Female employees (in % of workforce)	%	41.9446	39.4582	28.7807	6191	32.3692	277
	Investment Quota	Investments (devided by sales)	#	0.0922	0.0278	0.2583	6068	0.0692	264
	Workers Council?	Firm has workers council?	Dummy	0.5808	1	0.4935	3263	0.9431	281
	Exporting?	Firm exports?	Dummy	0.2685	0	0.4432	6234	0.4413	281

For my analysis, I use all available sales forecasts, except for some outliers which were excluded as explained below. The sample thus contains 6,234 verifiable forecasts of 2,511 different firms. For each year between 1993 and 1997, between 827 and 1,936 firm-year observations are available.

The firms differ widely (in terms of workforce numbers, legal form, sector, sales numbers et cetera), since the IAB aims at providing a representative subpopulation of German companies. As shown in Table 4.1, the average firm within the

³See for example Abberger et al. (2009) for a survey among the firms within the Ifo-Institute's Business Climate panel.

Table 4.2: Summary Statistics II

Legal Form	%	Sector	%	Origin	%
		Farming	1.99	Schlesw.-H.	2.39
Partnership	25.70	Energy	1.81	Hamburg	1.68
Limited	46.79	Industry	22.62	Niedersachsen	6.79
Joint-Stock Comp.	6.43	Manufact.	6.59	Bremen	2.46
Non-Profit	14.55	Construct.	8.23	Nordrhein-W.	12.90
State-Owned	6.53	Trade	12.79	Hessen	5.84
		Logistics	3.69	Rheinland-P.	3.17
		Finance	3.53	Baden-Würt.	8.40
		Services	27.73	Bayern	8.51
		Organiz.	3.14	Saarland	1.74
		Administr.	7.88	Berlin	6.32
				Brandenb.	7.66
				Mecklenb.-V.	7.33
				Sachsen	7.84
				Sachsen-Anh.	7.94
				Thüringen	7.73

panel had 614.82 (median firm: 129) employees, of whom 41.94 (39.46) percent were women; it newly hired 5.34 (1.69) and fired 5.76 (2.70) percent of its total workforce every year, paid an average yearly wage of 24,105 (23,685) Euros⁴ and registered sales of 774.87 million (28.44 million) Euros. The firms' inflation-adjusted sales grew by 6.94 (.65) percent every year. 26.85 percent of the firms exported part of its products or services and 58.08 percent have a workers' council. The firms' yearly investments amount to 9.22 (2.78) percent (compared to sales numbers). Table 4.2 presents data about the frequency of legal forms, sectors and origins expressed by the German state (Bundesland) in which the firms' headquarters are located.

Since the IAB only allows remote data access, outliers or improperly filled in data arrays could not be detected manually. Thus, correction rules which are able to exclude useless data had to be applied. I chose to conduct the following corrections of the data: I dropped observations when a firm's workforce or sales exploded or collapsed by more than ten times in one year's time in order

⁴Please note: All pecuniary values within this study are given inflation-adjusted and display prices of the year 2000. They are further converted into Euro using the official exchange rate: 1.95583 Deutsche Mark = 1 Euro.

to account for outliers and firms which did not complete the survey properly. Furthermore, I excluded observations, if the firm over- or underestimated its next year's sales by more than ten times. I did so because I assume that such escalations or misestimations are rather due to sudden existential changes (like insolvencies, mergers or acquisitions) than to actual forecast errors. Firms were completely left out in the analysis, when they gave exactly the same answers, either for their sales or for their workforce numbers⁵, three or more consecutive times, since I assume this to be a clear sign that the managers of these firms did not put enough effort into these questionnaires.⁶

4.5 Hypotheses and Empirical Analysis

Since I analyze secret company-internal instead of published management sales forecasts, firms cannot benefit from overglossed forecasts and, hence, have no incentives to cheat in order to influence analysts or investors. To test the suggestion that such pressures are the crucial foundation of the overoptimism established in previous literature, I formulate the following hypothesis:

- **H1: The mean and median forecast errors in this sample depict less overoptimism than it has been found in previous literature.**

Kahneman and Lovallo (2003a) as well as Flyvbjerg et al. (2002) state that the extent of internal and external organizational pressures to produce and report good news are an important source of overoptimism (although they differ in their assessment of its relative importance). Rogers and Stocken (2005) further provide evidence that overoptimism does not occur randomly, but varies with the companies' incentives to gloss over their estimations.

⁵The latter correction was only conducted for firms with more than 1,000 employees, as it is not unlikely that small firms have constant workforce numbers over time.

⁶To check for robustness, all analyses within this chapter were repeated with an otherwise corrected sample: Therefore, observations were only excluded, when the firm's sales and workforce numbers changed by more than a hundredfold or when the forecast misestimated future sales by more than this. The results do not differ qualitatively, so they are not reported here, but can be provided upon request.

Table 4.3: Forecast Errors

	Mean	StD	P25	Median	P75	Obs.
PFE (whole sample)	-5.5770***	42.2345	-11.6071	-0.3078	9.7007	6234
PFE (joint-stock comp).	-6.1306	39.5053	-12.7918	-1.0379	8.6666	2511
Mean PFE (whole sample)	-1.5726***	29.5955	-6.8682	1.9608	9.3364	281
APFE (whole sample)	20.5031	37.3418	4.6089	10.8681	23.6782	6234
Year	1993	1994	1995	1996	1997	
Mean PFE	-5.6446***	-6.1334***	-0.7376	-4.6641***	-8.7923***	
Obs.	911	907	827	1963	1626	

*, p < 0.1; **, p < 0.05; ***, p < 0.01; P-values refer to t-tests testing whether the mean is equal to zero.

Turning the focus on the firms' legal forms, I expect that joint-stock companies face more (at least external) pressure than for example self-employers. Accordingly, the fate of joint-stock companies depends much stronger on public judgments of its soundness, which are based on reports of past and predicted future financial results. Furthermore, its decision makers have to satisfy a greater number of owners (their shareholders) who are commonly more interested in short-term success than for example self-employers or strategic investors. Although the forecasts whose accuracy are analyzed in this chapter were retrieved under secrecy, it seems unreasonable to expect that forecaster are fully able to abstract from these organizational pressures while reporting their predictions to the IAB. Hence, I set up the following hypothesis:

- **H2: Management sales forecasts of joint-stock companies reveal more overoptimism than those of firms with other legal forms.**

To check the data for significant forecast biases, I focus on the distribution of the forecast errors. A median and mean value of zero would indicate that no significant statistical pattern existed, since it could not be rejected that all individual misforecastings were due to a random process.

Table 4.3 contains summary statistics of the firms' forecast errors (PFEs). The results clearly reveal that the average forecast is rather *overpessimistic* than *overoptimistic*. The mean forecast is 5.58 percent below the true value and is highly significantly different from this, as a t-test reveals (the p-value is smaller than 0.001). The median error for the whole sample is closer to zero than the

mean value, but slightly negative (-0.31) as well, showing that also a majority of sales forecasts is *overpessimistic*.

The results further reveal that the yearly mean values for PFEs are always negative, too, though differing in scale. They range from -8.79 in 1997 to -0.74 in 1995 and are always highly significantly smaller than zero (except for 1995). The yearly PFEs do not appear to follow a time trend, so there is no sign of common learning. However, the data covers a time frame of only five years and is thus of limited value for such analyses.

Since some individual firms occur in the sample up to five times, while others provided a forecast only once, I also report the distribution of the firms' individual mean forecast errors in Table 4.3. The results are qualitatively similar to those above which shows that the overpessimism bias is not due to the more or less frequent occurrence of some individual firms.

To sum up, these results clearly contradict previous evidence and allow support for hypothesis H1. The fact that the mean and median firm is *overpessimistic* is quite interesting considering that the phenomenon *overpessimism* has not gained much attention in economics thus far.

Table 4.3 further displays descriptive statistics for a subsample that is restricted to joint-stock companies. The mean PFE value is -1.57 percent, which also depicts an average, though smaller, bias towards overpessimism (but not statistically significant). However, the median forecast error for this subsample is 1.96, which shows that a majority of forecasts from these firms was overoptimistic. Both values allow some support for hypothesis H2 which will be further tested with different regression methods below.

Descriptive results also reveal that about 45 percent of the forecasts (in the whole sample) assumed declining sales numbers. Hence, managers are not trying to avoid to forecast negative values like it was found in Cho et al. (2011) for Japanese firms. I suggest that this is also due to the fact that the forecasts are kept secret by the IAB. However, it must be admitted that Cho et al. analyzed

earnings instead of sales forecasts, so the results are not perfectly comparable, since losses are commonly assumed to be more painful for firms than declining sales numbers.

The overall quality of forecasts within the sample is rather low: The 25-percent quartile of the PFE-distribution amounts to -11.61 percent, while the 75-percent quartile is 9.70, indicating that about 50 percent of the forecasts miss the mark by more than ten percent. The distribution's standard deviation (42.23) is also quite high, whereas broadly in line with previous literature about errors in earnings forecasts (McDonald, 1973; Imhoff, 1978; Pedwell et al. 1994).

To further assess the forecast quality, I compute the absolute percental forecast errors (APFE). APFE quantifies the forecast errors, disregarding the sign of the error, and is defined as:

$$APFE_{it}(X) = |PFE_{it}(X)| \quad (4.2)$$

The mean value of APFE amounts to 20.50 percent, as Table 4.3 shows. Since the firms' sales numbers on average change by 21.90 percent from year to year, the mean absolute forecast error is thus only slightly smaller. Hence, it can be stated that the managers are only able to correctly forecast a small fraction of their firms' development.

Using econometric regression methods, this section further aims to assess which firm characteristics drive overoptimism or -pessimism in general. Therefore, three additional hypotheses are set up in the following.

Camerer and Malmendier (2007) suggest that "misattribution of cause" may be one source which makes forecasters overly optimistic. Managers whose firms are currently succesful may wrongfully overestimate the proportion of this success which is based on their own skills and decisions, and underestimate the influence of luck and external factors like the general economic situation of their sectors. Hence they may underrate the possibility that the situation worsens due to external factors which they can hardly influence.

So I expect:

- **H3: Overoptimism is positively related to current success (measured by a firm's percental sales-growth).**

Previous literature reports extremely high failure rates for business startups (Camerer and Lovallo, 1999) and that entrepreneurs seem to be especially optimistic in character (Arabsheibani et al., 2000). Hence, I assume that overoptimism might be a bias that occurs especially frequent among younger firms. I thus expect:

- **H4: Younger firms reveal a higher extent of overoptimism.**

Previous behavioral research showed that women question their own skills more often than men and thus reveal less overconfidence, for example regarding their stock trading activity (Barber and Odean, 2001), and less overoptimism, for example with respect to their expectations of their future financial situation (Arabsheibani et al., 2000). Although the IAB Establishment Panel does not contain details about the forecasters' gender, data about the fraction of women among the workforce is available and reveals strong variations of this value. Thus, the fraction of female employees is used as a proxy for gender differences among the firms. I expect:

- **H5: Firms with a higher percental fraction of women among the workforce are less prone to overoptimism.**

The estimation strategy will be as follows: First, I use probit and logit models with clustered standard errors which estimate the probability that a forecast is too optimistic, all other things equal. Thus, the dependent variable is dichotomic and will be equal to 1, if the PFE is larger than zero, and 0, if the PFE is negative.⁷ A positive coefficient hence denotes a positive influence of a variable

⁷Here, 14 true forecasts (with a PFE of exactly zero) are left out. As robustness checks, the regressions were repeated with samples where true forecasts were either included in

on the probability that a firm issues an overoptimistic forecast, while a negative one indicates the opposite.

