Mining the Gold of the 21st Century: Studying Markets and Organizations Using Observational Data

Michael Hofmann

Dissertation

Ludwig-Maximilians-Universität Munich
Department of Economics
March 2023

Mining the Gold of the 21st Century: Studying Markets and Organizations Using Observational Data

Inaugural-Dissertation

zur Erlangung des Grades Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2023

vorgelegt von

Michael Hofmann

Referent:	Prof. Dr. Florian Englmaier
Korreferentin:	Prof. Dr. Claudia Steinwender
Promotionsabschlussberatung:	12. Juli 2023

Datum der mündlichen Prüfung: 05. Juli 2023

Namen der Berichterstatter: Prof. Dr. Florian Englmaier, Prof. Dr. Claudia Steinwender, Prof. Dr. Till Stowasser

Acknowledgments

First and foremost I would like to thank my doctoral advisor Florian Englmaier. His support and guidance was tremendously helpful at every stage of my dissertation. I am grateful for his encouragement and unbroken optimism about my research, especially in times when I was in doubt. Besides his direct support for me, his continuous efforts to make our chair and the whole department such a welcoming place has played a crucial part in the decision to start my doctoral studies. I would also like to thank my second advisor Claudia Steinwender. Her motivating and insightful feedback, especially on Chapter 2 and Chapter 3, helped me greatly improve my research. Finally, I would like to thank Till Stowasser, who coauthored Chapter 1 and kindly agreed to serve on my doctoral committee. Even more, I am grateful for his early and continued support — starting even before my bachelor's thesis — which not only helped me write better research papers today but inspired me to pursue a PhD in the first place.

I further had the chance to work together with Nadzeya Laurentsyeva and Thomas Fackler, who were a great source of motivation and inspiration. Chapter 2 is the result of this productive collaboration and I am proud of the work we have done together as a (sometimes co-located, sometimes mixed and sometimes distributed) team. I want to thank Stefanie Wolter for helping me understand and prepare the data for Chapter 3.

It was a pleasure to share my doctoral studies with a great group of colleagues and friends. I want to particularly thank Dominik Grothe, my officemate, Anne Niedermaier, Svenja Friess, Silvia Castro, Helene Strandt, Christoph Schwaiger, Clarissa Mang, Julian Heid, Thomas Überfuhr, Luisa Wallosek, and Lena Greska. However, I have greatly enjoyed spending time with each and every one and will keep countless fond memories from my time spent at the Munich Graduate School of Economics. Another thanks goes out to Manuela Beckstein, who patiently helped me battle the (perceived) jungle of bureaucracy that one faces when working at a German public institution.

I am grateful for the generous financial support from the DFG Research Training Group 1928: Microeconomic Determinants of Labour Productivity as well as from the Bavarian Research Institute for Digital Transformation (bidt).

Outside the academic environment I want to extend a special thanks to all my friends, who have made the last few years some of the best of my life. Especially in stressful times, spending time with my friends provided the necessary counterbalance for me to continue on with my research. Their contribution to my dissertation is invaluable. In this regard I am especially indebted to my loving soon to be wife Vera Thanner. For more than 10 years now, she has offered her continued support, understanding and love. Without her, I would not have been able to even start my doctoral studies, let alone complete them.

Finally, I am grateful to my parents, Astrid and Bernhard, who have always believed in me and whose support I can always count on.

Without you, all of you, I would not be where I am today.

Thank you!

Michael Hofmann, March 2023

Contents

List of Tables

List of Figures

Preface

"Acts demolish their alternatives, that is the paradox."

We can't know what lies at the end of the road not taken.

Joshua D. Angrist & Jörn-Steffen Pischke[1]

Uncovering causal relationships, the true impact of some treatment on an outcome, is the ultimate goal of most applied research in social sciences. This endeavor is always complicated by the fundamental problem of causal inference: For any individual unit, we can observe either the treated or the untreated value of the outcome, but never both. Therefore, it is impossible to observe the individual causal effect of a treatment on a single unit (Holland 1986).[2] At first glance, naive comparisons of treated and untreated units in observational data[3] might measure *average treatment effects*, but are virtually always plagued by a set of problems that are collectively coined *selection bias*. This describes the possibility that observed impacts of a treatment reflect underlying systematic differences between treated and non-treated units rather than the causal impact of the treatment. Therefore, the underlying causes of any observed correlation remain unclear. How can one take on this fundamental challenge of causal identification?

[1]This quote appears in the book *Mastering 'Metrics: The Path from Cause to Effect* (Angrist and Pischke 2014). The first sentence is taken from James Salter's novel *Light Years*.

[2]Holland (1986) has formulated the fundamental problem of causal inference. It is based on the framework of potential outcomes which was first proposed by Jerzy Neyman (Splawa-Neyman et al. 1923/1990) and was extended to a more general framework by Rubin (1974).

[3]Observational data is generated without any controlled interventions by the researcher.

One solution is to conduct randomized control trials (RCTs) and overcome the selection problem by actively randomizing who receives treatment. Due to their power of isolating causal treatment effects, RCTs have a long history in social sciences and produce findings of tremendous social relevance to this day: In 2019, Abhijit Banerjee, Esther Duflo and Michael Kremer received the Nobel Memorial Prize in Economic Sciences for their experimental approach to alleviating global poverty. Although RCTs still are the gold standard for achieving causal identification internally, through their nature of studying a limited number of subjects in a specific context, external validity remains a key challenge. An even more serious drawback is that RCTs require the researcher to randomly assign treatments, which in many situations is unethical, financially unaffordable or practically impossible.

In order to answer many pressing questions in social sciences researchers have to rely on observational data and find other ways to overcome the problem of selection bias. After a credibility crisis in the late 1980s (e.g. LaLonde 1986; Leamer 1983) empirical economists have increasingly focused on the quality of empirical research designs (Angrist and Pischke 2010). This design-based approach aims at analyzing non-experimental data in an experimental spirit by exploiting as good as random variation through *natural experiments* and by using *quasi-experimental* methods. The standard toolkit includes instrumental variables, regression discontinuity designs and difference-in-differences designs and is often complemented by matching methods. The beauty of these tools is that they focus a researcher's attention more on how to approach a causal question rather than applying evermore statistically sophisticated estimation models. Early and highly influential contributors are David Card (e.g., Card 1990, 1992) as well as Joshua D. Angrist and Guido W. Imbens (e.g., Imbens and Angrist 1994; Imbens and Rubin 1997). They collectively won the Nobel Memorial Prize in Economic Sciences in 2021, which highlights the importance of the design-based approach and credible identification strategies.

In more recent times a new arrow has entered the quiver of applied economics, or perhaps even an entirely new quiver of arrows: machine learning (ML). The ultimate goal of ML is (usually) not to estimate causal relationships, rather ML tools are designed to solve tasks of prediction, classification or data-grouping (Athey 2018). This does by no means undermine its value for empirical economic research. ML is often employed as an intermediate step to generate new variables, e.g. by summarizing vast amounts of data, which are subsequently used in econometric models. Further, ML can improve established econometric models, for example by automating specification searches in a data-driven way. Embracing this complementarity of econometrics and ML can help address the criticism of Leamer (1983) by focusing more on credible research designs and at the same time applying more sophisticated models. To put it in slightly more positive words than Susan Athey (2018): As ML automates some of the routine tasks of data analysis, it *frees up time for economists to practice* the art of creating credible and impactful empirical work.[4]

Whether we are engaging in social networks, navigating to the nearest café on our smartphone, or looking for a new place to live online, we generate a tsunami of data throughout every aspect of our life every single day. In doing so, we contribute to an ever-growing pile of data collected by governments, institutions and private firms. This pile not only consists of consumer data but contains data generated within organizations as well, for example through employee surveys or the systematic collection of team collaboration metadata. However, many organizations struggle to merely keep up with storing all the bits and bytes in adequate structure and high quality. Extracting and presenting useful information or, in other words, translating the data into actionable insights, is harder still.

With the help of econometric tools, we can utilize the vast amounts of data available to address issues of both economic and social relevance. Which market frictions further increase the social gap between rich and poor? What implications do remote work models

[4] The original quote in Athey (2018), which provides a great overview of ML in economic research, reads: "As ML automates some of the routine tasks of data analysis, it becomes all the more important for economists to maintain their expertise at the art of credible and impactful empirical work" (p. 543).

have on team collaboration and productivity? How can companies manage, motivate and retain employees in a world of increased skills shortage? The chapters of this dissertation address these questions applying quasi-experimental methods and machine learning, and thus contribute to mining the gold of the 21st century.

Chapter 1 is situated in the field of behavioral economics and addresses an issue of social relevance. In many market situations, individuals exhibit limited attention when processing information, a phenomenon that is well documented by now. We have much less knowledge, however, on whether some market participants deliberately exploit limited attention. Together with Till Stowasser, I study the German housing market to uncover whether landlords capitalize on a form of limited attention called left-digit bias, the tendency of individuals to focus on the leftmost digit of a number. We show that some landlords employ sophisticated strategies and exploit left-digit bias not only by setting prices just below multiples of €100, but they exploit it again (and more subtly) by discontinuously increasing rent prices if an apartment's size is just above a left-digit threshold (e.g. $60m^2$).[5] Other landlords follow a simpler strategy of rounding (up) rent prices without additionally increasing them at size thresholds. In order to arrive at these results, we employ an identification strategy based on a well established method from the applied economist's toolkit: regression discontinuity. The idea of our slightly modified version is that once we account for exact apartment size measures, changes in the leftmost digit should not play a role if it was not for left-digit bias. Our results are of substantial social relevance, since the strategic exploitation of limited attention further intensifies the distribution of wealth from relatively poorer tenants to relatively richer landlords. Uncovering this socially undesired market friction helps policymakers to counter the struggle that many face in the search of an affordable home. This would not have been possible without an increasing availability of microdata and the usage of

[5] To individuals who suffer from left-digit bias, these prices (e.g. €599) appear lower than they actually are. At the same time, the size difference of apartments just below and just above a left-digit threshold appears larger than it actually is and thus "justifies" a discontinuous price jump.

quasi-experimental methods.