The logit and probit estimations are then compared to a standard (robustly estimated, also with clustered standard errors) pooled OLS regression which uses PFE values as dependent variable. Since some firms occur in the dataset more often than others, a between-effects panel regression is conducted afterwards as a robustness check: Here, the individual firms' mean values are used to account for the possibility that the firms' different weights within the sample distort the results. Regarding these two models, the coefficients depict the estimated increase or decrease of the forecast errors in percentage points which follow an increase of the respective independent variable. A positive (negative) coefficient would hence show that a forecast error is estimated to be more (less) overoptimistic, while no statement can, of course, be made whether it is actually overoptimistic or -pessimistic.

The independent variables used in these specification are defined and summed up in table 4.1 as well. Regarding the hypotheses above, they include the firms' inflation-adjusted percental sales-growth rate ($Sales_Growth_{it}$), the firms' legal forms ($Form$, given as five different categories⁸) and a categorial variable of age (Age_{it})⁹.

Furthermore, control variables regarding the firms' size are used (the number of employees, $Workforce_{it}$, as well as the inflation-adjusted¹⁰ sales numbers, $Sales_{it}$), as well as dummy variables for the German state where the firm's headquarter is located, for the firm's sector and for the year of the observation. To

the overoptimism or -pessimism category. However, they are not reported here, since the results are qualitatively equal, but can be provided upon request.

⁸Joint-stock company, partnership, state-owned, limited (used as base level throughout the regressions) and others.

⁹Due to a lack of more precise information, only categorial data exists: Age_{it} equals 1, if the firm was founded before 1960; equals 2, if the founding date was after 1959 and before 1990; and equals 3, if the venture was launched after 1989. Throughout the regressions, dummy-variables are used for the two latter categories, while the first category is the base level.

¹⁰Given in prices of the year 2000.

control for macroeconomic influences, the real growth rate of the firm's industry¹¹ is included, too.

As a robustness check, the regressions are repeated using an additional range of control variables: The percental fraction of women among the workforce ($Women_{it}$), the persistence of the workforce (measured by the percentage of the workforce that has been newly hired, $Hired_{it}$, or fired, $Fired_{it}$, during the respective year, as well as the percentage of current vacancies, $Vacant_{it}$), the firm's investment ratio ($Invest_{it}$, measured as the sum of investments divided by the sales numbers), its wage costs ($Wagecost_{it}$, also as a fraction of sales) and its inflation-adjusted average wage ($AvWage_{it}$). Furthermore, additional dummy variables are included: $Export$, which equals 1, if the firm exports parts of its products, and $Council$, which is 1, if the firm allows its employees to let their interests be represented by a workers council.

Table 4.4 depicts the regression results. Columns (1) and (5) show the probit regression results and columns (2) and (6) the logit regression. Both models allow support for hypothesis H2 (highly significant at the one percent level). Compared to limiteds, the probability of issuing an overoptimistic forecast is much higher for joint-stock companies, while the probabilities of firms with other legal forms do not differ significantly from that of limiteds.

Furthermore, also hypothesis H3 cannot be rejected: The greater a firm's current growth-rate, the higher is the estimated probability for an overoptimistic forecast (significant at least at the five percent level). However, no support can be stated for hypothesis H4 about the influence of the firms' age.

Moreover, columns (5) and (6) show significantly negative influences on the probability of issuing an overoptimistic forecast for the investment ratio (significant at the one percent level) and the wage-sales ratio (five percent level). Hence, companies which are active in people-intensive businesses or which make

¹¹Due to changes of the sector-classification, the firms can only be classified into eleven industrial sectors. The growth rates are collected from the German central bank (Bundesbank) and the German Federal Bureau of Statistics.

relatively high investments seem to be more cautious when predicting their own future.

The logit and probit analyses further provide no support for H5, while the the OLS and between-effects models in columns (7) and (8) do. Here, the fraction of women among the workforce has a clearly significant negative influence on the forecast error (five percent level), as suggested by H5. All other things equal, the forecast error is estimated to be about 2.25 percentage points lower, if the fraction of women grows by the value of one standard deviation.

Table 4.4: Regression Results

	Column Method	(1) LOGIT	(2) PROBIT	(3) OLS	(4) BE	(5) LOGIT	(6) PROBIT	(7) OLS	(8) BE
Category	Variable	Dummy	Dummy	PFE	APFE	Dummy	Dummy	PFE	APFE
Size	<i>Workforce_{it}</i>	0.0000 (0.6445)	0.0000 (0.6318)	0.0004 (0.1201)	0.0003 (0.5353)	0.0000 (0.8347)	0.0000 (0.8242)	0.0004 (0.1277)	0.0004 (0.4524)
	<i>Sales_{it}</i>	0.0000 (0.5286)	0.0000 (0.5062)	0.0000 (0.4209)	0.0000 (0.8929)	0.0000 (0.8287)	0.0000 (0.8368)	-0.0001 (0.2823)	-0.0001 (0.7709)
Situation	<i>Sales_{it}</i>	0.0024** (0.0103)	0.0014*** (0.0064)	0.0740*** (0.0004)	0.0866*** (0.0000)	0.0032*** (0.0069)	0.0019*** (0.0037)	0.0614*** (0.0028)	0.0684*** (0.0000)
	<i>Growth_{it}</i>	-0.0215 (0.4308)	-0.0132 (0.4335)	-0.5667 (0.2705)	-0.0520 (0.9335)	-0.0316 (0.2704)	-0.0196 (0.2714)	-0.6411 (0.2176)	-0.4307 (0.4146)
Age	<1960	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)
	>1959 & <1991	0.0433 (0.8810)	0.0266 (0.8821)	-1.7463 (0.7764)	-0.9703 (0.8686)	-0.0512 (0.8688)	-0.0266 (0.8886)	-5.4791 (0.3816)	-4.9241 (0.3140)
	>1990	0.0862 (0.5364)	0.0547 (0.5277)	1.1883 (0.5824)	1.9735 (0.5301)	0.0814 (0.5853)	0.0526 (0.5699)	-0.2210 (0.9152)	0.3627 (0.8918)
	Partnership	0.0395 (0.6998)	0.0249 (0.6972)	3.6660* (0.0671)	3.3982 (0.1082)	0.0207 (0.8622)	0.0126 (0.8650)	2.9700 (0.1763)	2.8068 (0.1560)
Form	Limited	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)	(base) (base)
	Joint-Stock	0.5223*** (0.0014)	0.3231*** (0.0013)	3.7955* (0.0743)	3.5131 (0.2834)	0.5124*** (0.0046)	0.3150*** (0.0046)	1.2787 (0.5649)	0.5570 (0.467)
	State-Owned	-0.0501 (0.7472)	-0.0334 (0.7287)	2.2310 (0.4684)	2.4432 (0.4516)	-0.0110 (0.9520)	-0.0092 (0.9352)	3.8404 (0.1268)	3.3374 (0.2720)
	Other	0.1159 (0.5319)	0.0721 (0.5345)	-1.1748 (0.7776)	-0.1172 (0.9760)	0.2258 (0.2639)	0.1413 (0.2615)	2.7369 (0.3998)	2.9047 (0.3885)
Further Key Figures	<i>Women_{it}</i>					-0.0016 (0.3942)	-0.0010 (0.3836)	-0.0776** (0.0408)	-0.0815*** (0.0083)
	<i>Invest_{it}</i>					-0.7052*** (0.0075)	-0.4157*** (0.0046)	-5.0189* (0.0996)	-5.1176** (0.0296)
	<i>Vacant_{it}</i>					0.0019 (0.7273)	0.0012 (0.7214)	0.0918 (0.2454)	0.0733 (0.4054)
	<i>Hired_{it}</i>					-0.0035 (0.4339)	-0.0021 (0.4458)	-0.0744 (0.4018)	-0.0560 (0.4270)
	<i>Fired_{it}</i>					0.0036 (0.4281)	0.0022 (0.4269)	0.0951 (0.3181)	0.0675 (0.3387)
	<i>AvWage_{it}</i>					-0.0038 (0.4650)	-0.0023 (0.4716)	0.0738 (0.3823)	0.0673 (0.4361)
	<i>Wagecost_{it}</i>					-0.0061** (0.0240)	-0.0038** (0.0214)	-0.2317*** (0.0002)	-0.2379*** (0.0000)
	<i>Council_{it}</i>					0.0513 (0.6503)	0.0297 (0.6721)	2.1668 (0.2088)	2.4912 (0.1758)
Control-Dummies	<i>Export_{it}</i>					0.0788 (0.5574)	0.0511 (0.5403)	2.0302 (0.3280)	1.3575 (0.5342)
	Years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Further Control-Dummies	Sector	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	State	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Constant	0.0938 (0.7778)	0.0584 (0.7784)	-9.5173 (0.2199)	-10.1786 (0.1411)	0.3111 (0.4275)	0.1917 (0.4334)	1.6722 (0.8224)	1.5562 (0.8100)
	Obs.	2975	2975	2989	2989	2592	2592	2606	2606
	R2			0.0246	0.0272			0.0505	0.0535
	Pseudo R2	0.0238	0.0237			0.0325	0.0323		

p-values in brackets; *: p < 0.1; **: p < 0.05; ***: p < 0.01

Regarding hypothesis H3, the OLS- and between-effects-models (3), (4), (7) and (8) further confirm the support of the probit and logit regressions. All other things equal, firms whose growth rate is larger by one standard deviation provide forecasts which are about three percentage points more overoptimistic. Furthermore, also the investment ratio has an, at least weakly significant, influence in the OLS and between-effects models (columns (7) and (8)). The significantly negative impact of the wage ratio on a firm's forecast error can be confirmed as well (one percent level).

Regarding the firms' legal forms, the results of the OLS and between-effects models reveal only mixed evidence. While model (3) reports weakly significantly higher amounts of overoptimism for joint-stock companies and partnerships (compared to limiteds), models (5), (7) and (8) do not find any significant differences. However, this might be due to joint-stock companies' better forecast quality, since their errors show much less variation (as can be seen in the second line of table 4.3).

To sum up: The regression results provide clear and robust support for H3: Currently successful firms appear to be more vulnerable for issuing too optimistic forecasts about their future sales numbers. However, hypothesis H4 had to be clearly rejected, as no significant influence was found for the age of the firms.

At least partly support was found for hypotheses H2 and H5: While the probability of issuing an overoptimistic forecast is estimated to be significantly higher for joint-stock companies in the probit and logit models and in the OLS-model (3), this results do not hold for the other OLS model and the between-effects regressions.

For hypothesis H5, the opposite situation occurs: Models (7) and (8) provide evidence that the forecast error declines significantly with the fraction of women among the workforce, while no significant influence could be found in the logit (5) and probit (6) models. Furthermore, all models reveal a robustly significant negative influence for the wage ratio and the investment ratio.

4.6 Conclusion

This study is, to my best knowledge, the first one about management forecast errors that states a significant statistical pattern towards *overpessimism*. My first suggestion is that the non-existence of a general overoptimism bias in this sample is due to the lack of external pressure to report good news. Unlike forecasts issued in public disclosures, the forecasts analyzed here cannot be used as strategic signals to investors and analysts. Hence, firms have no benefit from intentionally overglossing their stated prospects. This allows some support for Flyvbjerg's (2003) hypothesis that strategic deception is one, if not the main, cause of overoptimism. If, however, cognitive biases were the main reasons for overoptimism, it could be expected that the forecasts analyzed here were too optimistic on average, too, since it would not matter for which purpose they were made.

However, the reasoning above is not able to sufficiently explain why the results actually show a tendency towards *overpessimism*, as it can only explain why the mean PFE is less positive, but not why it is actually negative. Thus, my second suggestion is that overpessimism may be a sign of loss aversion. I argue that decision makers may dislike negative surprises more than they like positive ones and thus hedge against rude surprises by being especially cautious, when estimating their firms' future prospects.

Yet, further research is indicated to solidify the findings and conclusions of this study. It would be most preferable to directly compare undisclosed and disclosed forecasts of the same firms, but this appears to be impossible due to a likely lack of data.¹² However, the analysis could be repeated on the one hand with company-internal forecasts from other countries and on the other hand with public disclosures of German firms. This approach would come closer to a *ceteris-paribus* analysis.

¹²Firm panels are typically highly anonymized for data protection reasons, so matching panel datasets with data from public disclosures appears to be impossible.

Using different regression methods, this study could further provide some support for different hypotheses related to behavioral research. It could be shown that currently successful firms display a tendency towards overoptimism, which I argue might be due to the cognitive bias “misattribution of cause”. If management skills and decisions are held accountable for current success too much, firms may underrate the influence of randomly occurring external factors and thus the possibility of a worsening situation.

At least some support could further be found for the hypotheses that women are more pessimistic when forecasting the future, which can be seen as in line with previous behavioral evidence, and that joint-stock companies are more prone to be overoptimistic than firms with other legal forms. I suggest that managers of joint-stock companies face stronger pressures to deliver positive numbers, since their shareholders first and foremost demand persistent returns and dividends. Thus, these forecasters may not be able to fully abstract from these omnipresent pressures, when forecasting their firms’ future.

4.7 References

- Abberger, K., Birnbrich, M. and Seiler, C. (2009), Der “Test des Tests” im Handel - eine Metaumfrage zum Ifo Konjunkturtest, Ifo Schnelldienst, 62(21), 34–41.
- Arabsheibani, G., de Meza, D., Maloney, J. and Peasron, B. (2000), And a Vision Appeared Unto Them of a Great Profit: Evidence of Self-Deception Among the Self-Employed, Economics Letters, 67(2000), 35–41.
- Barber, B. and Odean, T. (2001), Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment, The Quarterly Journal of Economics, 116(1), 261–292.
- Brown, P., Clarke, A., How, J. and Lim, K. (2000), The Accuracy of Management Dividend Forecasts in Australia, Pacific-Basin Finance Journal, 8, 309–331.
- Camerer, C. and Lovallo, D. (1999), Overconfidence and Excess Entry: An Experimental Approach, American Economic Review, 89, 306–318.
- Camerer, C. and Malmendier, U. (2007), Behavioral Economics of Organizations, In: Diamond, P. and Vartiainen, H. (Eds.), Behavioral Economics and Its Applications, Princeton University Press, 235–290.
- Cho, M., Hah, Y. and Kim, O. (2011), Optimistic Bias in Management Forecasts by Japanese Firms to Avoid Forecasting Losses, The International Journal of Accounting, 46(1), 79–101.
- Firth, M. and Smith, A. (1992), The Accuracy of Profits Forecasts in Initial Public Offering Prospectuses, Accounting and Business Research, 22(87), 239–247.
- Fischer, G., Janik, F., Müller, D. and Schmucker, A. (2009), The IAB Establishment Panel - Things Users Should Know, Schmollers Jahrbuch, 129, 133–148.

- Flyvbjerg, B., Holm, M. and Buhl, S. (2002), Underestimating Costs in Public Works Projects - Error or Lie?, *Journal of the American Planning Association*, 68(3), 279–295.
- Flyvbjerg, B. (2003), Delusions of Success, *Harvard Business Review*, 81(12), 121–122.
- Imhoff, E. (1978), The Representativeness of Management Earnings Forecasts, *The Accounting Review*, 53(4), 836–850.
- Kahneman, D. and Lovallo, D. (2003a), Delusions of Success: How Optimism Undermines Executives’ Decisions, *Harvard Business Review*, 81(7), 56–63.
- Kahneman, D. and Lovallo, D. (2003b), Delusions of Success, *Harvard Business Review*, 81(12), 122.
- Kasznik, R. (1999), On the Association between Voluntary Disclosure and Earnings Management, *Journal of Accounting Research*, 37(1), 57–81.
- Kato, K., Skinner, D. and Michio, K. (2009), Management Forecasts in Japan: An Empirical Study of Forecasts that are Effectively Mandated, *Accounting Review*, 84(5), 1575–1607.
- Matsumoto, D. (2002), Management’s Incentives to Avoid Negative Earnings Surprises, *The Accounting Review*, 77(3), 483–514.
- McDonald, C. (1973), An Empirical Examination of the Reliability of Published Predictions of Future Earnings, *The Accounting Review*, 48 (3), 502–510.
- Ota, K. (2010), The Value Relevance of Management Forecasts and Their Impact on Analysts’ Forecasts: Empirical Evidence From Japan, *Abacus*, 46(1), 28–59.
- Pedwell, K., Warsame, W. and Neu, D. (1994), The Accuracy of Canadian and New Zealand Earnings Forecasts: A Comparison of Voluntary Versus Compulsory Disclosures, *Journal of International Accounting, Auditing & Taxation*, 3(2), 221–236.

- Rogers, J. and Stocken, P. (2005), Credibility of Management Forecasts, *The Accounting Review*, 80(4), 1233–1260.
- Straub, T. (2007), *Reasons for Frequent Failure in Mergers and Acquisitions: A Comprehensive Analysis*, Gabler Edition Wissenschaft - Business and Economics.

Chapter 5

Counter- or Pro-Cyclical Public Spending? The Impact of Political Competition and Government Partisanship

Abstract

For several decades now, economists and politicians have been arguing whether and to which extent governments should actively use economic policy to counterbalance the fluctuations of the economic cycle. From a scientific point of view, also the question arises which economic or political factors make governments choose a more or a less counter-cyclical spending policy. Referring to the political business cycle literature, two hypotheses are of interest. First, that the different parties of the political spectrum have different preferences regarding counter-cyclical fiscal policy (the partisan hypothesis). Second, that governments, by signalling that they care for those suffering from economic hardships, tend to counterbalance economic fluctuations more strongly when facing political competition in election years (the opportunism hypothesis). Seitz (2000), using data of the German federal states, had to reject the partisan hypothesis. The present chapter extends this approach by testing both hypotheses against each other, using similar, though more recent data of the German states. While the partisan hypothesis has to be rejected again, some evidence can be provided that the state governments' spending policies are significantly more counter-cyclical when they are campaigning for reelection.

Key Words: *Political Business Cycles, Activist Fiscal Policy*

JEL-Classification: *D72, E61, E62, H61*

5.1 Introduction

Throughout recent decades, there has been an extensive debate among economists about the key determinants of governments' economic policies. Two concurrent schools have evolved. On the one hand, the partisan school, which assumes that a government chooses a certain policy first and foremost according to its fundamental beliefs, disregarding whether this is currently popular among the the majority of the voters (Hibbs, 1977). It is thus predicted that the policies chosen by leftist governments differ significantly from those chosen by right-wing governments.

On the other hand, the opportunism school claims that political competition is the key driver. The approach is based on the assumption that a government first and foremost maximizes its own utility (Downs, 1957; Nordhaus, 1975) and does what is needed to please the swing voters and thus to assure its reelection. The hypothesis predicts electoral cycles, which means that governments will conduct popular policy strategies when facing political competition (thus, in election years) and unpopular policies afterwards.

While early theoretical or empirical studies mostly analyzed why governments choose different monetary policies (thus, different choices regarding the trade-off between low inflation and low unemployment characterized by the Philipps-curve), more recent works focused on fiscal policy instead. This shift was due to the objection that the monetary policy is nowadays defined by independent central banks in most western countries, and cannot be directly manipulated by the governments any more (Drazen, 2000).

Hence, most recent empirical works aiming to test the partisan hypothesis analyzed whether governments led by a left-wing party are prone to raise public spending or indebtedness to a stronger extent than right-wing governments. Thereby, it is assumed that leftists prefer "big government", thus a more active state and a larger public spending ratio. The opportunism hypothesis, in turn, searches for electoral cycles in fiscal policies, meaning that governments

increase public spending in election years in order to please the swing voters. In non-election year, spending cuts are expected.

However, the underlying assumptions appear to be too simplistic: Regarding the opportunism school, it does not seem to be reasonable to assume that voters can be repeatedly deluded before every election by simply raising fiscal activity.¹ Furthermore, voters are not likely to prefer one certain type of policy in every economic situation, but may rather have context-dependent preferences (Franzese, 2002; Cusack, 1999) which differ between booms and busts. Hence, their preferences may be sensitive to changes in economic conditions. Accordingly, one might hypothesize that, first, voters' preferences vary with the state of the economy and that, second, opportunistic governments may exploit this by counterbalancing economic fluctuations more actively when facing political competition in election years.

Regarding the partisan school, though it may be reasonable at first sight to assume that left-wing politicians prefer a more active role of the state, it appears to be unrealistic that this leads to steady spending increases when leftist governments are in power, while right-wing governments do the opposite.² Rather, ideological differences between the different political blocks should be more sophisticated.

In order to overcome these shortcomings, Seitz (2000) suggested to distinguish fiscal policy types differently. Instead of analyzing how strongly different governments raise public spending and indebtedness, he focused on the question to which extent fiscal policies are adjusted to the current overall economic situation. Hence, Seitz tested the partisan hypothesis by assuming that left-wing governments follow a more proactive, thus more "Keynesian-style" fiscal policy

¹Some theorists argue that this strategy is successful not because voters are myopic, but rather because they cannot assess the political contestants' credibility in advance (Rogoff, 1990). Accordingly, they may be rational, but uncertain about what follows after the elections and must hence rely on signals of competence. However, the question remains, why voters should prefer the same type of fiscal policy in every economic situation.

²Franzese (2002) furthermore states that there is no convincing evidence "for naive views of the left (right) as unconditional deficit (surplus) producers".

strategy than right-wing governments, meaning that spending (and indebtedness necessarily too) are raised strongly during economic downturns, while they are significantly cut during upswings. He thus assumes that parties first and foremost disagree about the question of how much governments are able (or should try) to counterbalance economic fluctuations.