Chapter 2 is coauthored with Thomas Fackler and Nadzeya Laurentsyeva. It addresses the timely question of how to organize collaboration in virtual work environments. To achieve this, we exploit the largest natural experiment of recent history: the COVID-19 pandemic. Sending the entire world into lockdowns, the pandemic caused an explosion of hybrid and fully remote work models. This sudden change in work arrangements combined with a setting in which online collaboration for some teams was common already before creates a rare opportunity to study the consequences of switching to remote work and the determinants of successful online collaborations. We use comprehensive data from GitHub, the world's largest platform for open-source software development, and study the performance of teams before and during the pandemic. In a difference-in-differences design, we exploit variation in the impact of COVID-19 on teams with different spatial organizations before the shock. In particular, we differentiate between teams that were geographically distributed before the pandemic and those that were geographically co-located. Without the option of face-to-face collaboration, co-located teams had to adapt to virtual work models. In contrast, the production process of distributed teams was hardly affected because they already relied on online collaboration before the pandemic. While the productivity of co-located teams took a substantial hit during the pandemic, otherwise similar distributed teams remained resilient. Our results highlight that moving previously non-remote teams to the virtual world entirely can cause significant performance losses even for digitally skilled workers such as software developers. What can be done to mitigate these negative effects? We investigate potential mechanisms and show that access to remote talent as well as experience are important factors of success, but they cannot explain the performance differences between co-located and distributed teams. Instead, our results highlight the crucial role of communication for productive online collaborations. Therefore, we conclude that setting up systems for effective online communication can help avoid productivity loss in remote teams.

Chapter 3 builds on the literature of management practices which dates back almost as far as the field of economics itself. It has gained more attention in recent years, owing to the increasing availability of microdata. This new strand of the literature documents consistent (causal) evidence that structured management has positive effects on firm performance. I use German panel survey data to analyze whether these results lead to a trend toward structured management styles. A key challenge of analyzing such survey data is its high dimensionality. Therefore, I utilize machine learning in order to uncover latent groupings in the data, which I call management styles. The algorithm uncovers two styles: One is characterized by intensive use of structured practices while the other lacks these practices. Uncovering these latent structures would have been impossible without adding ML to the standard toolkit of empirical economics. I document two main patterns in the German management landscape: First, management varies widely across firms with some firms employing the structured style while others do not. Second, in the latter group, no secular trend toward more structured styles can be observed. Surprisingly, I find that although many managers introduce relevant practices, they dismiss them again after a short time. My results, together with those of the remaining literature, carry important lessons for managers, as many firms will not reach their full potential if they fail to adopt structured management styles. However, the questions raised in this chapter constitute a significant contribution to the literature on their own: Why are management styles so rigid? What keeps managers from adopting beneficial practices? Which factors within firms are key contributors to the success of structured practices? I discuss two potential factors holding back firms: a miscalculation of cost-benefit trade-offs and a lack of appropriate corporate culture. These factors should be investigated in future research to help managers overcome the obstacles of adopting structured management styles.

Chapter 1

Charmers vs Rounders: Rent-price Discontinuities in the German Housing Market[1]

Abstract

We study the role of left-digit bias — a tendency to focus on the leftmost digit of a number — in the rent-price-setting behavior of landlords. We show that there are two types of landlords — which we name *charmers* and *rounders* — and that their strategic behavior differs systematically. Using web data on German apartment listings we document that charmers exploit left-digit bias by increasing rent prices at salient apartment-size measures. In contrast, rounders aim to increase rent prices for all apartments, not taking advantage of limited attention. In addition, we provide evidence that landlords exploit behavioral biases to a larger extent in markets that grant them higher market power. Being able to identify heterogeneity in bias-exploiting behavior allows us to further our understanding of how price discontinuities evolve, even in high-stakes settings such as the housing market.

[1]This chapter is based on joint work with Till Stowasser.

1.1 Introduction

Facing complex economic decisions individuals often reach their cognitive limits and are not able to take all relevant information into account (Simon 1955). Instead, they rely on mental shortcuts to reduce complexity of information, which leads to biased decisions. One of these biases, the so called left-digit bias, implies that people fully process the leftmost digit of a number but fail to pay full attention to the remaining digits. It has been shown that left-digit bias can lead to discontinuous price jumps if economic agents use this mental shortcut when they process product characteristics. For example, prices of used cars drop substantially at left-digit mileage thresholds (Englmaier et al. 2018b; Lacetera et al. 2012). However, it is not yet sufficiently understood which agents are subject to limited attention and whether some agents actively exploit limited attention of others.

In this paper we use German housing-market data to study the role of left-digit bias in the price-setting behavior of landlords. We show that rent prices discontinuously jump at left-digit thresholds of apartment-size measures[2] (multiples of $10m^2$) and investigate whether this pattern is driven by landlords who exploit limited attention.[3] In our setting we are able to distinguish landlord types and identify those who are more likely to exploit left-digit bias. This makes it possible to study whether the above-described discontinuity patterns are *primarily* driven by these landlords. Hence our data provides a unique opportunity to study heterogeneity in price-setting strategies of economic agents and whether some exploit limited attention of others.

To identify landlord types, we study clustering patterns in rent prices and find two bunching

[2]In the German housing market it is common to report the size of apartments in square-meter measures and this is one of the main factor tenants take into account.

[3]For example, consider two almost identical apartments which only slightly differ in size, say $59m^2$ versus $60m^2$. If tenants focus on the leftmost digit of apartment size, they will perceive the size difference of these apartments as larger than it actually is. Landlords can exploit this left-digit bias by increasing the rent price for the larger apartment by a higher amount than the actual size difference would justify. This would lead to discontinuous rent-price jumps at $10m^2$ marks of apartment-size measures.

regions: *Round prices* are set to multiples of €50. So-called *charm prices* exploit left-digit bias by setting digits further to the right to a 9, thus keeping the leftmost digit low. We argue that these price formats reflect landlords' overall strategies and thus can be used to classify landlords into strategic types — *charmers* and *rounders*. While charmers exploit left-digit bias in the price dimension, rounders (and other landlords) are either not aware of the bias or are aware of it but do not exploit it.

Distinguishing these landlord types, we are able to study heterogeneity in their price-setting behavior with respect to a second dimension: apartment size. We employ a regression-discontinuity design to test whether (some) landlords discontinuously increase rent prices at left-digit thresholds of size measures.[4] According to our results, rounders do not increase prices at left-digit thresholds of apartment size and neutral landlords increase prices only slightly, but charmers increase prices sharply by up to €25.[5] Because we have already established that charmers exploit left-digit bias in the *price dimension*, we can rule out that they are biased themselves. Therefore, our results strongly suggest that the discontinuity patterns in the *size dimension* are also driven by exploitive behavior of charmers. This is further supported by the contrasting results we find for rounders, who do not show this behavior in either dimension. If the discontinuities were driven by biased landlords, we would expect that (some) rounders are also biased and therefore increase prices at left-digit thresholds of size measures, as well. Although rounders do not appear to exploit biases, we document that they in general demand higher rent prices than other landlords, characterizing them as a profit-seeking type of landlord. Identifying landlord types and heterogeneity in

[4]Because we observe bunching in reported size measures, we explicitly account for size rounding to alleviate concerns of manipulative sorting. We find that rent prices are systematically higher if size measures are rounded and that this pattern is mainly driven by rounders. However, we refrain from interpreting this result as size-rounding premium, but instead we find suggestive evidence that this result appears to be driven by a sorting mechanism.

[5]A linear approximation in the spirit of Lacetera et al. (2012) implies that these price jumps amount to 36% of the average rent-price increase between two thresholds and are comparable to previous findings in other markets.

bias-exploiting behavior helps us to better understand how price discontinuities evolve.

Our data allows us to study a second type of heterogeneity, namely across local levels of market tightness and thus landlords' market power. If we observe that landlords exploit left-digit bias, they will likely intensify their behavior when they have high market power. Our results show that this is the case as charmers (and neutral landlords) increase rent prices at left-digit thresholds of size by a larger amount in tight markets. While rounders do not exploit left-digit bias even in tight markets, we find that our previous result — they demand higher rent prices in general — is exclusively driven by tight markets. This is in line with our interpretation of profit-seeking rounders, as extracting profits is easier when their market power is high. These results highlight the economic and social relevance of our findings, as the economic burden for tenants, which is already high in tight markets, is further increased by landlords' exploitive behavior.

Our paper relates to several strands of a literature documenting various behavioral biases in economic decision-making. One part of the literature analyzes clustering of numeric measures at focal numbers, which are often round numbers. For example, Lynn et al. (2013) use data on pay-what-you-want purchases, tipping in restaurants, as well as gasoline purchases. In all three cases customers show a tendency to choose round prices, which the authors explain by a subjective preference for round numbers. Round numbers also play a role in stock and foreign exchange markets (Sonnemans 2006; Sopranzetti and Datar 2002) as well as in non-economic situations (Allen et al. 2017). More closely related to this study, Pope et al. (2015) document extensive clustering of negotiated house prices at round numbers, especially multiples of $50,000. We add to this literature by not only confirming previous results, but also documenting a predictive correlation of rounding behavior across dimensions.

Further, this paper contributes to research that focuses on the implications of limited attention and, in particular left-digit bias. A theoretical model of limited attention is provided by DellaVigna (2009), who decomposes the value of a good into a visible and an opaque component. Due to limited attention, the latter is not fully taken into account, which leads to discrepancies between the perceived value and the true value of a good. Lacetera et al. (2012) apply this model to the case of left-digit bias. In this model, individuals fully process the leftmost digit of a number, but only pay partial attention to the remaining digits. They find empirical evidence for left-digit bias in a used-car market, where prices discontinuously drop at 10,000-mile odometer marks. Englmaier et al. (2018b) confirm these results in a German used-car market and show that people are also inattentive to exact first-registration dates. Left-digit bias and, more generally, limited attention has been shown to play a role in several other occasions, such as stock markets (Gilbert et al. 2012), online games (Englmaier et al. 2018a), hospital ratings (Pope 2009) taxation (Chetty et al. 2009) and energy labels (Palmer and Walls 2015). We contribute to this literature by documenting heterogeneity in discontinuity effects and providing evidence that some agents exploit inattention of others.

Another branch of the literature has focussed on analyzing left-digit bias specifically in housing markets. In an early study, Allen and Dare (2004) analyze the effects of charm list prices (e.g. €499,000), and show that these prices lead to higher final negotiated transaction prices in the Florida housing market. Repetto and Solís (2019) document similar results and argue that charm pricing attracts more bidders in Swedish apartment auctions. Chava and Yao (2017) further document that charm prices can lead to a 3.8% increase in sale likelihood and 3.5 days (5%) less time on the market. Meng (2019) analyzes repeat house sales in the greater London market and finds that houses are sold for a premium of 4% if the previous sales price was round rather than charm. She explains these results by a combination of left-digit bias and reference dependence. Due to left-digit bias, the perceived value of an owned house is lower if it was bought at a charm price rather than a round price. Following the implications of prospect theory, reference dependence and loss aversion, homeowners who bought at a charm price are thus willing to sell at a lower price than those who bought at a round price. In contrast to the studies above, this paper does not analyze the direct market outcomes of price formats. Instead, we focus on house characteristics and show that left-digit bias also plays a role in the size dimension, leading to rent-price discontinuities.