The suggestion to focus on the cyclical policy appears to be very suitable, since it is one of the key strategic decisions a government has to take. Indeed, the question whether and to what extent governments should try to counterbalance cyclical fluctuations of their economies is one of the most recurring disputes among economists and politicians. As Auerbach et al. (2010) point out, no consensus has evolved yet, neither within academia, nor in public. Furthermore, empirical evidence about the success of counter-cyclical measures is mixed. Econometric approaches typically try to estimate a multiplier effect which measures how much additional growth can be created with one additional unit of public spending during downswings.

However, according to Auerbach et al. (2010), the range of measured coefficients is “embarrassingly large” and differs strongly with respect to covered time periods, political measures and analyzed countries. Thus, it is no surprise that the optimal cyclical policy has always remained a matter of public dispute.³ While supporters often argue that a pro-active spending policy is able to substitute missing private demand and thus to cure economic downswings, opponents believe that state interventions exacerbate the economic situation in the long run.

³In the United States, the necessity of the stimulus packages introduced by the Obama administration after the Lehman-crisis in 2007 was a main point of issue during the 2012 presidential election campaign. In Germany, parties argued heavily during 2009, whether another stimulus package should be adopted by the Merkel administration to counterbalance the severe economic downturn. Later, in 2011 and 2012, when the German economy experienced a powerful upswing, the debate was the other way round, as several politicians urged the government to save more revenues. On the state level, the same debate occurred for example during the electoral campaign in the state of Nordrhein-Westfalen in 2012, when the state government was criticized for its high budget deficits despite the powerful economic recovery.

The aim of the present chapter is to extend Seitz's (2000) approach using similar data. Seitz employed a dataset containing budget and economic data of eleven German states from the time between 1976 and 1996 and analyzed whether the state governments' partisanship significantly affects the cyclicity of fiscal policy. Therefore, he adopted an empirical approach by Bayoumi and Eichengreen (1995) who developed a measurement for the elasticity of public spending with respect to GDP-growth. The German states provide a well-fitting data basis, as they are constitutionally homogeneous, but relatively sovereign in their fiscal policy. Furthermore, their expenditures amount to 14.5 percent of the GDP (in 2010) which is more than the federal or the municipal level administrations have at their disposal. Thus, state governments should have the chance to affect economic conditions with their budget decisions.

While Seitz (2000) only tested the partisan hypothesis, I focus on both hypotheses in order to reveal which one has more predictive power. Therefore, I use a dataset of all German states for the time between 1991 and 2010. The opportunism hypothesis is tested by analyzing whether the timing of elections influences the governments' choices of fiscal policies. To be precise, it is tested whether governments conduct more "Keynesian-style" fiscal policies in election years. This approach seems appropriate, as governments may aim to signal that they care about the economic situation and about possible economic hardships during downswings when they face political competition.

The following results can be reported: Econometric regressions reveal that German state governments in general conduct weakly pro-cyclical spending policies, meaning that spending-growth is changed to a lesser extent than the GDP-growth, but in the same direction. Furthermore, supporting Seitz's results, I do not find clear hints that the cyclicity of spending policies differs between governing coalitions. The partisan hypothesis thus has to be rejected. The opportunism hypothesis, however, can be supported. In election years, governments seem to adjust their spending policy significantly more counter-cyclically to current economic fluctuations.

The remainder of this chapter is structured as follows: Section 5.2 provides a brief overview about the related literature. In the following, section 5.3 describes the dataset used throughout this chapter, while section 5.4 introduces the empirical approach and reports the regression results. Finally, section 5.5 concludes.

5.2 Related Literature

There is already a vast amount of empirical literature which tests whether partisanship or opportunism are the main factors that determine Western governments' economic policy choices. Since a complete overview would be beyond the scope of this chapter, I will only focus on the most important theoretical works and some selected empirical ones which are directly linked to the approach used here. For more exhaustive overviews, see for example Drazen (2000), Franzese (2002) or, more recently, Markwardt (2008)⁴.

This section is organized as follows: I will first summarize the evidence regarding the opportunism school and sketch its pathway. Then, the partisan school is described. Finally, I will outline the political and academic debate about the success of activist fiscal policy.

5.2.1 Opportunism School

Downs (1957) suggested that public economics researchers should no longer assume that governments try to maximize social welfare, but rather that they maximize their own utility. Thus, he claimed that they do everything to stay in office. Downs called for a positive rather than an idealistic modelling of government behavior, and laid the foundations for the opportunism school. The central suggestion was that "political parties in a democracy formulate policy strictly as a means of gaining votes" (Downs, 1957).

⁴In German language.

While Downs proposed that governments will first and foremost try to win elections, the question arose what this means with regard to economic policy choices. In his seminal paper, Nordhaus (1975) refers to the Philipps-curve and states that governments always face a trade-off between present and future welfare, respectively stable prices and low unemployment. Right before elections, governments would stick to present welfare and ease monetary policy in order to reduce unemployment. He provided some empirical evidence that unemployment rates indeed followed the predicted trend in nine western countries between 1947 and 1972.

However, many following works rejected the hypothesis, as is summed up in Alesina and Sachs (1988), while more recent analyses backed it again: Abrams and Iossifov (2006), for example, provided evidence that the FED's policy followed political monetary cycles, thus that it is changed before and after presidential elections. Vaubel and Dreher (2009) show that central banks in developing countries reduce their reserves of foreign exchanges right before elections in order to be able to boost the money supply while keeping exchange rates stable.

Many researchers, however, later criticized that the Nordhaus-model was based on monetary policy interventions, as this field is nowadays largely independent from governmental influence, at least in most industrialized countries (Drazen, 2000). Eventual theoretical works (see for example Rogoff, 1990) then rather focused on fiscal instead of monetary policy, thus on expansions or cuts of government spending, tax revenues or public borrowing. The underlying assumption is again that voters prefer short-term to long-term welfare, and hence reward higher transfers or a smaller tax burdens (Alesina, 1989).

Various works tested the budget data of different countries and time-periods for electoral cycles, revealing at most mixed evidence, at least for western democracies (Drazen, 2000). Alesina et al. (1993) found that deficits and expenditures in the OECD countries between 1960 and 1987 were higher close before elections. Alesina et al. (1998), reversing the analysis, however showed that voters do not seem to punish austerity policies. Jochimsen and Nuscheler (2011), using data

from the western German federal states, could even reveal that governments tend to reduce borrowing in pre-election years instead of boosting it.

Furthermore, focusing on studies of German state budgets reveals a similar mixed impression: Galli and Rossi (2002), using data from between 1974 and 1994, provided some evidence that governments raise the ratio of expenditures to GDP in election years which supports the opportunism approach. Schneider (2007 and 2010), while not finding a general effect of elections on public expenditures, at least revealed some opportunistic shifting within the composition of state expenditures. Social transfer payments, for example, are increased in election years, while other less visible items of expenditure are reduced. Berger and Woitek (1997), however, analyzing budget data from the German federal level, found no evidence for opportunism in fiscal policy.

5.2.2 Partisan School

In his seminal Paper, Hibbs (1977) argued that the political parties may have differently rank-ordered preferences regarding important political topics because they represent different constituencies and thus different social classes. Regarding the trade-off between low inflation and low unemployment as stated by the Philipps-curve, the governments' choices in terms of monetary policy should hence differ. Hibbs reasons that left-wing parties like the Democrats in the United States or the Labour Party in the United Kingdom should prefer low unemployment over low inflation, as they represent the working class, while it should be vice versa for the Republicans respectively the Tories who are backed by the rentiers.

Hibbs found support for his hypothesis using post-war data from the United States. He further found no evidence of electoral cycles and thus rejected the opposing opportunism hypothesis. Alesina (1987) extended Hibbs's theoretical approach using a Neo-Keynesian macroeconomic model and showed that it can be consistent with the assumption of rational expectations. However, later evidence

could not support Hibbs's (1977) results with more recent data from the United States (Alesina et al. 1997; Faust and Irons, 1999).

Like the opportunism school, the partisan school then turned its attention to the field of fiscal instead of monetary policy. As Drazen (2000) stated, empirical evidence, though not totally convincing, is at least more robust when fiscal policy is analyzed. Blais et al. (1993) provided some evidence that, in industrialized countries, left-wing governments tend to spend more than right-wing governments, although the difference is rather small. More robust results are reported by Cusack (1997) for a similar set of OECD countries, but a longer period.

With respect to the German states, the evidence is rather weak. Galli and Rossi (2002) tested the states' expenditures (as well as a variety of single expense items like those for infrastructure or health care) for partisan effects, but had to reject the hypothesis that the governments' political leaning significantly affects budget decisions. The same holds for Jochimsen and Nuscheler (2011) who found no effect of government color on budget deficits. Potrafke (2011a), however, found some differences as he showed that left-wing German state governments tend to raise expenditures for basic education stronger than right-wing coalitions, which in turn spend more on higher education.

5.2.3 Activist Fiscal Policy

While previous work about political cycles in fiscal policy analyzed whether elections or ideologies affect the *levels* of public spending or borrowing, Seitz (2000) and this chapter focus on its *cyclical*ity, thus its reaction to changes in overall economic conditions. The question whether or to what extent governments should try to counterbalance cyclical fluctuations of their economies (or should install automatic stabilizers) has been one of the most prominent disputes among politicians and economists, if not the most important one.

Keynesian economists, on the one side, prefer an active role of the state. Accordingly, governments should raise spending during economic downturns and

cut it during upswings in order to flatten the overall effect of economic cycles. On the other side, disciples of the neoclassical school are pessimistic about the governments' capabilities and blame counter-cyclical policies for the crowding out of private investment and hence for harming the economy in the long run. Thus, they advocate a passive role of the state.

As Auerbach et al. (2010) or Gechert and Will (2012) have pointed out, no consensus about the impact of counter-cyclical fiscal policy has evolved yet. According to Auerbach et al. (2010), the range of multiplier effects reported in the literature is inconclusive and differs strongly regarding covered periods, political measures and analyzed countries.

Bayoumi and Eichengreen (1995) have shown that American and German state governments indeed counterbalance economic fluctuations of their economy, with the extent being greater in Germany. Lane (2003), using data of the OECD countries' central governments, compared the cyclicity of fiscal policy and also showed that German governments tend to adjust their spending policy quite heavily to economic conditions compared to other countries. Galí and Perotti (2003), analyzing similar data from European countries, further showed that the introduction of fiscal restrictions via the Maastricht Treaty has not led to less counter-cyclical policies.

Seitz (2000) and Cusack (1999) provided the only attempts to link the political business cycle literature to that of activist fiscal policy. Seitz showed that the German states are conducting counter-cyclical (or at least weakly pro-cyclical) fiscal policies and reports values for the elasticity similar to those which Bayoumi and Eichengreen (1995) found for the American states. However, Seitz has to reject the partisan hypothesis that left-wing governments try to counterbalance economic fluctuations more strongly than right-wing governments. Contradictory, Cusack (1999) found some supporting evidence for a partisan effect on the cyclicity of fiscal policy, using federal level data of 14 OECD countries. He showed that left-wing governments raise the public deficit to a stronger extent in times of high unemployment than other governments do. However, his approach

is hardly comparable to the one used in this chapter, since Cusack used changes in employment as a proxy for economic conditions instead of GDP-growth.