The remainder of the paper is organized as follows. Section 1.2 introduces the data and Section 1.3 explores clustering patterns in apartment-size measures and rent prices. Section 1.4 presents our empirical results and documents heterogeneities across landlord types and across different market situations. Section 1.5 provides robustness tests and Section 1.6 concludes.

1.2 Data

1.2.1 Data source

We obtained the apartment data from ImmobilienScout24[6] (IS24), the leading German online real-estate portal, with about 12 million visitors and roughly 500,000 real-estate offerings a month.[7] IS24 takes the role of an intermediary between landlords and tenants. For a fixed fee, landlords can place their apartment listings. IS24 then provides a free vacancy search tool for apartment seekers.

Figure A.1 in the Appendix contains screenshots of the search process on IS24. On the starting page, shown in Figure A.1a, IS24 presents a simple search mask with several filter options. Tenants can specify whether they wish to rent or buy, the search area, the type of real estate, the maximum rent as well as the minimum room number and apartment-size requirements.[8] After entering these search criteria, a result list is presented, highlighting the most important characteristics of each apartment. An example can be found in Figure A.1b, which reveals that basic rent price[9] (in € per month), apartment size (in square meters) and

[6] `https://www.immobilienscout24.de`

[7] Source: `https://www.immobilienscout24.de/unternehmen/immobilienscout24.html`; last accessed: May 14, 2018

[8] Note that we focus on apartments for rent, excluding commercial property and property for sale.

[9] Basic rent refers to the German term "Kaltmiete" (cold rent) and excludes any additional costs such as for heating, water supply or waste disposal.

the number of rooms are most prominently specified. Selecting one of the offers leads to a detailed apartment page such as in Figure A.1c. These apartment listings vary in their appearance, but the vast majority includes a standardized table, adding further information. Among these are the year of construction or last modernization, self-reported quality of the apartment itself and its equipment, ancillary cost, as well as heating and energy type and cost. In addition, landlords may enter text descriptions, to provide further details about the apartment, its equipment or the location.

The data was collected with a web scraper, which allows to download a current cross-section of all apartment postings on IS24. This was done in 5 waves: November 2017, January 2018, March 2018, October 2018 and November 2018. After removing duplicates, the sample comprises roughly 450,000 apartment listings. We augment our apartment data with data on several area-specific characteristics, because previous research has shown that these location characteristics are important rent-price determinants.[10] These additional variables are measured at the county-level (in German: “Landkreise & kreisfreie Städte”) and were obtained from the regional database of the German Statistics Office.[11] Furthermore, a measure of local market conditions — capturing supply and demand for housing — was gathered from a study by the European research institute Prognos (Koch et al. 2017).

[10]See for example the review of Sirmans et al. (2005).

[11]On January 16, 2018 the data was obtained from: `https://www.regionalstatistik.de/genesis/online`.

Table 1.1 *Summary statistics*

Variable	Observations (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Basic rent*	329,049	538.14	280.29	64.84	3,800.00
Basic rent/m^2	329,049	8.58	3.84	2.00	63.64
Total rent*	329,049	689.90	312.95	108.39	4,978.00
Round rent price (dummy)	329,049	0.18	0.38	0.00	1.00
Charm rent price (dummy)	329,049	0.13	0.34	0.00	1.00
Other rent price (dummy)	329,049	0.69	0.46	0.00	1.00
Apartment size (m^2)	329,049	63.35	18.19	20.00	100.99
Integer apartment size	329,049	0.53	0.50	0.00	1.00
Apartment size mult. $5m^2$	329,049	0.18	0.39	0.00	1.00
Number rooms	329,049	2.38	0.79	1.00	8.00
Floor level	268,933	2.02	1.49	0.00	9.00
Age	329,049	34.86	34.82	0.00	199.00
Apartment condition	329,049		— Categorical —		
Apartment type	329,049		— Categorical —		
Equipment quality	329,049		— Categorical —		
Number photos	135,463	9.04	5.65	1.00	79.00
Number words	136,881	205.45	133.58	5.00	1,848.00
Landlord size**	8,036	17.20	89.11	1.00	5,189.00
GDP p.c.	329,049	39.47	16.54	15.35	131.57
HH income p.c.	329,049	20.72	2.73	16.27	35.66
Population/area	329,049	1,340.94	1,261.80	36.27	4,668.11
Apartments/population	329,049	9.43	8.68	0.00	36.48
Market tightness	329,049	3.49	1.72	1.00	6.00
East Germany (dummy)	329,049	0.38	0.49	0.00	1.00
State	329,049		— Categorical —		

Notes: This table shows summary statistics of key variables. Location-specific variables are measured at the district level. Population/area indicates a district's population per km^2. Apartments/population is measured as number of apartment ads per 1,000 people.

* Basic rent refers to the German term "Kaltmiete" (cold rent) and excludes additional costs such as for heating, water and waste disposal. Total rent refers to the German term "Warmmiete" (warm rent) and includes the above mentioned additional costs.

** Our working sample comprises 8,036 landlords/real-estate agencies, which we can identify by their names. We define landlord size as the number of apartment offerings we observe from the same landlord.

1.2.2 Sample composition

Table 1.1 reports summary statistics for our working sample. To obtain this working sample, we performed the following steps. First, we truncated the raw data to apartments ranging from $20m^2$ to $100m^2$. The reason for the minimum size is to exclude atypical offerings, as apartments below this threshold are mostly rooms in shared apartments or temporary offers, which might not be comparable to whole apartments. The maximum size was chosen due to increasing data scarcity from $100m^2$ on. We further cleaned the data by removing observations with erroneous data. For example, apartments reporting construction years that lie far in the future or apartments with ten rooms but only $30m^2$ apartment size were removed during this step.

To deal with missing values in some of the apartment characteristics we took the following steps. We added a "not reported" category to the three categorical variables, apartment condition, apartment type and equipment quality. The floor level is left out of subsequent analyses as it would reduce the sample by 60,000 apartments and as its exclusion does not change our main results. We keep apartment age in the analysis, as it is likely to influence the probability of observing rounded apartment sizes and rent prices.[12] After removing apartments with missing age, the main working sample comprises roughly 330,000 apartments, which we use for the analysis in Section 1.4.2.

We observe landlord identities, the number of words (in the apartment description) and the number of photos only in a subset of roughly 130,000 observations.[13] This subsample comprises 8,036 landlords which on average offer 17 apartments. As we use landlord identities to construct our landlord-type measure the analyses in Section 1.4.3 and 1.4.4 are based on

[12]Age could affect the rent price as both very new and rather old (but renovated) apartments might be in high demand. Further, it could be the case that the likelihood of not knowing the exact size of the apartment increases with its age, and thus the probability of apartment-size rounding.

[13]Our initial version of the web scraper did not collect these information. After updating the web scraper we obtained these information in the last two sampling waves in October 2018 and November 2018.

this subsample.

A geographic distribution of apartment ads is illustrated in Figure A.2. The map illustrates the number of apartment listings in all German administrative districts. Darker shades indicate a higher number of listings. Naturally, most apartments are located in and around urban centers like Berlin, Munich, Frankfurt or the Rhein-Rhur area.

1.3 Clustering around salient numbers

In this section we graphically document clustering patterns in rent prices and apartment-size measures. Further, we show how these patterns are correlated across the two dimensions.

Clustering in rent prices

Figure 1.1a shows the distribution of rent prices, capped at €1,000 for improved readability.[14] Columns illustrate the actual number of apartments in €10 bins while the red line shows an illustrative counterfactual distribution in the absence of any spikes.[15] This allows to visualize and quantify the excess mass of apartments at bunching regions. In general, the distribution is heavily right-skewed with a median rent price of €469. The distribution of rent prices is highly discontinuous and two distinct cluster patterns exist. On the one hand, a large number of rent prices is rounded to multiples of €50, at which we observe 20% more apartments than the counterfactual distribution would predict. On the other hand, spikes occur just below multiples of €100, with a relative excess mass of 34%. These observations are in line with previous findings on house prices, see for example Chava and Yao (2017) or Meng (2019).

[14]Above this cap of €1,000 only few observations are in the data but the clustering patterns can still be observed.

[15]We employ a simplified version of an approach which was first proposed by Chetty et al. (2009). The counterfactual distribution is estimated by fitting a high-order Chebyshev polynomial to the actual distribution, excluding the bunching regions. The estimates are then uniformly shifted upwards, such that the area beneath equals the one under the actual distribution.

The observed pattern clearly suggests the presence of two prevalent price formats: *Round rent prices* at multiples of €50 and *charm rent prices*, which we define as prices ending between €90.00 and €99.99. The latter are just marginally smaller than their neighboring round counterparts, but keep the next digit to the left smaller by one unit. We categorize the remaining rent prices, which are neither round nor charm, as "unformatted".

Because rent prices are non-negotiated and set by landlords[16], we argue that landlords reveal their type — charmer or rounder — by their choice of price formats. While rounding rent prices could result out of convenience or out of a lack of experience (Repetto and Solís 2019), setting a charm rent price indicates more sophisticated and strategic behavior. Using charm prices landlords seem to exploit left-digit bias, in order to make their apartments appear cheaper than they actually are. Our interpretations are backed by previous findings that higher educated and better performing landlords often choose charm house prices and that this strategy leads to better market outcomes (Chava and Yao 2017; Repetto and Solís 2019).

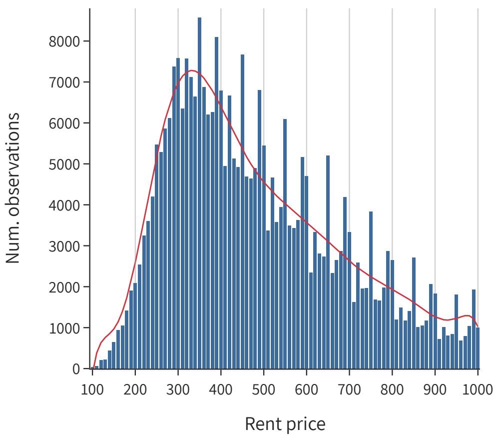
Clustering in apartment sizes

Although apartment size is measurable, rounding behavior can still be expected in our data of reported size measures. German legislation accepts a certain degree of imprecision for apartments, allowing the reported size to deviate up to 10% from the true size. Figure 1.1b shows the total number of apartments per $1m^2$ bin. These numbers are illustrated in stacked columns, where gray columns indicate non-integer (e.g. $50.67m^2$) and blue columns indicate integer (e.g. $50.00m^2$) size measures. The distribution is roughly symmetric around $60m^2$, albeit slightly right skewed. It is immediately visible that the distribution is highly discontinuous with large spikes at focal apartment size measures. These spikes emerge at multiples of $10m^2$ and to a slightly lesser extent $5m^2$, which we both call *round size measures*. Again, we simulate a counterfactual distribution in the absence of spikes (red line), which allows to

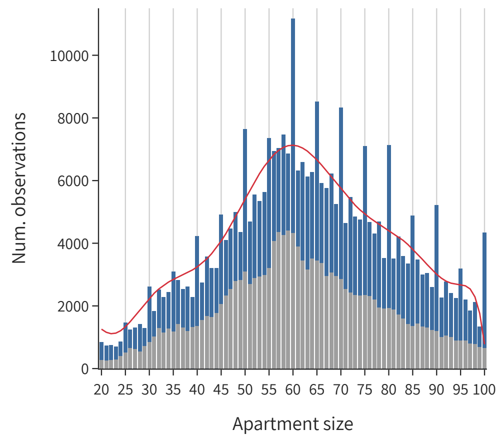
[16] In recent years a legal cap on rent price increases ("Mietpreisbremse") was introduced. However, at the time the data was collected, this cap is only binding for a very small number of apartments.

Figure 1.1 *Clustering patterns: rent price & apartment size*

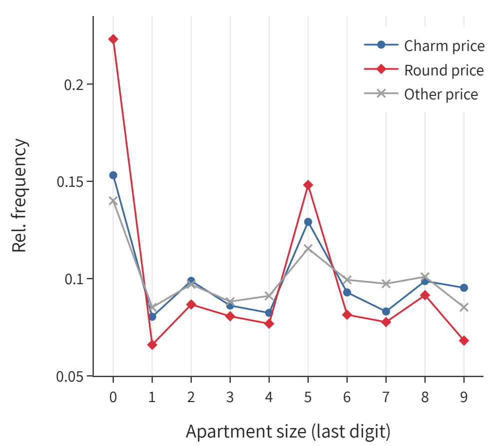
(a) *Basic rent*



(b) *Apartment size*



(c) *Apartment size and rent price formats*



Notes: Plotted are distributions of apartment size and rent prices. Panel (a) shows the total number of apartments in $1m^2$ bins. Rent-price distributions are shown in panel (b) which plots the number of apartments in €10 bins. The red lines in (a) and (b) indicate adjusted counterfactual distributions to better emphasize the magnitudes of spikes. In (a) the columns are stacked and indicate integer (blue, e.g. $55.00m^2$) as well as non-integer (gray, e.g. $55.64m^2$) apartment size indications. To facilitate the readability of panel (b) it is capped at €1,000, but note that above the patterns remain highly similar. Panel (c) shows relative frequencies of last digits in size measures, for different rent price formats. For example, the leftmost red marker in panel (c) indicates that of all round-priced apartments, 22% have a reported apartment size that ends on a 0.

quantify the excess mass at multiples of $5m^2$. At $60m^2$ the absolute excess mass is largest, with about 4,000 (55.54%) more apartments than the counterfactual distribution would predict. The largest relative deviation is at $90m^2$ with 73.10% (2,200) more apartments than predicted.

It is not immediately clear whether landlords round for non-strategic or strategic reasons.[17] The latter case would lead to manipulative sorting and consequently pose a threat to our identification strategy in Section 1.4. Strategic landlords could exaggerate apartment sizes to make them appear more attractive, with focal numbers representing a psychological cap for the degree landlords are willing to overstate size measures. If this was the case, the excess mass at salient numbers would be drawn from the left rather than the right side of the focal value. In Figure 1.1b this would lead to a larger lack of mass in adjacent bins left of multiples of $5m^2$ compared to bins on the right side. Using the counterfactual distribution, we compute this lack of mass and on average it is equal on both sides, with 19% fewer apartments than predicted to the left and 20% fewer apartments to the right of round size measures.[18] Although we cannot test these observations statistically and we cannot rule out manipulative sorting outright, we argue that our observations are more in line with convenience rounding. Nevertheless, we address potential issues due to manipulative sorting by explicitly accounting for size-rounding in all our regressions in Section 1.4.

Correlation of clustering patterns

Having established considerable clustering in both numeric dimensions, Figure 1.1c illustrates how these patterns are related. It shows the distribution of last digits of apartment-

[17]Note, that an alternative explanation for the observed patterns could be that apartments were originally built to have salient apartment-size measures. However, the distribution of non-integer apartment sizes is very smooth, without spikes or drops around focal numbers. Assuming it to be highly implausible that landlords falsely report a non-integer apartment size, this part of the distribution likely reflects actual apartment sizes as constructed. Consequently, it is more plausible that the observed clustering is due to misreporting.

[18]These numbers are calculated based on one bin to the left and one to the right of multiples of $5m^2$. For example, at $50m^2$ we compare the predicted relative lack of apartments in the $49m^2$ and $51m^2$ bins. Considering two bins on each side, the average relative lack off mass amounts to 12% on the left and 13% on the right of round size measures.

size measures, separately for the three price formats defined above: Round, charm and unformatted. Overall, the clustering exists regardless of the price-setting strategy, with all three lines showing spikes at last digits of zero and five. However, the spikes are rather small when rent prices are unformatted (gray line). In comparison, the spikes are slightly larger when rent prices are charm (blue line) and significantly larger when rent prices are round (red line). Of those apartments with a round rent-price format, almost 40% also report a round size measure ending on either zero or five. This share amounts to roughly 30% and 26% for charm and unformatted prices, respectively.

Figure 1.1c clearly reveals that the rounding behavior of landlords is highly correlated across dimensions. This is an important observation as it supports our view that landlord types — charmers and rounders — exist and that we can distinguish them by the way they report numeric measures. We investigate these types and their price-setting behavior more thoroughly in the following sections.

1.4 Empirical analysis

We now focus our attention on the apartment-size dimension and analyze rent-price discontinuities at left-digit thresholds of size measures. First, we describe our empirical strategy in Section 1.4.1 and provide full-sample estimates of discontinuities and a split-sample analysis in Section 1.4.2. Then we analyze heterogeneous effects across landlord types (Section 1.4.3) as well as market tightness levels (Section 1.4.4).

1.4.1 Empirical strategy

In order to identify rent-price discontinuities at salient measures of apartment size, we combine a hedonic pricing model with a regression discontinuity design.[19] This leads to the

[19]This strategy was originally proposed in Lacetera et al. (2012).

specification in Equation (1.1)

$$RentPrice_i = \alpha + \beta_1 LeftDigit_i + \beta_2 Round_i + f(Size_i) + \mathbf{X_i}'\boldsymbol{\theta} + \varepsilon_i. \tag{1.1}$$

$RentPrice_i$ is the basic rent price[20] of an apartment and $\mathbf{X_i'}$ is a vector of apartment and location characteristics. $f(Size_i)$ is a flexible function of apartment size and captures the continuous relationship between apartment size and rent prices. $LeftDigit_i$ is a dummy variable indicating whether an apartment's size measure surpasses a left-digit threshold and $Round_i$ indicates whether size is reported as an exact multiple of $5m^2$.

The idea behind this identification strategy is that the leftmost digit of apartment size and whether the size is a round number should not matter, once the much finer apartment size information is accounted for. However, if individuals focus on the leftmost digit of apartment size, the perceived value of apartments will increase discontinuously at left-digit thresholds. This can lead to discontinuous rent-price jumps at left-digit thresholds of size (positive β_1). $Round_i$ explicitly controls for the size-rounding pattern we discovered in Section 1.3 to ensure our estimates at left-digit thresholds are not driven by this pattern. In addition, it allows us to test whether size-rounding affects rent prices in a discontinuous way. However, as we could not identify strategic motives for rounding, we are agnostic about the direction of any effects at round thresholds and let the data speak.

We use two different specifications of $f(Size_i)$

$$f(Size_i) = \gamma_0 Size_i + \gamma_1 Size_i^2 + \sum_{j=3}^{9}\left[\gamma_{2j}w_j + \gamma_{3j}\left(Size_i * w_j\right) + \gamma_{4j}\left(Size_i^2 * w_j\right)\right] \tag{1.2a}$$

$$f(Size_i) = \gamma_0 Size_i^1 + \gamma_1 Size_i^2 + \gamma_2 Size_i^3 + \gamma_3 Size_i^4. \tag{1.2b}$$

In our main analysis, in Sections 1.4.3 and 1.4.4, we employ a local-quadratic approach, which is shown in Equation (1.2a). We divide the range of size measures into windows (w_i) of

[20] In Section 1.5 we use total rent and show that our results are robust.

size $10m^2$ centered around each left-digit threshold.[21] For each window we include threshold-specific intercepts γ_{2j} to capture the level effects of the average size differences across windows. $Size_i$ is centered around the left-digit thresholds, added in linear and quadratic form, and we allow the effect of $Size_i$ to vary across windows.[22] This local-estimation approach allows us to include one single indicator for all left-digit thresholds and another one for all round size measures. Consequently, β_1 and β_2 estimate average discontinuities across the whole range of size measures. The advantage of this approach is that we are able to interact both indicators with other covariates to study effect heterogeneities, without overcomplicating, and potentially oversaturating, our model.

We also employ an alternative approach in which we define $f(size_i)$ globally, choosing a 4^{th}-order polynomial as shown in Equation (1.2b).[23] In Section 1.4.2 we follow this strategy to estimate discontinuities at left-digit thresholds individually and to visualize the discontinuities. Further, this approach provides additional robustness to our main results.

The control vector $\mathbf{X_i'}$ contains apartment and location characteristics. Apartment-specific characteristics include the number of rooms, age, age squared, as well as the three categorical variables equipment quality, apartment type and apartment condition.[24] Apartment age is added in linear and quadratic form to account for potentially non-linear price effects, as both brand new and historic apartments are typically in high demand. In Sections 1.4.3 and 1.4.4 the number of words and photos as well as indicators for various amenities like balconies,

[21]For example, the window around the $50m^2$-threshold ranges from $45.00m^2$ to $54.99m^2$.

[22]This specific functional form was chosen based on the Akaike Information-Criterion. In Section 1.5 we show that our results are robust to a set of alternative specifications.

[23]Again, in Section 1.5 we show that our results are robust to a set of alternative specifications. Choosing higher-order polynomials would only marginally improve the Akaike Information-Criterion. Further, the disadvantages of using higher-order polynomials has recently been discussed in the literature (Gelman and Imbens 2018).

[24]We further add an indicator for apartment size measures that are rounded to an integer and interact this indicator with apartment size. We observe in the data that apartments are more expensive if the size is reported as an integer (e.g. $50m^2$) as opposed to a non-integer (e.g. $50.25m^2$). However, this rent price difference fades for larger apartments. Including the integer-rounding dummy enables to distinguish a rounding-to-multiples effect from an integer-rounding effect.

basements or lifts are included as additional controls.[25] Among the location characteristics are household income per capita, population/km^2, the number of apartments per 1,000 inhabitants, and state indicators. Lastly, we add an indicator for the five sampling waves to mitigate concerns about time-of-the-year-specific market conditions that might bias the results.