5.3 Data

The panel dataset used in this study contains data of the 16 German federal states from the time between 1991 and 2010, thus all available years after the unification in 1990. In detail, it includes yearly macroeconomic (state-specific GDP) and budget data (public spending) as well as information about the ruling coalitions and the dates when state elections took place. Macroeconomic data was collected from the states' statistical agencies, while budget data was gathered from a database of the Ministry of Finance of Nordrhein-Westfalen (the largest German state). Throughout the following analyses, all values are given inflation-corrected in (Euro-) prices of the year 2010. A GDP-deflator time series was collected from the German Federal Office of Statistics. Finally, the election dates were gathered from various sources.

The German states provide an excellent background for testing the hypotheses set up above, as they do not differ much in their constitutional structure and the compulsory tasks they have to fulfill (Potrafke, 2011a; Seitz, 2000). Their main areas of responsibility are education, media, inner security, infrastructure and the (financial and political) supervision of local authorities. Thereby, the states feature a high amount of sovereignty: They have their own constitution, a sovereign parliament, government, judiciary and power of the purse. Although the state governments are not allowed to determine tax rates⁵ because this is decided on at the federal level⁶, and pay or receive equalization payments, they feature a wide scope to determine their own budget policy with their own priorities and are allowed to contract public debt on their own.⁷ Furthermore, their expenditures

⁵There are some few exceptions of minor importance like the land transfer tax.

⁶However, a weighted majority of the state governments has to approve tax rate changes in the second federal chamber, the Bundesrat.

⁷However, due to the adoption of a public sector-wide "tax brake", this will change after the year 2020.

account for a considerable share of total public expenditures in Germany, and their spendings amounted to 14.5 percent of Germany's GDP in 2010. Since different coalitions⁸ have governed in the states at different times, and state elections are held at different times, the panel dataset provides a good amount of within- and between-cluster variation which is needed to reliably estimate the influence of partisanship or electoral cycles.

The states differ widely in size: While the largest state, Nordrhein-Westfalen, has about 18 million inhabitants today, the smallest one, the city state of Bremen, has only less than one million residents. Furthermore, the states differ in economic power. In 2010, the GDP per capita amounted to an average of 22,271 Euros for the five "new" states in Eastern Germany, while it is 49,639 for Hamburg (the wealthiest state) and 37,101 for Hessen (the wealthiest territorial state in Germany). However, since the econometric analysis below uses logarithmized values and fixed-effects methods, cross-country differences should not bias the analyses.

I define the following variables: $GDP_{i,t}$ is state i 's inflation-adjusted gross domestic product in year t (in Euro-prices of the year 2010), while $SPEND_{i,t}$ is the corresponding amount of (state-level) public spending. Note that $SPEND_{i,t}$ depicts ex-post values of public spending, not ex-ante planned ones. Table 5.1 shows the mean values for all variables, divided by the single states.

The categorial variable $ELECTION_{i,t}$ is equal to 1, if state i holds state elections in the corresponding year t (and 0 otherwise). Table 5.2 shows when state elections took place in which state. Furthermore, the dummy variables $COALITION_{i,t}^k$ label the different governing coalitions.⁹ Throughout the covered timeframe, 14 different types of coalitions were in power in the German states. However, since some coalition types solely occurred in single states or for a short time, I follow Seitz (2000) and define eight different types of k :

⁸As Germany's electoral law is based on proportional representation, single-party governments are very seldom.

⁹When governments changed, $COALITION_{i,t}^k$ takes the value of the government which was in power for the major part of the year.

Table 5.1: Summary Statistics: Macroeconomic Values (Means)

State	Pop.	GDP	Growth	Spend.	Sp./GDP
	mill.	bill.	perc.	bill.	perc.
Baden-Württemberg	10.49	322.71	1.14	34.92	10.89
Bayern	12.19	386.41	1.56	38.78	10.11
Berlin	3.42	86.18	0.92	24.80	28.88
Brandenburg	2.56	46.40	4.31	11.29	25.71
Bremen	0.67	24.61	0.99	4.80	19.73
Hamburg	1.72	79.24	1.18	11.38	14.47
Hessen	6.03	199.57	1.24	20.65	10.40
Mecklenburg-Vorpommern	1.77	31.38	3.65	8.10	26.82
Niedersachsen	7.85	193.59	1.03	24.82	12.89
Nordrhein-Westfalen	17.91	495.99	0.84	53.90	10.91
Rheinland-Pfalz	4.00	98.12	0.84	13.19	13.49
Saarland	1.06	27.34	0.78	3.83	14.13
Sachsen	4.42	81.60	3.93	18.14	23.25
Sachsen-Anhalt	2.60	45.62	3.76	11.73	26.81
Schleswig-Holstein	2.77	69.54	0.89	9.00	12.99
Thüringen	2.42	42.45	4.42	10.67	26.46

Table 5.2: Summary Statistics: Election Dates

State	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10
Baden-Württemb.		X				X					X					X				
Bayern				X				X					X					X		
Berlin					X				X		X					X				
Brandenburg				X					X					X					X	
Bremen	X				X				X				X				X			
Hamburg	X		X				X				X			X				X		
Hessen	X				X				X				X					X	X	
Mecklenburg-V.				X				X				X				X				
Niedersachsen				X				X					X					X		
Nordrhein-W.					X					X					X					X
Rheinland-Pf.	X					X					X					X				
Saarland				X					X					X					X	
Sachsen				X					X					X					X	
Sachsen-Anhalt				X				X				X				X				
Schleswig-H.		X				X				X					X				X	
Thüringen				X					X					X					X	

1. Single-party governments solely formed by the Christian Democratic Union (CDU) or the Christian Social Union (CSU)¹⁰.
2. Single-party governments solely formed by the Socialdemocratic Party (SPD).
3. Two-party coalitions comprising the CDU (respectively, CSU) and the liberal Free Democratic Party (FDP).¹¹
4. Two-party coalitions formed by the SPD and the Green Party.
5. Two-party coalitions formed by the SPD and the Left Party.
6. Grand coalitions (Social and Christian Democrats) led by the CDU.
7. Grand coalitions led by the SPD.
8. Other two- or three-party coalitions made up of parties from both political blocks.¹²

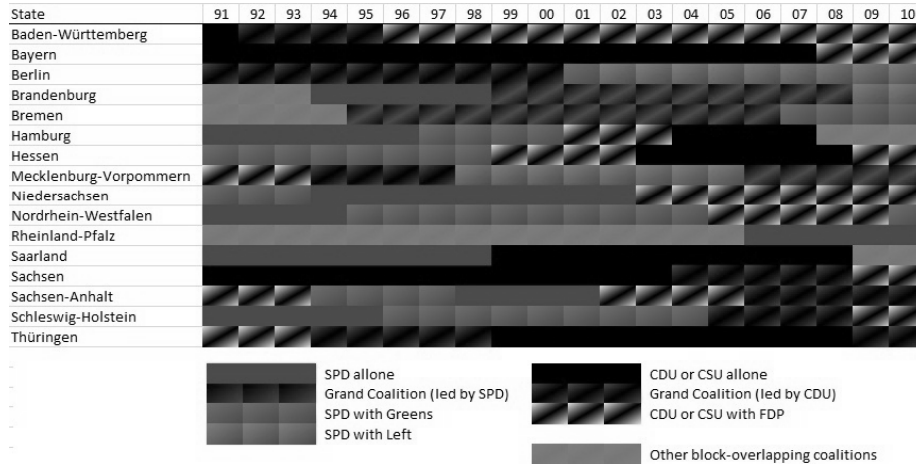
According to the self-conception of the political parties (and the convention of the German media and academia), I rate the CDU, CSU and FDP among the right-wing parties and the SPD, the Greens and the Left party among the left-wing block. Thus, coalition types 1 and 3 are right-wing governments; 2, 4 and 5 are left-wing and 6, 7 and 8 are block-overlapping coalitions formed with at least one party of both blocks. Figure 5.1 shows when which types of coalitions were in power in the 16 states.

¹⁰This party is only active in the state of Bavaria. It is further the sister party of the CDU, which in turn is not active in Bavaria, but in every other state.

¹¹Between 2001 and 2003, CDU and FDP formed a state government together with the right-wing populist Schill Party in the city-state of Hamburg. This coalition is also included here.

¹²This includes socialdemocratic-liberal, socialdemocratic-liberal-green, christiandemocratic-green and christiandemocratic-liberal-green coalitions which all were solitary cases.

Figure 5.1: Summary Statistics: Coalitions



5.4 Estimations

This section introduces the estimation strategy which is used to test the hypotheses set up above. First, I will present an empirical concept by Bayoumi and Eichengreen (1995) who developed a sensitivity-parameter measuring the cyclicity of public spending. Second, using my updated dataset of the German states' fiscal policy, I will repeat Seitz's (2000) analysis who tested whether the governments' partisanship affects the cyclicity of public spending. Afterwards, the opportunism hypothesis is examined, and, at last, both competing hypotheses are analyzed in one single empirical setting. Thereby, I chose to focus on public spending (instead of public deficits or single items of public spending like welfare payments) as it is the key aggregated setcrew of governmental behavior.

5.4.1 Cyclicity

Using data from the American states (timeframe: 1959-1992), Bayoumi and Eichengreen (1995) estimate a sensitivity-parameter β_0 , which measures the counter-cyclicity of the state governments' spending policy. The parameter depicts to which extent a government uses its budget policy to compensate the ups and downs of the economy.

Table 5.3: Correlations

	$GDP_{i,t}$	$SPEND_{i,t}$	$\ln(GDP_{i,t})$	$\ln(SPEND_{i,t})$	$\Delta \ln(GDP_{i,t})$	$\Delta \ln(SP_{i,t})$	$ELEC_{i,t}$	$TIME$
$GDP_{i,t}$	1							
$SPEND_{i,t}$	0.9509***	1						
$\ln(GDP_{i,t})$	0.9275***	0.9051***	1					
$\ln(SPEND_{i,t})$	0.8443**	0.9320***	0.9146***	1				
$\Delta \ln(GDP_{i,t})$	-0.1462**	-0.0980*	-0.1984***	-0.0441	1			
$\Delta \ln(SPEND_{i,t})$	0.0958*	0.087	0.0919	0.0824	0.1684***	1		
$ELECTION_{i,t}$	-0.0072	-0.0154	0.0045	-0.0107	0.0187	0.0429	1	
$Year$	0.0828	-0.0223	0.1059*	-0.0435	-0.2911***	0.0786	-0.0118	1
$COALITION_{i,t}^1$	0.0888	0.0553	0.0897	0.0454	0.07	-0.0096	-0.0259	0.0407
$COALITION_{i,t}^2$	-0.0641	-0.0673	-0.0477	-0.0812	-0.0695	-0.0665	-0.0229	-0.2641***
$COALITION_{i,t}^3$	0.2904***	0.2473***	0.2752***	0.2710***	0.1156**	0.1253**	0.0105	0.1948***
$COALITION_{i,t}^4$	0.1668***	0.1557***	0.1570***	0.1044*	-0.0508	0.1062*	0.0302	-0.1165**
$COALITION_{i,t}^5$	-0.1407**	-0.0663	-0.1281**	-0.028	-0.0718	-0.1192**	-0.005	0.1731***
$COALITION_{i,t}^6$	-0.1367**	-0.0336	-0.0866	0.0268	0.0024	-0.0686	0.0108	-0.0224
$COALITION_{i,t}^7$	-0.2234***	-0.2537***	-0.3082***	-0.3148***	-0.0267	-0.0605	0.0025	0.1316**
$COALITION_{i,t}^8$	-0.1472***	-0.1769***	-0.1209**	-0.1569***	-0.0225	0.0407	0.0025	-0.1004*

*: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$

Bayoumi and Eichengreen set up an empirical model that predicts the first difference of the states' spending quota, with the latter defined as the quotient of $SPEND_{i,t}$ and $GDP_{i,t}$. Thus, quota changes in percentage points are measured. As key independent variable they use the growth-rate of state i 's $GDP_{i,t}$ in year t in real terms (given as the first difference of the logarithmized GDP values) in order to be able to estimate the influence GDP-growth changes on the spending quota.¹³

Beside the GDP-growth, the first lag of the spending quota ($\frac{SPEND_{i,t-1}}{GDP_{i,t-1}}$) is also included in order to take account of possible level-effects, as it seems reasonable to argue that spending is raised less, if it has already been outstandingly high in the years before. Moreover, the regression includes a time-trend effect δ , a state-specific intercept μ_i and an idiosyncratic error term $\epsilon_{i,t}$. Pairwise correlations can be obtained from table 5.3.