1.4.2 Rent-price discontinuities at apartment-size thresholds

This section provides results for the entire sample. Further, we document heterogeneity across rent-price formats in a split-sample analysis, in which we analyze rent-price discontinuities for charm, round and unformatted rent prices separately. The RD-type setting of the empirical analysis makes it possible to visualize rent-price discontinuities. Figure 1.2 shows average adjusted residuals resulting from a regression of rent prices on an apartment-size polynomial and control variables.[26] To show the trend and replicate real rent-price patterns, the polynomial is added back and rent prices are averaged within $1m^2$ bins. The residuals are plotted for the full sample as well as for charm round and unformatted rent prices separately. In order to obtain parametric estimates of price discontinuities, Table 1.2 shows results from estimating Equation (1.1) in combination with Equation (1.2b). We include a full set of controls, a 4^{th}-order polynomial of apartment size and estimate the discontinuities at left-digit thresholds individually.

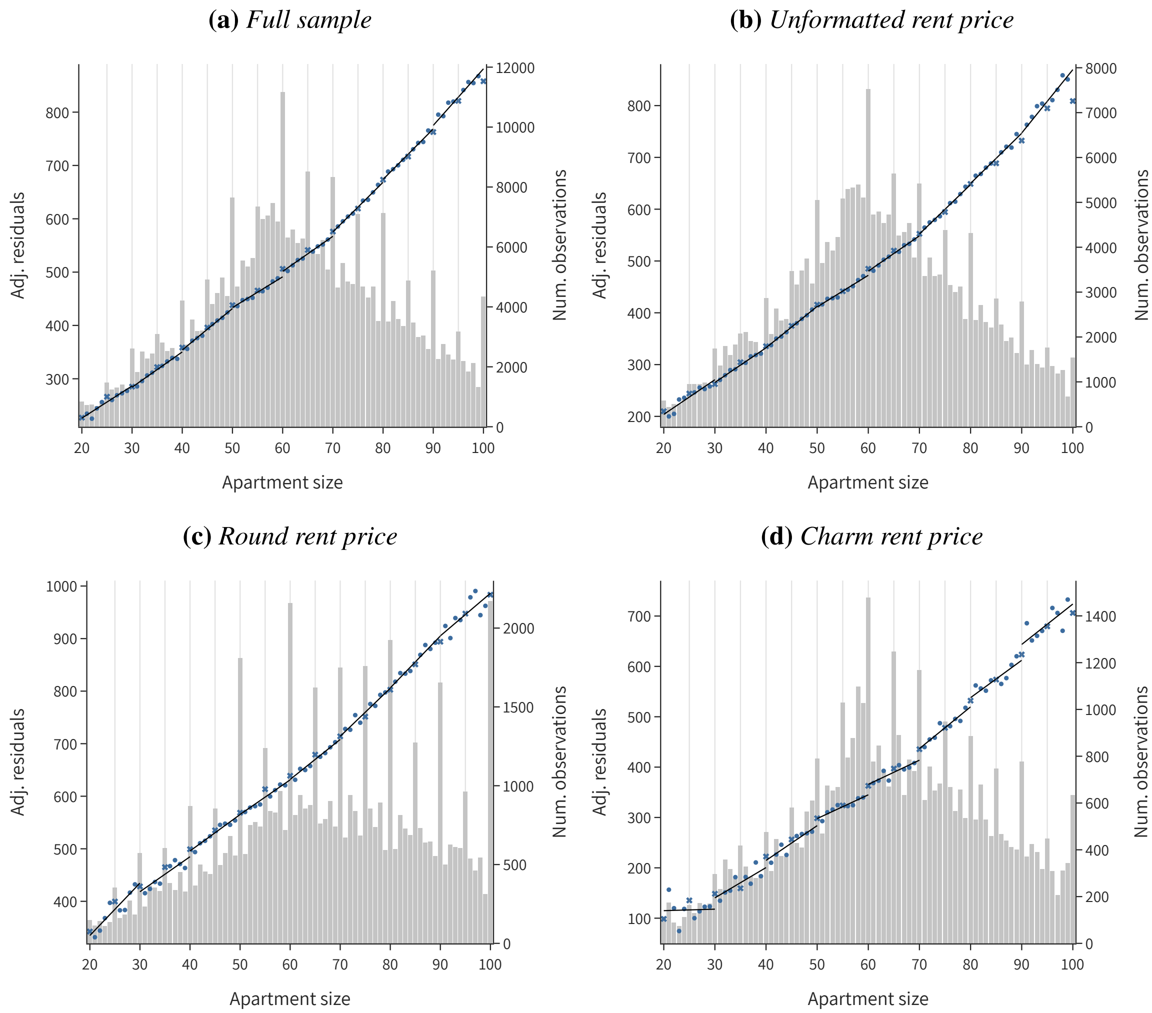
Full sample results

Figure 1.2a plots residuals for the full sample. As one would expect, apartment size affects rent prices positively in an almost linear way. To make price jumps at size thresholds more visible, we estimate local linear trends between the thresholds, illustrated by the black lines.

[25] These controls are only available for the same subsample in which landlord identities are known. We therefore exclude them in Section 1.4.2, in which we do not use landlord identities, to analyze the full working sample.

[26] Figure A.3 in the Appendix shows raw data. The patterns in the raw data are similar and even more pronounced than in Figure 1.2.

Figure 1.2 *Average adjusted rent-price residuals*



Notes: Plotted are average adjusted residuals from regressions of basic rent on a 4^{th}-order apartment size polynomial as well as apartment and location specific controls. In order to replicate real rent price developments the residuals are adjusted by adding back the polynomial. Further, residuals are averaged within $1m^2$ bins. Solid lines represent linear fits between $10m^2$ multiples. Gray columns indicate the total amount of apartments per $1m^2$ bin in the respective subsample. Note that the plots show average predicted values net of the price effects of control variables and thus do not coincide with real average rent prices.

Lines that create a broken pattern at left-digit thresholds rather than smoothly transitioning into each other imply rent-price discontinuities at these thresholds. From $30m^2$ onwards clear signs of positive rent-price jumps are visible at all thresholds and from $60m^2$ onwards they become larger. This pattern strongly suggests the role of inattentive behavior, in which the left-most digit of apartment size measures is given too much weight compared to the remaining digits. To show potential price effects of apartment-size-rounding, we mark multiples of $5m^2$ by crosses. Figure 1.2a reveals that many of these crosses lie above local average price-levels, for example at $60m^2$ and $65m^2$, suggesting that apartment-size rounding could positively affect rent prices.

Parametric estimates of price discontinuities are shown in column (1) of Table 1.2 and confirm the observations made above. With one exception at $30m^2$, all left-digit-threshold coefficients are positive and statistically significant. The average price jump amounts to €5.16. If we assume a linear relationship between rent prices and apartment size, we are able to approximate an inattention parameter in the spirit of Lacetera et al. (2012). The average price increase between left-digit thresholds is €70.10, which implies an inattention parameter of $\frac{5.16}{70.10} = 0.074$. This suggests that 7.4% of the rent-price increase between two $10m^2$-thresholds materializes right at the threshold, which is sizeable but smaller than inattention parameters in other markets.[27] Further, Table 1.2 column (1) confirms that apartment-size-rounding is associated with higher rent prices of on average €4.41. This is a puzzling result and it is not clear whether this discontinuity reflects an actual size-rounding premium or is driven by a sorting mechanism. In Sections 1.4.3 and 1.4.4 we find evidence that is more in line with a sorting mechanism in which profit-seeking landlords are more likely to report rounded size measures.

The full-sample results highlight that inattention plays a role in the German housing market and therefore confirms previous findings in the literature. However, our main research

[27] See DellaVigna (2009) for an overview.

question is whether the discontinuities are driven by landlord behavior and whether they vary across landlord types. As argued in Section 1.3, landlords reveal their type — charmer or rounder — by their choice of rent prices. We therefore split our sample across price formats to test whether the choice of rent prices affects the price-discontinuities we document in the size dimension.

Table 1.2 *Rent-price discontinuities: rent-price format*

	Dependent variable: Basic rent			
	Full sample	Unformatted	Round	Charm
	(1)	(2)	(3)	(4)
$\geq 30m^2$	−10.427***	−11.930***	−19.649**	14.018
	(2.715)	(2.777)	(8.332)	(8.658)
$\geq 40m^2$	11.378***	7.781***	10.185*	30.452***
	(1.720)	(1.786)	(5.383)	(5.330)
$\geq 50m^2$	5.638***	4.220***	3.492	15.143***
	(1.335)	(1.374)	(4.477)	(4.095)
$\geq 60m^2$	2.853**	2.580**	−2.452	9.156***
	(1.134)	(1.197)	(3.849)	(3.209)
$\geq 70m^2$	7.329***	4.213**	−0.415	22.589***
	(1.518)	(1.652)	(4.503)	(4.222)
$\geq 80m^2$	12.403***	10.880***	−0.865	31.235***
	(1.948)	(2.165)	(5.303)	(5.308)
$\geq 90m^2$	6.956**	7.023**	−4.017	28.960***
	(2.817)	(3.396)	(6.046)	(7.423)
Mult. $5m^2$	4.412***	−0.823	5.819***	3.487
	(0.851)	(0.926)	(2.033)	(2.400)
Adj. R^2	0.741	0.752	0.719	0.739
Observations	329,049	227,301	58,529	43,219
% Integer	55.61	50.07	73.83	58.52
% Mult. $5m^2$	20.07	16.23	33.28	21.27

Notes: The dependent variable is basic rent and all regression include a 4^{th}-order apartment size polynomial as well as apartment and location specific controls. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%. The two bottom rows indicate the shares of apartments with integer and multiple-of-$5m^2$ size measures.

Heterogeneity across rent-price formats

Figure 1.2b plots price residuals for the subsample of unformatted rent prices. The patterns in this figure are very similar to those in the full sample, but slightly less pronounced. Parametric estimates in Table 1.2 column (2) confirm this observation, with an average price-jump of €3.54 at left-digit thresholds, leading to an inattention parameter of 0.05. The coefficient for size-rounding becomes statistically insignificant. These observations are not surprising, as unformatted rent prices do not indicate any specific landlord type.

Figure 1.2c plots price residuals for round rent prices. The linear fits between the thresholds (solid black lines) create an almost perfectly smooth upward trend. Thus, there is no sign of discontinuous shifts in prices at left-digit thresholds. In contrast, many of the residuals at exact multiples of $5m^2$ are slightly above the general trend. Table 1.2 column (3) shows parametric estimates of the discontinuities. Almost all of the threshold estimates become statistically insignificant, whereas the size-rounding coefficient stays highly significant and increases to €5.82.