The model by Bayoumi and Eichengreen (1995) is thus defined as:

$$\Delta \frac{SPEND_{i,t}}{GDP_{i,t}} = \alpha + \gamma \cdot \frac{SPEND_{i,t-1}}{GDP_{i,t-1}} + \delta \cdot TIME_t + \mu_i + \epsilon_{i,t} + \beta_0 \cdot \Delta \ln(GDP_{i,t}) \quad (5.1)$$

¹³One could argue that the quota should rather be affected by the GDP-growth of the previous year than of the current year, as budget plans are set up in the year before. However, using the simultaneous growth-values appears to be justified, since $SPEND_{i,t}$ depicts ex-post true spending instead of ex-ante planned spending. Furthermore, ex-post realisations regularly differ broadly from ex-ante predictions, as governments tend to adjust spendings regularly in the short-term (Heinemann, 2006).

A negative β_0 would thus reveal that governments try to counterbalance the economy's fluctuations. During upswings, they do so by raising their spending to a lesser extent than the GDP-growth increases, or even by cutting it. During downswings, the situation is vice versa. If β_0 is positive, governments adjust their spending pro-cyclically.

Bayoumi and Eichengreen estimated the cyclicity of the state governments' budget policy for the American states. They report significantly negative values for the cyclicity-parameter β_0 of -0.08 and -0.10 (depending on the covered time-frame). Following Bayoumi and Eichengreen, Seitz (2000) conducted the same analysis using data from the west German states (timeframe: 1976-1996) and reported a fairly similar cyclicity parameter of -0.07 .

When repeating the estimation¹⁴ with my dataset which contains more recent data from all German states (timeframe: 1991 to 2010), a twice as large cyclicity-parameter of -0.14 can be revealed. Thus, a change of GDP growth of one percentage points leads to a spending quota-change in the opposite direction of 0.14 points. It can thus be concluded that the German states' spending policy has become more counter-cyclically over time.¹⁵ The results are presented in column (1) of table 5.4.

However, some econometric problems occur: Using the first difference and the lag of the spending quota ($\frac{SPEND_{i,t-1}}{GDP_{i,t-1}}$) as dependent respectively independent variable appears to be problematic for two reasons. On the one hand, since the current GDP-values are used on both sides of the equation, dependent and independent variables move simultaneously and highly likely cause a bias.¹⁶ On the other hand, the emerging value of β_0 does not allow an unambiguous interpretation, at least if it is negative. A negative β_0 reveals that the spending quota declines if GDP-growth rises. However, no distinction can be made whether the

¹⁴As described below, a robustly estimated fixed-effects panel model with clustered standard errors is used.

¹⁵Since Seitz excluded data from the east German states, I repeat my analysis with a subsample of west German states only. The coefficients are now less negative, amounting to -0.10 . However, they are still more negative than in Seitz (2000).

¹⁶Jochimsen and Nuscheler (2011) further elaborate on this.

amount of spending (in Euros) is actually lowered or raised compared to the previous year, thus whether the policy is truly counter-cyclical (meaning that spending is reduced) or just weakly pro-cyclical (meaning that spending is only raised less than GDP-growth grows). As Seitz (2000) points out, as long as β_0 is smaller than $(-\frac{SPEND_{i,t}}{GDP_{i,t}})$, the Euro-amount of spending is reduced. However, if β_0 instead is smaller than zero, but larger than $(-\frac{SPEND_{i,t}}{GDP_{i,t}})$, spending rises only weakly pro-cyclically. Thus, the question whether spending is adjusted truly counter- or just weakly pro-cyclically depends on the idiosyncratic values of $SPEND_{i,t}$ and $GDP_{i,t}$.

In order to overcome both shortcomings, I apply a modification of this estimation strategy throughout this chapter's remainder. Instead of using the first difference of the spending quota $(\frac{SPEND_{i,t}}{GDP_{i,t}})$, I will follow Lane (2003) and Potrafke (2011b) and incorporate the growth rate of spending in real terms, operationalized by the first difference of the logarithmized spending values, $\Delta \ln(SPEND_{i,t})$, as regressand. Accordingly, the lagged logarithmized level of spending is inset as a control variable instead of the lagged spending quota.

The modified equation is now defined as:

$$\Delta \ln(SPEND_{i,t}) = \alpha + \gamma \cdot \ln(SPEND_{i,t-1}) + \delta \cdot TIME_t + \mu_i + \epsilon_{i,t} + \beta_0 \cdot \Delta \ln(GDP_{i,t}) \quad (5.2)$$

Since both GDP- and spending-values are now given in logarithms, the cyclicity-parameter β_0 can be interpreted more intuitively, as it forms a simple elasticity which measures by how many percentage points the governments change their spending-growth when GDP-growth rises by one point compared to the long-term trend. In any case, a negative elasticity would now reveal a counter-cyclical, thus truly Keynesian budget policy which raises expenditures during recessions and cuts it during upswings. Furthermore, a parameter larger than 0 and smaller than 1 reveals a weakly pro-cyclical adjustment, meaning that spending reacts in the same way as the GDP, but to a lesser extent. A value larger than 1 stands for a truly pro-cyclical adjustment.

5.4.2 Estimation Strategy

To conduct the estimation of equation 5.2, two different estimation techniques are used. These two models will always be used throughout the remainder of this chapter whenever an econometric equation is to be estimated.

Firstly, to account for the panel structure of the data and for the likely occurrence of unobserved heterogeneity, a fixed-effects panel estimation model (FE) is conducted. It uses the differences to the state-specific mean values as dependent and independent variables in order to let the state-specific intercepts μ_i drop out. A further advantage is that the fixed-effects model only accounts for the within-variation of the data and thus eliminates a possible size-bias which may occur because the included level of $\ln(SPEND_{i,t-1})$ is correlated with the states' size. Using a fixed- instead of a more efficient random-effects model is indicated according to a Hausman test which reveals the latter to be inconsistent in this case. All fixed-effects models will be estimated robustly with clustered standard errors, since a Breusch-Pagan test discloses some heteroscedasticity. At last, a unit root test as proposed by Levin et al. (2002) clearly reveals the process to be stationary, as it clearly rejects the corresponding null-hypothesis of non-stationarity.

However, due to the inclusion of the lagged value $\ln(SPEND_{i,t-1})$, equation 5.2 forms a dynamic panel model. The FE approach sketched above may thus suffer from inconsistency, since the lagged value is correlated with the error terms and thus endogenous. To overcome this shortcoming, a fixed-effects regression model with instrumental variables (IV FE) is used as a second estimation technique throughout this chapter. Therefore, an instrument that is uncorrelated with the disturbances, but correlated with the lag has to be found. Following Anderson and Hsiao (1981) and others, I choose to use the lagged first-difference, thus $\Delta \ln(SPEND_{i,t-1})$ as instrument for $\ln(SPEND_{i,t-1})$. First stage results of the estimations confirm that the instrument is relevant.¹⁷ Furthermore, F-statistics

¹⁷See column (3) in 5.4 and columns (2), (4) or (6) in table 5.5.

on the excluded instruments in the first stage reveal that the instrument is not weak, which is also supported by the Stock-Yogo weak identification test.¹⁸ To correct for heteroscedasticity and the serial correlation of the dynamic model, heteroscedasticity- and autocorrelation-consistent standard errors are computed.

Table 5.4: Regression Results I

Column	(1)	(2)	(3)
Model	5.2	5.3a	5.3b
Method	FE	FE	FE IV
Dependent Variable	$\Delta \frac{SPEND_{i,t}}{GDP_{i,t}}$	$\Delta \ln (SPEND_{i,t})$	$\Delta \ln (SPEND_{i,t})$
$SPEND_{i,t-1}/GDP_{i,t-1}$	-0.1905*** (<0.0001)		
$\ln (SPEND_{i,t-1})$		-0.1573*** (0.0004)	-0.0214 (0.8694)
$\Delta \ln (GDP_{i,t})$	-0.1431*** (<0.0001)	0.2045*** (0.0015)	0.2144*** (0.0014)
$TIME_t$	-0.0404*** (0.0053)	0.0000 (0.9237)	0.0005 (0.5916)
Constant	3.5989*** (<0.0001)	3.6726*** (0.0004)	0.4851 (0.8102)
Observations	304	304	288
R ²	0.7727	0.1224	0.0562
p-values in brackets; *: p < 0.1; **: p < 0.05; ***: p < 0.01			
IV: First stage			
Instrumented: $\ln (SPEND_{i,t-1})$			
Instrument: $\Delta \ln (SPEND_{i,t-1})$			-0.1919** (0.0380)
All other instruments?			Yes
First stage R ²			0.3670
F-Test of excluded instruments:			44.65 (<0.0001)
Underidentif. test (Kleibergen-Paap statistic)			23.96 (<0.0001)
Weak identification test (Kleibergen-Paap statistic)			44.65
Compared to Stock-Yogo critical values			<10%

Columns (2) and (3) of table 5.4 contain the computed values which result when equation 5.2 is estimated with the FE and the IV FE model respectively. The values for the cyclical parameter β_0 are highly significant and range between 0.20 and 0.21. A change of GDP-growth of one percentage points hence leads to a spending-growth change in the same direction of a fifth of a point.

Thus, German state governments adjust their spending policy weakly procyclical to GDP-growth. The numbers indeed reveal a considerable amount of counterbalancing which can be shown by comparing the elasticity of spending with the elasticity of the states' tax revenues. When the growth of tax revenues is regressed on GDP-growth in a simple fixed-effects setting, an elasticity of 1.27

¹⁸First stage- and test-results are reported in the respective regressions tables 5.4 and 5.5.

emerges. Since β_0 amounts only to 0.20, respectively 0.21, the state governments indeed tend to withhold a huge amount of their revenues during upswings and to spend additional (borrowed) money during downswings.

5.4.3 Partisan Hypothesis

As mentioned above, Seitz (2000) modified Bayoumi and Eichengreens' (1995) approach and tested whether the cyclicity of public spending is affected by the different types of ruling coalitions. This procedure shall be repeated in the following. Therefore, coalition-dummies ($COALITION_{i,t}^k$) are included into the equation, as well as interaction terms which multiply the coalition-dummies with GDP-growth. The dummies are equal to 1, if the respective coalition is in power.¹⁹ The cyclicity of the fiscal policy of coalition of type K is thus given by $(\beta_0 + \beta_K)$. Negative values of β_K thus reveal a stronger-than-average counter-cyclicity, positive ones less a less-than-average one.

Introducing Seitz's modifications into equation 5.2 thus leads to:

$$\begin{aligned} \Delta \ln (SPEND_{i,t}) = & \alpha + \gamma \cdot \ln (SPEND_{i,t-1}) + \delta \cdot TIME_t + \mu_i + \epsilon_{i,t} \\ & + \beta_0 \cdot \Delta \ln (GDP_{i,t}) + \sum_{k=1}^8 (\beta_k \cdot COALITION_{i,t}^k \cdot \Delta \ln (GDP_{i,t})) \\ & + \sum_{k=1}^8 (\lambda_k \cdot COALITION_{i,t}^k) \end{aligned} \quad (5.3)$$

Columns (1) and (2) of table 5.5 report the regression results. Significant values of β_k are only reported for the coalitions of type 3 (right-wing coalitions formed by the Christian Democrats and the Liberals). The corresponding coeffi-

¹⁹To avoid the dummy variable trap and to be able to estimate β_k -coefficients for all types of coalitions, Seitz specified the dummies using effect coding. Thereby, the dummies are revalued and take the value of -1, if a reference coalition is in power (in my analysis a single-party government formed by the Christian Democrats). The corresponding dummy of the reference coalition is excluded from the estimation, but is later computed by obtaining it from the variance-covariance matrix, since it is the negative value of the sum of the other coalition-coefficients. Thus, the sum of all coalition-coefficients is by definition equal to zero. In the following, this approach will be used, too.

cient β_3 amounts to -0.27 and -0.24, respectively, showing a tendency towards a more counter-cyclical policy. As the elasticity ($\beta_0 + \beta_3$) is just marginally larger than zero, the growth of spending is hardly affected by changes in GDP-growth. However, this result does not support the typical partisan hypothesis which predicts *left*-wing governments and not *right*-wing governments to be more active, thus to conduct more “Keynesian-style” counter-cyclical policies. Since no other coefficient β_k is significant, Seitz’ partisan hypothesis (2000) has to be rejected again.^{20, 21}

Regarding the λ_k -coefficients of the coalitions’ dummy variables (which would depict possible tendencies to change spending-growth independently from economic conditions), the picture is unclear as well. Positive and significant values for λ are reported for coalitions of type 3 and, at least in the FE IV specification, type 4 (Socialdemocrats and Greens), while negative ones are revealed for type 5-coalitions (Socialdemocrats and Left Party) in the FE IV model. However, only the weakly significant coefficient for type 4-coalitions fits with the standard hypothesis that left-wing governments prefer more “big government” than right-wing governments.

5.4.4 Opportunism Hypothesis

I now turn to the opportunism hypothesis. It shall be analyzed whether political competition affects the degree of cyclicity, thus whether the latter differs significantly between election and non-election years. I hypothesize that governments

²⁰If Seitz’ own estimation strategy (with $\frac{SPEND_{i,t}}{GDP_{i,t}}$ instead of $\ln(SPEND_{i,t})$ as key variable) is used, the picture is similar. Beside β_3 and λ_3 , only β_8 , the coefficient for the non-standard, block-overlapping coalitions is significant. The overall picture that the partisan hypothesis has no convincing predictive power in this setting is thus supported again.

²¹Further, it should be noted that Seitz’ empirical approach differs in another aspect, too. He does not incorporate the coalitions’ possible effects as a coefficient β_k which is added to or subtracted from the cyclicity-parameter β_0 like in equation 5.2, but rather as a multiplier of it. Accordingly, the cyclicity of the coalition K ’s fiscal policy would then be given by $\beta_0 \cdot (1 + \beta_K)$, while it is $\beta_0 + \beta_K$ in equation 5.2. However, when those kind of multiplicative coefficients are computed, again only β_3 is significant.

may try to please voters by adjusting economic fluctuations more strongly during an election year. Therefore, I include the $ELECTION_{i,t}$ dummy and an interaction term which multiplies it with GDP-growth. The cyclicity of the fiscal policy in election years is then given by $(\beta_0 + \beta_e)$.

Modifying equation 5.2 accordingly leads to:

$$\begin{aligned} \Delta \ln(SPENDING_{i,t}) = & \alpha + \gamma \cdot \ln(SPENDING_{i,t-1}) + \delta \cdot TIME_t + \mu_i + \epsilon_{i,t} \\ & + \beta_0 \cdot \Delta \ln(GDP_{i,t}) + \beta_e \cdot ELECTION_{i,t} \cdot \Delta \ln(GDP_{i,t}) + \lambda_e \cdot ELECTION_{i,t} \end{aligned} \quad (5.4)$$

Columns (3) and (4) of table 5.5 contain the results. The coefficient β_e is significant at the five percent level, however only in the IV FE model. It amounts to -0.23. Thus, governments seem to conduct a more counter-cyclical spending policy in election years. The total cyclicity is reduced from 0.28 in non-election years to 0.05 in election years, thus it is only marginally pro-cyclical then.

To sum up, the regressions provide at least some support for the adjusted opportunism hypothesis (that governments act more counter-cyclically while facing political competition). Furthermore, the standard hypothesis (that governments simply raise spending during election years), is not supported, as λ_e is not significant in both estimations.

5.4.5 Partisan versus Opportunism Hypothesis

In the last step, I will test both competing hypotheses within one single setting. Therefore, equations 5.3 and 5.4 are combined as follows:

$$\begin{aligned} \Delta \ln(SPENDING_{i,t}) = & \alpha + \gamma \cdot \ln(SPENDING_{i,t-1}) + \delta \cdot TIME_t + \mu_i + \epsilon_{i,t} \\ & + \beta_0 \cdot \Delta \ln(GDP_{i,t}) + \beta_e \cdot ELECTION_{i,t} \cdot \Delta \ln(GDP_{i,t}) + \lambda_e \cdot ELECTION_{i,t} \\ & + \left(\sum_{k=1}^8 \beta_k \cdot COALITION_{i,t}^k \cdot \Delta \ln(GDP_{i,t}) \right) + \left(\sum_{k=1}^8 \lambda_k \cdot COALITION_{i,t}^k \right) \end{aligned} \quad (5.5)$$

The results, shown in columns (5) and (6) of table 5.5, support the previous picture. On the one hand, the coalition types do not seem to have a conclusive effect on the cyclicity of the spending policy. On the other hand, the event of an election clearly influences the cyclicity of the fiscal policy (significant at the five percent level), according both to the FE and the FE IV model.

While the spending policy is found to be weakly pro-cyclical in non-election years (with a significant elasticity β_0 of 0.35, respectively 0.37), it is much less pro-cyclical in election years (with an elasticity $\beta_0 + \beta_e$ of only 0.07, respectively 0.06). Thus, a change of GDP-growth of one percentage point leads to a spending-growth change of more than a third of a percentage point in non-election and of less than a tenth of a point in election years. Since the election-dummy coefficients λ_e are at least weakly significant and positive, elections further seem to have a small positive effect on spending-growth which is independent from overall economic conditions.

5.5 Conclusion

The present chapter is to the best of my knowledge the first economic study which tries to disentangle whether the cyclicity of public spending policy is determined by different ideological preferences or by the timing of elections. The chapter extended Seitz's (2000) approach who tested only the former hypothesis, using data of the German states. Employing similar, however more recent data of the German states between 1991 and 2010, this chapter was able to show that the opportunism hypothesis seems to have more predictive power than the partisan hypothesis. While the composition of the ruling coalitions had almost no (and by all means no clear patterned) influence on the cyclicity of public spendings (as was also shown by Seitz, 2000), this was different when political competition was considered. During election years, public spending was revealed to be noteworthy more counter-cyclical. This conclusion was at least backed by three out of four econometric estimations.

Table 5.5: Regression Results II

Column Model Method Dependent Variable: $\Delta \ln (SPEND_{i,t})$	(1) 5.3a FE	(2) 5.3b FE IV	(3) 5.4a FE	(4) 5.4b FE IV	(5) 5.5a FE	(6) 5.5b FE IV
$\ln (SPEND_{i,t-1})$	-0.1946*** (0.0006)	-0.0938 (0.4441)	-0.1558*** (0.0004)	-0.0245 (0.8523)	-0.1920*** (0.0005)	-0.1048 (0.3932)
$\Delta \ln (GDP_{i,t})$	0.2669*** (0.0001)	0.2750*** (0.0008)	0.2322*** (0.0002)	0.2785*** (0.0013)	0.3516*** (<0.0001)	0.3742*** (0.0001)
$\Delta \ln (SPEND_{i,t}) \cdot ELECTION_{i,t}$			-0.1498 (0.1123)	-0.2333** (0.0422)	-0.2856** (0.0276)	-0.3147** (0.0131)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^1_{i,t}$	-0.1023 (0.4190)	-0.1581 (0.3830)			-0.1324 (0.2990)	-0.1669 (0.3840)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^2_{i,t}$	-0.1270 (0.3112)	-0.1464 (0.2130)			-0.1139 (0.3809)	-0.1402 (0.2704)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^3_{i,t}$	-0.2688** (0.0143)	-0.2359* (0.0598)			-0.3348*** (0.0067)	-0.3083** (0.0204)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^4_{i,t}$	0.1050 (0.3398)	0.0961 (0.6099)			0.1482* (0.0624)	0.1384 (0.3619)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^5_{i,t}$	-0.0309 (0.7583)	0.0892 (0.7843)			-0.0059 (0.9580)	0.0964 (0.7595)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^6_{i,t}$	0.0325 (0.5828)	-0.0172 (0.8753)			0.1198 (0.1351)	0.0776 (0.5064)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^7_{i,t}$	0.3200 (0.1935)	0.3138 (0.3148)			0.2764 (0.2728)	0.2591 (0.4044)
$\Delta \ln (SPEND_{i,t}) \cdot COALITION^8_{i,t}$	0.0723 (0.4651)	0.0585 (0.7208)			0.0425 (0.6774)	0.0439 (0.7777)
$ELECTION_{i,t}$			0.0067 (0.1125)	0.0083 (0.1176)	0.0077* (0.0538)	0.0087* (0.0869)
$COALITION^1_{i,t}$	-0.0016 (0.8920)	-0.0021 (0.8470)			-0.0004 (0.9720)	0.0000 (0.9980)
$COALITION^2_{i,t}$	0.0072 (0.3911)	0.0068 (0.2203)			0.0077 (0.3484)	0.0076 (0.1769)
$COALITION^3_{i,t}$	0.0223** (0.0179)	0.0171* (0.0707)			0.0222** (0.0145)	0.0176* (0.0566)
$COALITION^4_{i,t}$	0.0121 (0.1830)	0.0126* (0.0719)			0.0122 (0.1381)	0.0125* (0.0624)
$COALITION^5_{i,t}$	-0.0236 (0.1808)	-0.0230** (0.0109)			-0.0237 (0.1480)	-0.0234*** (0.0057)
$COALITION^6_{i,t}$	-0.0044 (0.6098)	-0.0031 (0.6157)			-0.0054 (0.4871)	-0.0038 (0.5256)
$COALITION^7_{i,t}$	-0.0163 (0.2763)	-0.0148 (0.1211)			-0.0152 (0.2559)	-0.0138 (0.1424)
$COALITION^8_{i,t}$	0.0044 (0.7504)	0.0065 (0.5925)			0.0021 (0.8576)	0.0037 (0.7360)
$TIME_t$	-0.0002 (0.7424)	0.0002 (0.8237)	0.0000 (0.9856)	0.0004 (0.6639)	-0.0002 (0.6681)	0.0001 (0.9446)
Constant	4.5448*** (0.0006)	2.1819 (0.3002)	3.6372*** (0.0004)	0.5581 (0.7815)	4.4829*** (0.0005)	2.4397 (0.2394)
Observations	304	288	304	288	304	288
R ²	0.1983	0.1514	0.1294	0.0715	0.2126	0.1731
p-values in brackets; *: p < 0.1; **: p < 0.05; ***: p < 0.01						
IV: First stage Instrumented Variable: $\ln (SPEND_{i,t-1})$						
Instrument: $\Delta \ln (SPEND_{i,t-1})$		-0.3071** (0.0290)		-0.2485** (0.0160)		-0.3875** (0.0210)
All other instruments?		Yes		Yes		Yes
First state R ²		0.4589		0.3694		0.4620
F-Test of excluded instruments:		45.52 (<0.0001)		44.37 (<0.0001)		45.71 (<0.0001)
Underidentif. test (Kleibergen-Paap statistic)		27.07 (<0.0001)		23.93 (<0.0001)		27.16 (<0.0001)
Weak identification test (Kleibergen-Paap statistic)		45.52		44.38		45.71
Compared to Stock-Yogo critical values		<10%		<10%		<10%