In contrast, Figure 1.2d plots price residuals for charm rent prices and reveals a strong pattern of discontinuous price jumps at left-digit thresholds. According to the estimates in column (4) of Table 1.2 the average discontinuity at left-digit thresholds of apartment size amounts to €21.65 and is therefore more than five times as large as in the full sample. This leads to an inattention parameter of 0.309, which is comparable to inattention parameters measured in used-car markets (Englmaier et al. 2018b; Lacetera et al. 2012). Round size-measures are not systematically associated with higher rent-price residuals.

The results in Figures 1.2c and 1.2d reveal that the choice of rent-price formats strongly affects the discontinuities we find in the apartment-size dimension. Landlords who exploit left-digit bias in the price dimension, by choosing a charm rent price, sharply increase rent prices at left-digit thresholds of size measures. In stark contrast, landlords who do not exploit left-digit bias in the price dimension, by choosing a round number, also do not discontinuously increase rent prices at left-digit thresholds of apartment size. This correlation in price-setting behavior across dimensions has two important implications. First, it supports our view that two distinct landlord types exist — charmers and rounders — which can be distinguished by their price-setting behavior. Second, the results strongly imply that price discontinuities at left-digit thresholds of size are the result of exploitive behavior by charmers. Taken together, these findings motivate our approach to define a landlord-type measure in order to more directly quantify the effect of landlord types on rent-price discontinuities.

1.4.3 Heterogeneity across landlord types

We now directly and simultaneously estimate differences in the price-setting behavior of charmers and rounders. For this, we introduce a landlord type measure which we add to our main specification in Equation (1.1). Our measure of landlord type is defined in Equation (1.3)

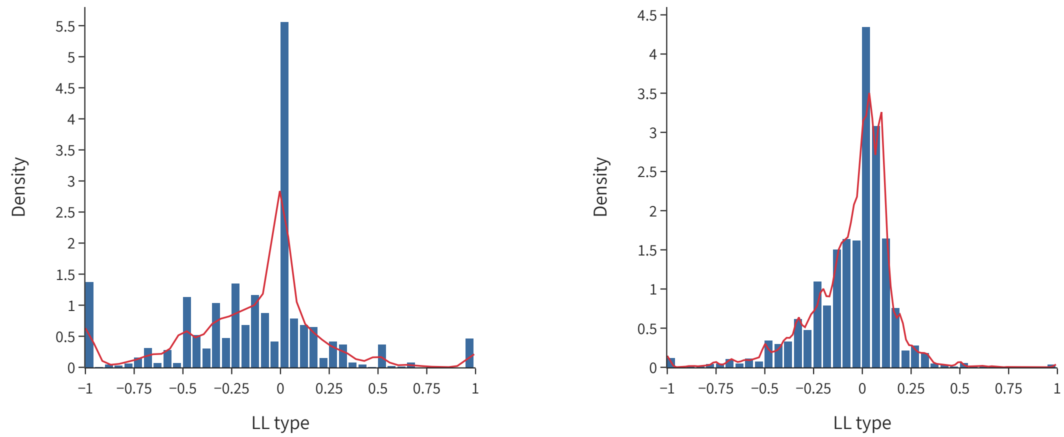
$$LL\,type = \frac{Num.\ charm\ rent\ price\ format - Num.\ round\ rent\ price\ format}{Num.\ total\ offerings}. \tag{1.3}$$

For each landlord it takes the difference between the number of charm and round rent prices and divides it by the total number of this landlord's offerings. Thus, the measure captures the relative preference for charm prices over round prices and ranges from -1 for pure rounders (100% round rent prices) to $+1$ for pure charmers (100% charm rent prices). Note, that a neutral value of zero can either mean that a landlord chose an equal number of charm and round prices or only chose unformatted prices. In our analysis below we also use alternative definitions which calculate the share of charm and round prices individually. This way we do not force both landlord types onto the same scale.

Figure 1.3 shows the distribution of our landlord-type measure. In Figure 1.3a the distribution of landlord type is calculated at the landlord level. Many landlords are of a neutral type (close to zero), but the distribution also indicates a considerable number of both charmers

Figure 1.3 *Landlord types: distributions at apartment & landlord level*

(a) *Landlord level* **(b)** *Apartment level*



Notes: Shown is the distribution of our landlord-type measure at the landlord (a) and apartment level (b). The histograms show the normalized frequency of the measure with a bin size of 0.025 and red lines indicate kernel density estimates. The landlord-type measure takes the difference between the number of charm and round rent prices and divides it by the total number of a landlord's offerings. It ranges from -1 (100% round rent prices) to $+1$ (100% charm rent prices).

(positive values) and rounders (negative values).[28] Figure 1.3b shows the distribution of landlord type at the apartment level, which is more concentrated around zero. The reason for this is that large real-estate agencies are less consistent in their chosen rent-price formats than landlords with smaller portfolios. Nevertheless, we argue that our landlord-type measure offers sufficient variation to distinguish landlords in a meaningful way.

We now test whether landlord types can explain the heterogeneity in rent-price discontinuities in the apartment-size dimension. For this, we add our landlord-type measure to our main specification and interact it with the indicators for left-digit thresholds and size rounding.

[28]The distribution indicates a higher number of rounders than charmers. This is in part driven by our definition of round rent prices as multiples of €50, leading to a smaller distance between two round rent prices than between two charm rent prices. Therefore, most rent prices more likely lie in close proximity to round numbers than charm numbers, creating a stronger incentive to round. For example, a "fair" rent price of €640 is only €10 smaller than the next round number but €50 smaller than the next charm number.

This extended specification is shown in Equation (1.4)

$$\begin{aligned} RentPrice_i = & \beta_1(LeftDigit_i * LLtype_i) + \beta_2(Round_i * LLtype_i) \\ & + \beta_3 LeftDigit_i + \beta_4 Round_i + \beta_5 LLtype_i \\ & + f(Size_i) + \boldsymbol{X_i'\theta} + \varepsilon_i. \end{aligned} \tag{1.4}$$

The main effects (β_3 and β_4) capture the overall rent-price discontinuities irrespective of landlord type and the interaction coefficients (β_1 and β_2) capture whether these discontinuities are more or less pronounced for certain landlord types.

Table 1.3 summarizes the results. Column (1) includes only the main effects and thus estimates overall discontinuities. Both kinds of discontinuities are positive and significant and confirm previous findings. The landlord type measure enters statistically significantly and is negative and sizeable. Because negative values of our landlord-type measure indicate rounders, this implies that for the same apartment rounders demand higher rent prices than charmers or neutral landlords.

Column (2) adds the interaction terms and confirms that landlord types matter for rent-price discontinuities in the apartment-size dimension. The average discontinuity at left-digit thresholds amounts to €5.41, but charmers additionally increase rent prices by up to €19.63, leading to a maximum inattention parameter of $\frac{5.41+19.63}{70.10} = 0.36$. The main effect of size rounding is insignificant, however, according to our estimates, rounders increase prices at round size multiples by up to €20.78.[29]

One could be concerned that these results are driven by either charmers or rounders alone, as our landlord-type measures forces both types onto the same scale. To alleviate these concerns, we use alternative measures of landlord type and analyze the behavior of charmers and rounders individually. Column (3) re-estimates Equation (1.4) using the share of charm rent prices as landlord-type measure. It confirms that charmers increase rent prices at left-digit

[29]Note that negative values of our landlord-type measure indicate rounders. Therefore, a negative coefficient of ($LL\ Type \times Round$) implies that rounders increase rent prices at round apartment-size multiples.

Table 1.3 *Rent-price discontinuities: landlord type*

	Dependent variable: Basic rent			
	(1)	(2)	(3)	(4)
LL Type × Left Digit		19.632***		
		(4.981)		
LL Type × Round		−20.776***		
		(6.145)		
LL %Charm × Left Digit			30.730***	
			(8.678)	
LL %Charm × Round			−33.262***	
			(10.214)	
LL %Round × Left Digit				−15.904***
				(6.123)
LL %Round × Round				21.561***
				(8.112)
Left Digit	4.801***	5.409***	0.528	7.160***
	(1.597)	(1.596)	(1.889)	(1.779)
Round	3.713***	1.837	9.617***	−2.963
	(1.385)	(1.392)	(1.988)	(2.064)
LL Type	−53.301***	−58.269***		
	(2.688)	(3.905)		
Adj. R^2	0.764	0.764	0.763	0.767
Observations	127,965	127,965	127,965	127,965

Notes: Shown are coefficients from local-quadratic estimations using windows of size $10m^2$ centered at multiples of $10m^2$. *Left Digit* indicates whether the size measure includes the larger leftmost digit within each window. *Round* indicates whether the size measure is a multiple of $5m^2$. *LL %Charm* and *LL %Round* measure the share of a landlord's apartments with charm and round rent prices, respectively. *LL Type* is the landlord type measure as constructed in Equation (1.3). It ranges from -1 (100% round prices) to 1 (100% charm prices). Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

thresholds of size measures by up to €30.73, while neutral and rounder landlords do not increase prices at these size thresholds. Further, the positive price-effect of size-rounding is offset for charming landlords. Column (4) re-estimates Equation (1.4) using the share of round rent prices and shows that rounders behave the opposite way. They increase prices at round size-multiples by up to €21.56, but do not discontinuously increase rent prices at left-digit thresholds.

In summary, these results are consistent with our hypothesis that charmers and rounders systematically differ in their price-setting behavior. *Rounders*, who do not exploit left-digit bias in the price dimension, *do not increase* rent prices at left-digit thresholds of size. *Charmers*, who clearly are aware of left-digit bias and exploit it in the price dimension, *increase prices discontinuously* at left-digit thresholds of size. Consequently, the most plausible explanation of our results is that charmers exploit left-digit bias both in the price dimension and the size dimension. This reflects a very sophisticated strategy: Assuming inattentive tenants who focus on the leftmost digit of apartment size, charmers discontinuously increase rent prices at these left-digit thresholds. Exploiting left-digit bias again, charmers are able to (partially) hide rent-price discontinuities by choosing a charm price format.

Rounders, on the other hand, do not exploit left-digit bias in any of the two dimensions, but employ a simpler approach choosing round numbers. Although rounders employ a simpler approach, we find that they in general demand higher prices than other landlords. We interpret this observation as suggestive evidence that rounders are rather careless[30] and are more profit-seeking than others. Why rounders seem to increase prices at multiples of $5m^2$ remains puzzling, but we refrain from interpreting this pattern as evidence for a size-rounding premium. Instead, our results suggest that these price-discontinuities at round size measures

[30]Section 1.3 suggests that landlords round size measures also out of convenience rather than following a sophisticated strategy. Apartment-size measures could be rounded due to a lack of knowledge of exact measures and a lack of motivation to obtain exact measures. Further, previous literature has shown that charm prices perform better than round prices. Therefore we argue that it is also more likely that rounders round rent prices out of convenience rather than strategically.

are driven by a sorting mechanism. In Section 1.3 we show that rounders are more likely to report rounded size measures and they most likely round out of convenience rather than strategic motives to extract a premium. This section, however, shows that rounders in general demand higher prices than charmers and neutral landlords. In other words, the share of profit-seeking rounders is larger at multiples of $5m^2$ than at any other size measure, which likely leads to the observed rent-price discontinuities.