While an increase (or decrease) of GDP-growth by one percentage point is followed by an increase (or decrease) of spending-growth of between 0.28 and 0.37 points in non-election years, this elasticity is reduced to between 0.05 and 0.06 points in election years. Thus, governments tend to adjust their spending policy much stronger to economic conditions. However, since the elasticity is still positive, governments conduct no truly counter-cyclical, but rather weakly pro-cyclical policies (meaning that spending-growth is adjusted in the same way as GDP-growth changes, but to a lesser extent).

However, more research on this topic is indicated to solidify the results above. First, one could analyze whether the identified pattern is specific to the political system of the German states or whether it holds for a broader set of countries and/or for the central states or municipalities as well. Furthermore, it does not seem to be unrealistic that the partisan hypothesis could hold in countries with a more diverse political spectrum than the German one. Second, electoral cycles within the cyclicity of economic policy might also be identifiable regarding monetary policies (at least in countries with dependent central banks) or certain single budget items like public spending on social security or for civil servants. Third, it would surely be of interest, whether and in which situations governments benefit from conducting more counter-cyclical policies with regard to their chance to get reelected. However, finding a proper econometric identification strategy for the latter might be an obstacle due to endogeneity problems.

5.6 References

- Abrams, B. and Iossifov, P. (2006), Does the Fed Contribute to a Political Business Cycle?, *Public Choice*, 129(3), 249–262.
- Alesina, A. (1987), Macroeconomic Policy in a Two-Party System as a Repeated Game, *The Quarterly Journal of Economics*, 102(3), 651–678.
- Alesina, A. and Sachs, J. (1988), Political Parties and the Business Cycle in the United States, *Journal of Money, Credit and Banking*, 20(1), 63–82.
- Alesina, A. (1989), Politics and Business Cycles in Industrial Democracies, *Economic Policy*, 1989(8), 55–98.
- Alesina, A., Cohen, G. and Roubini, N. (1993), Electoral Business Cycles in Industrial Democracies, *European Journal of Political Economy*, 1993(9), 1–24.
- Alesina, A., Roubini, N. and Cohen, G. (1997), *Political Cycles and the Macroeconomy*, Cambridge, The MIT Press.
- Alesina, A., Perotti, R. and Tavares, J. (1998), The Political Economy of Fiscal Adjustments, *Brookings Papers on Economic Activity*, 197–266.
- Anderson, T.W. and Hsiao, C. (1981), Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association*, 76(375), 598–606.
- Auerbach, A., Gale, W. and Harris, B. (2010), Activist Fiscal Policy, *Journal of Economic Perspectives*, 24(4), 141–164.
- Bayoumi, T. and Eichengreen, B. (1995), Restraining Yourself: The Implications of Fiscal Rules for Economic Stabilization, *Staff Papers - International Monetary Fund*, 42(1), 32–48.
- Berger, H. and Woitek, U. (1997), Searching for Political Business Cycles in Germany, *Public Choice*, 91(2), 179–97.

- Blais, A., Blake, D. and Dion, S. (1993), Do Parties Make a Difference? Parties and the Size of Government in Liberal Democracies, *American Journal of Political Science*, 37(1), 40–62.
- Cusack, T. (1997), Partisan Politics and Public Finance: Changes in Public Spending in the Industrialized Democracies, *Public Choice*, 91, 375–395.
- Cusack, T. (1999), Partisan Politics and Fiscal Policy, *Comparative Political Studies*, 32(4), 464–486.
- Downs, A. (1957), An Economic Theory of Political Action in a Democracy, *Journal of Political Economy*, 65(2), 135–150.
- Drazen, A. (2000), The Political Business Cycle After 25 Years, *NBER Macroeconomics Annual*, 2000(15), 75–138.
- Faust, J., and Irons, J. (1999), Money, Politics, and the Post-War Business Cycle, *Journal of Monetary Economics*, 43, 61–89.
- Franzese, J. (2002), Electoral and Partisan Cycles in Economic Policies and Outcomes, *Annual Review of Political Science*, 5, 369–421.
- Galí, J. and Perotti, R. (2003), Fiscal Policy and Monetary Integration in Europe, *Economic Policy*, 18(37), 533–572.
- Galli, E. and Rossi, S. (2002), Political Budget Cycles: The Case of the Western German Länder, *Public Choice*, 110(3/4), 283–303.
- Gechert, S. and Will, H. (2012), Fiscal Multipliers: A Meta Regression Analysis, *IMK Working Paper* 97.
- Heinemann, F. (2006), Planning or Propaganda? An Evaluation of Germany’s Medium-term Budgetary Planning, *FinanzArchiv / Public Finance Analysis*, 62(4), 551–578.
- Hibbs, D. (1977), Political Parties and Macroeconomic Policy, *The American Political Science Review*, 71(4), 1467–1487.

- Jochimsen, B. and Nuscheler, R. (2011), The Political Economy of the German Länder Deficits: Weak Governments Meet Strong Finance Ministers, *Applied Economics*, 43(19), 2399–2415.
- Lane, P. (2003), The Cyclical Behaviour of Fiscal Policy: Evidence from the OECD, *Journal of Public Economics*, 87(12), 2661–2675.
- Levin, A., Lin, C. and Chu, C. (2002), Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, 108, 1–24.
- Markwardt, G. (2008), *Wahlen und Konjunkturzyklen*, Gabler Edition Wissenschaft, Gabler Verlag, Wiesbaden, 2008.
- Nordhaus, W. (1975), The Political Business Cycle, *The Review of Economic Studies*, 42(2), 169–190.
- Potrafke, N. (2011a), Public Expenditures on Education and Cultural Affairs in the West German States: Does Government Ideology Influence the Budget Composition?, *German Economic Review*, 12(1), 124–145.
- Potrafke, N. (2011b), Does Government Ideology Influence Budget Composition? Empirical Evidence from OECD Countries, *Economics of Governance*, 12(2), 101–134.
- Rogoff, K. (1990), Equilibrium Political Budget Cycles, *The American Economic Review*, 80(1), 21–36.
- Schneider, C. (2007), Politischer Opportunismus und Haushaltsdefizite in den Westdeutschen Bundesländern, *Politische Vierteljahresschrift*, 48(2), 221–242.
- Schneider, C. (2010), Fighting with one Hand Tied Behind the Back: Political Budget Cycles in the West German States, *Public Choice*, 142, 125–150.
- Seitz, H. (2000), Fiscal Policy, Deficits and Politics of Subnational Governments: The Case of the German Laender, *Public Choice*, 102(3/4), 183–218.

Vaubel, R. and Dreher, A. (2009), Foreign Exchange Intervention and the Political Business Cycle: A Panel Data Analysis, *Journal of International Money and Finance*, 28, 755–775.

Chapter 6

Final Remarks

This thesis aimed to add some new insights to different aspects of competition and behavioral economics. In chapter 2, we provided new evidence that sequential contests can be severely biased by order effects. Analyzing data from a TV cooking contest where participants perform in turn, we revealed that the first contestant wins significantly less frequently. Since the starting order was always randomly defined by the TV producers, the effect should not be due to the contestants' characteristics, but rather due to psychological order biases. We thus recommended to define the starting order as randomly as possible, for example, by reversing the order, if a contest is comprised of more than one heat.

Chapter 3 revealed that price regulations may not have the intended positive effects in terms of strengthening competition between gas stations. Using a controlled laboratory experiment, we could show that, quite the contrary, two out of three currently discussed regulations would rather harm competition and consumer welfare. These are the Austrian rule which forces gas stations to increase their prices at most once a day (while price cuts are always possible) and the Luxembourg rule which introduces price ceilings for gasoline. A third proposed rule, the Western-Australian rule, which allows only one price change per day, was not found to significantly change price levels compared to the non-regulated baseline scenario. Since none of the proposed measures was found to significantly reduce consumer prices, we recommended not to introduce any of them.

Chapter 4 provided some evidence that the reason why companies tend to publish overoptimistic forecasts of their own future success may likely be their incentives to do so. This would support the strategic deception-hypothesis of Flyvbjerg et al. (2002) and reject the competing hypothesis that overoptimism is mainly due to cognitive biases. While all previous studies analyzed forecasts from public disclosures and unanimously found them to be too optimistic on average, I evaluated the accuracy of secret, company-internal forecasts and revealed them to be *overpessimistic* on average. I argued that the reason for this divergent result may likely be that the forecasts are secret and not for publishing purposes. However, additional research is necessary to further solidify this conclusion. Furthermore, econometric regressions revealed that joint-stock companies are prone to be more overoptimistic, while a greater share of women among the workforce tends to reduce overoptimism.

In Chapter 5, it could be shown that German state governments tend to counterbalance economic fluctuations to a greater extent during election years compared to non-election years. This provided some support for the opportunism hypothesis that governments are prone to change their policies in election years in order to please the swing voters and to assure reelection. The competing partisan-hypothesis, stating that differences in fiscal policies can be explained by the parties' differing basic beliefs, had to be rejected. The latter was in line with Seitz (2000), who had previously tested the partisan-hypothesis with similar data.