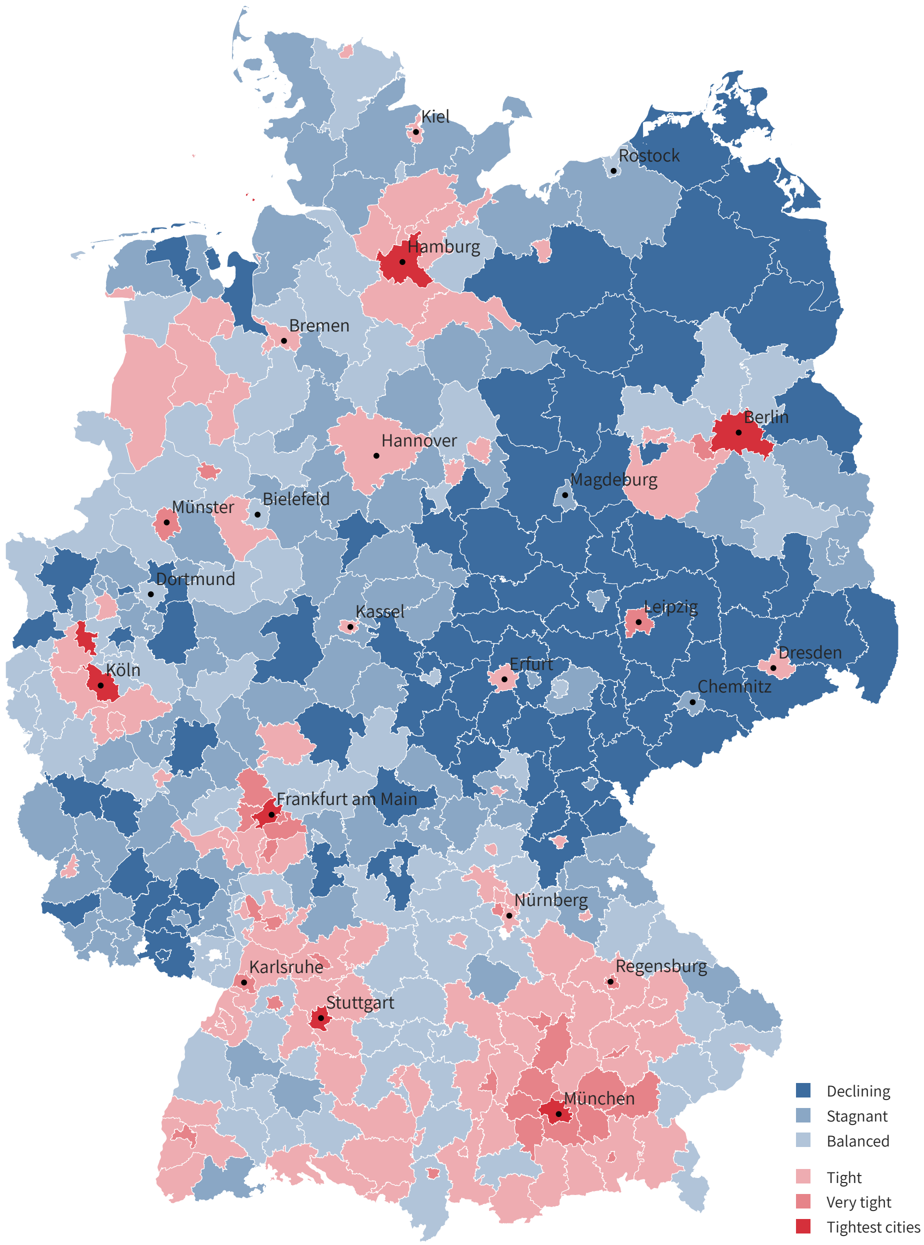
1.4.4 Heterogeneity across market power

In this section we test whether there is more scope for exploitation in tight housing markets. Landlords can take advantage of their market power in tight markets and can more easily extract rent-price premia. Consequently, if the rent-price discontinuities that we find in the previous sections are truly driven by exploitive behavior of landlords, these discontinuities should be more pronounced in tight markets.

In order to assess the situation in local housing markets, we introduce a market-tightness measure, which we take from Koch et al. (2017). The measure is based on developments in supply and demand for housing as well as on other indicators of labor markets and wealth trends.[31] Market tightness is a categorical indicator at the county level and ranges from one to six. Categories one to three indicate stagnant or balanced rental markets, while the higher categories stand for tight markets. For our purpose we reduce the six categories into two groups: relaxed markets and tight markets. Forming two groups assures large enough strata to get meaningful estimates while still capturing differences in market tightness. Figure 1.4 shows a map indicating the market situation across German districts. Tight markets, shown in red, are located mostly in the main metropolitan areas of Germany such as Munich, Hamburg or Berlin. Relaxed markets, shown in blue, comprise mostly rural areas as well as less booming urban areas in the Rhine-Ruhr area.

[31] For a detailed description refer to Koch et al. (2017).

Figure 1.4 *Map of Germany: market situation*



Notes: This map shows the geographic distribution of the market-tightness indicator developed by Koch et al. (2017). The shades indicate the respective market-tightness categories as described in the legend. Relaxed markets are shown in blue and tight markets in red.

To estimate the effect of market tightness on rent-price discontinuity patterns, we add the market-tightness dummy to our main specification and interact it with both discontinuity indicators. Further, we add triple interactions of market tightness, the discontinuity indicators and our landlord-type measure to test whether market power affects landlord types in different ways. The full specification is shown in Equation (1.5)

$$\begin{aligned} RentPrice_i = {} & \beta_1(LeftDigit_i * TightMarket_i) + \beta_2(Round_i * TightMarket_i) \\ & + \beta_3(LeftDigit_i * LLtype_i) + \beta_4(Round_i * LLtype_i) \\ & + \beta_5(LeftDigit_i * LLtype_i * TightMarket_i) + \beta_6(Round_i * LLtype_i * TightMarket_i) \\ & + \beta_7 LeftDigit_i + \beta_8 Round_i + \beta_9 TightMarket_i + \beta_{10} LLtype_i \\ & + \beta_{11}(LLtype_i * TightMarket_i) \\ & + f(Size_i) + \boldsymbol{X_i'\theta} + \varepsilon_i. \end{aligned} \tag{1.5}$$

We estimate different versions of Equation (1.5) and the results are shown in Table 1.4. To get a first idea of how market tightness affects rent-price discontinuities, they are estimated separately for relaxed and tight markets in columns (1) and (2), respectively. While the coefficients for both discontinuities are small (size-rounding) or even insignificant (left-digit thresholds) in relaxed markets, they are positive and much larger in tight markets. These results imply that the rent-price discontinuities are almost entirely driven by the price-setting behavior of landlords in tight markets. Column (3) adds the market-tightness dummy and confirms the results of the split-sample analysis. Not surprisingly, rent prices are on average €38.16 higher in tight markets, which means that landlords can realize higher rent prices if they have high market power. The main effect of left-digit thresholds is statistically insignificant, but in tight markets rent prices discontinuously jump by €8.35 at thresholds. Similarly, the size-rounding effect is estimated negatively in relaxed markets, but in tight markets turns positive and amounts to €18.35 (€23.47−€5.12). Taken together, the results in columns (1) through (3) show that rent-price discontinuities are more pronounced if landlords have higher market power and are thus consistent with our interpretation of active bias exploitation.

Table 1.4 *Rent-price discontinuities: market power*

	Dependent variable: Basic rent				
	(1)	(2)	(3)	(4)	(5)
TightMkt × Left Digit			8.348***	5.929***	5.999***
			(0.961)	(1.338)	(1.327)
TightMkt × Round			23.470***	10.289***	9.322***
			(1.417)	(1.970)	(2.022)
LL Type × Left Digit				19.864***	18.803***
				(4.964)	(5.631)
LL Type × Round				−22.712***	−15.817**
				(6.112)	(6.969)
LL Type × TightMkt × LD					1.654
					(9.032)
LL Type × TightMkt × R					−11.056
					(11.206)
Left Digit (LD)	−0.235	4.647**	−1.610	2.619*	2.573
	(0.962)	(1.878)	(1.085)	(1.590)	(1.576)
Round (R)	1.463*	13.694***	−5.120***	−4.629***	−4.077***
	(0.868)	(1.582)	(0.944)	(1.486)	(1.466)
TightMkt			38.158***	30.997***	31.055***
			(0.877)	(1.252)	(1.246)
LL Type				−5.859	−7.135*
				(4.121)	(4.102)
LL Type × TightMkt				−82.100***	−80.135***
				(4.584)	(6.721)
Tight markets		**X**	**X**	**X**	**X**
Relaxed markets	**X**		**X**	**X**	**X**
Adj. R^2	0.706	0.683	0.736	0.768	0.768
Observations	150,122	160,040	310,162	127,965	127,965

Notes: Shown are coefficients from local-quadratic estimations using windows of size $10m^2$ centered at multiples of $10m^2$. *Left Digit (LD)* indicates whether the size measure includes the larger leftmost digit within each window. *Round (R)* indicates whether the size measure is a multiple of $5m^2$. *LL Type* is the landlord-type measure as defined in Equation (1.3). *TightMkt* indicates tight local markets. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Column (4) adds back our landlord-type indicator and its interactions with the discontinuity measures. This way we are able to mitigate concerns that our results merely capture sorting of different landlord types into different market conditions. The effects of market tightness on the discontinuities are now smaller in size but still positive and highly significant. At the same time, the coefficients of the interactions of both discontinuity indicators with landlord type are almost identical to previous results in Table 1.3. These results contradict a sorting mechanism and instead imply that both market tightness and landlord type affect rent-price discontinuities independently of each other.

An interesting side result can be observed for the interaction of landlord type and market tightness, which enters with high statistical significance and strongly negative (€-82.10). Our previous result, that rounders in general demand higher rent prices than charmers and neutral landlords, is therefore entirely driven by differences in price-setting behavior in tight markets. This is consistent with our interpretation that rounders aim to extract high profits and they are able to achieve this in tight markets. Again this underlines our preferred interpretation that the size-rounding effect is mainly driven by a sorting mechanism rather than reflecting a true size-rounding premium. Rounders demand higher rent prices than other landlords, especially in tight markets, and more often round size-measures to multiples of $5m^2$, which could lead to the observed patterns.

Finally, in column (5) we add triple interactions of market tightness, landlord type and both discontinuity measures. Both coefficients of these triple interactions are not statistically significant, whereas the coefficients of the double interactions remain largely intact. Our interpretation of these results is that all landlords, irrespective of their type, exploit behavioral biases to a larger degree if they have high market power.

1.5 Robustness

In this section we perform a range of robustness tests and document that our results are valid. Table A.1 repeats our main specifications for alternative dependent variables, namely total rent and logged basic rent. The estimates for total rent are very close to those for basic rent. Thus, the discontinuities in basic rent prices do not appear to be offset by utilities, but are fully borne by tenants. Results for logged basic rent can be interpreted as relative changes and confirm the heterogeneity across landlord types. Regarding the heterogeneity across market conditions, the results are a bit mixed but not contradicting our main findings. In tight markets, the rounding premium still increases relative to basic rents, but the discontinuity at left-digit thresholds in relative terms becomes slightly smaller. This means that although discontinuities increase in absolute terms, they do not increase as much as the general rent price levels.

In Table A.2 we test alternative functional forms of our local-estimation approach. Columns (1) to (3) test different versions of local-linear approaches and columns (4) and (5) test local-quadratic versions. The results are virtually unchanged irrespective of the functional form of $f(Size_i)$. Based on the Akaike Information-Criterion our preferred specification is the local-quadratic approach of column (5). Table A.3 shows results for alternative specifications of the size polynomial, which we use in Section 1.4.2. Although the discontinuity estimates are slightly instable for the linear and quadratic specification they are robust for size-polynomials of degree three and above, which according to the Akaike Information-Criterion provide a better fit.

In Table A.4 we repeat our main analysis with placebo thresholds after converting apartment size to square yards. Columns (1) to (3) include only placebo-thresholds while columns (4) to (6) show estimates for a horse-race specification in which we add back square-meter thresholds. Placebo left-digit thresholds seem to cause negative rent-price jumps in relaxed markets and placebo round measures are associated with positive price jumps, again only in relaxed markets. The effect heterogeneity across landlord types is less pronounced and only moderately statistically significant. These results are less systematic than those of the actual thresholds and (partly) at odds with well established findings of previous literature. We suspect that using converted size measures and placebo thresholds misspecifies the actual relationship between rent prices and apartment size and thus leads to these odd results. In the horse-race specifications, estimates of the actual square-meter thresholds remain virtually unchanged with the exception of $TightMkt \times Left\ Digit$. Overall, we argue that the placebo test supports the validity of our main results.

To provide graphical evidence that the discontinuities are not driven by changes in the market composition around apartment-size thresholds, Figure A.4 shows observable heterogeneity of control variables at salient size measures. As none of these figures indicate systematic variation at left-digit thresholds of apartment size, we argue that the discontinuity estimates are not driven by changes in the market composition. However, some control variables appear to be correlated with apartment-size rounding.[32] For example, apartments with round size measures are associated with a higher number of photos, fewer words in text descriptions and are more likely to be located in West Germany. These observations constitute another reason why we refrain from interpreting rent-price discontinuities at round size measures as size-rounding premium, but rather consider them as the result of a sorting mechanism.

In Table A.6 we estimate regressions of landlord type (and price formats) on control variables, but most coefficients are very small. Thus, Table A.6 does not uncover any systematic variation across landlord types additional to those we already know — rounding of apartment size and market tightness. This suggests that our landlord-type measure captures

[32] Since we control for size-rounding this does not pose a threat to the identification of price discontinuities at left-digit thresholds.

actual differences in landlords' price-setting behavior and does not simply pick up systematic variation of apartment or location features. For the interested reader, Table A.7 in the Appendix shows the coefficient estimates of the hedonic-pricing model, which we use to control for apartment- and location-specific characteristics. There are no surprising results, with most coefficients being statistically significant and in line with common expectations.

1.6 Conclusion

We study the price-setting behavior of landlords in the German housing market and find empirical evidence that (some) landlords exploit limited attention of tenants. Based on their choice of rent prices, we categorize landlords into charmers and rounders, and show that their pricing behavior with respect to apartment-size measures differs systematically. Charmers increase rent prices discontinuously at left-digit thresholds of size measures. While this behavior is more pronounced in tight markets, it also occurs in relaxed market situations. Rounders on the other hand do not increase prices at left-digit thresholds of size measures. Instead, they demand higher rent prices in general, but only in tight markets.

Our results show that landlords follow different price-setting strategies. Although we cannot fully proof it, charmers appear to exploit limited attention of tenants to their own advantage, which constitutes a rather sophisticated approach. They strategically set rent prices just below multiples of €100 to make their apartments appear cheaper than they actually are and, in contrast to rounders, they increase rent prices at left-digit thresholds in a discontinuous way. This is a cunning way of increasing rent prices, because tenants who rely on left-digit heuristics are unlikely to detect these price discontinuities. Although rounders are profit-seeking as well, they follow a simpler approach, not exploiting left-digit bias in any dimension. They round rent prices, not taking advantage of left-digit bias[33], and they

[33] If landlords round to multiples of €100, price-charming could be employed at virtually no cost (€0.01).

tend to round apartment-size measures, more likely out of convenience than strategic reasons. Further, they simply increase rent prices for all apartments, not aiming to exploit limited attention to size measures.

Our interpretation of charmers and rounders is in line with previous research which found that price charming in housing markets is positively correlated with education and experience of real estate agents (Chava and Yao 2017; Repetto and Solís 2019). However, we are limited by our data to further characterize charmers and rounders.[34] Investigating which characteristics are associated with landlord types and what leads landlords to become either charmers or rounders is an interesting topic for future research.

Another limitation of our data is that we cannot measure how successful the different price-setting strategies are and in particular whether the arguably more cunning approach of charmers is superior in terms of market outcomes.[35] In the German housing market, success could be measured by time on the market or whether landlords reduce rent prices of unsuccessful apartment postings, but we do not observe these information in our data. What we can see in our data is that rounders only increase prices in tight markets, in which they have high market power. This suggests that their strategy could be less successful than charming in more relaxed markets. Further investigating in which market situations a charming strategy with respect to apartment size, or more generally product characteristics, is superior to other strategies is an interesting task for future research.

The implications of our results are not limited to the housing market as behavioral biases can be exploited in other market situations, as well. Our results not only underline previous findings that price discontinuities in used-car markets are driven by limited attention of final customers (Busse et al. 2013; Englmaier et al. 2018b; Lacetera et al. 2012). Even more, they

[34]We observe the number of apartments of each landlord, which can be a (crude) proxy for experience. However, our data does not suggest that this measure is correlated with charming, rounding or rent-price discontinuities.

[35]Previous literature has shown that price charming leads to better market outcomes than price rounding (Chava and Yao 2017; Repetto and Solís 2019).

suggest that some car dealers could actively exploit biased customers. Limited attention can be exploited in any other domain in which continuous measures are mapped into discrete categories such as star-based ratings. For example, restaurants can do so by increasing prices if their star rating reaches a certain threshold.[36] In a similar way, e-commerce platforms could apply a flexible price-setting approach in which they increase prices for products that reach certain rating thresholds.

Finally, our results are of important economic and social relevance. The observed price patterns could lead to market distortions, but these are likely limited to very relaxed markets in which vacancies are abundant and tenants are able to shun overpriced offerings. However, the price-setting behavior of landlords certainly leads to a socially undesired redistribution from (potentially less wealthy) tenants to (potentially wealthy) landlords. Because expenses for housing constitute a major part of household spending, any systematic exploitation of inattention leads to significant economic impacts for a large share of the population. Our findings therefore carry important policy implications. Educating market participants or raising the attention to exact size measures might not suffice to prevent exploitive behavior. Due to a shortage of alternatives, especially in tight markets, tenants will likely have to accept an overpriced offer even if they detect price discontinuities. Further, IS24 might be reluctant to change their platform design[37] as most of their revenue consists of fees paid by landlords. Consequently, to prevent the unwanted redistribution regulation might be necessary. Further investigating policy responses and finding the best measures to prevent exploitation of inattention, not only in the housing market but in general, will be an important task for future research.

[36]Luca (2011) shows that consumers are more responsive to star ratings than to precise information.

[37]One option to redesign the platform is to change how size filtering works. If tenants for example set minimum size requirements to $10m^2$ marks, IS24 could nevertheless show apartments slightly smaller than this threshold. Although this might not fully prevent exploitive behavior, it would make it easier for tenants to detect price discontinuities and opt for slightly smaller but considerably less expensive apartments.

Chapter 2

Defying Gravity: What Drives Productivity of Remote Teams?[1]

Abstract

How can teams organize for productive online collaboration? The coronavirus pandemic has led to a large and persistent shift toward remote work. Using fine-grained data from the world's largest platform for open-source software development, we find that although its importance has slightly decreased, geographic proximity still matters for knowledge workers. We find that the pandemic reduced the productivity of previously co-located teams substantially, whereas teams with remote work experience remained resilient. A large set of controls and matching approaches show that this result is not driven by pre-existing differences between co-located and distributed teams. While access to remote talent and experience are important for overall success, our results highlight the crucial role of communication for productive online collaboration. We find suggestive evidence that, with their peers shifting to online work, remote workers become better integrated into their teams' communication. We conclude that while teams' performance may suffer from the shift to remote work, setting up systems for effective online communication can help mitigate productivity loss.

[1]This chapter is based on joint work with Thomas Fackler and Nadzeya Laurentsyeva.

2.1 Introduction

The growing importance of immaterial goods and increasing digitalization have enabled virtual production processes and have made virtual teamwork possible. Digital technologies not only reduce communication costs over long distances but also help create powerful environments with no physical barriers to collaboration and the exchange of knowledge and ideas. The COVID-19 pandemic has accelerated the adoption of technologies and new work practices by forcing teams to work remotely, at least during the lockdowns. In the (post-pandemic) future of work, remote work will likely be much more frequent than in the past, ranging from entirely remote companies to hybrid models that blend working from the office with working from anywhere (Barrero et al. 2021).

Do remote and hybrid work modes represent long-term viable solutions for the organization of knowledge teams? On the one hand, remote-work arrangements allow firms to access a larger talent pool of skilled workers and, hence, to address skill shortages, foster innovation, and expand geographically. Simultaneously, many individual knowledge workers value flexibility and express their preference for a remote or hybrid job (Aksoy et al. 2022; Bloom et al. 2022). On the other hand, the existing research has stressed the importance of co-location for knowledge transfer, complex problem-solving and knowledge production, idea generation, and coordination (Bahar et al. 2022; Emanuel et al. 2022; Gibbs et al. 2021; Hu and Jaffe 2003; Yang et al. 2022).

This project aims at identifying which characteristics make remote collaboration in knowledge teams more or less productive. Our empirical setting is GitHub — the world's largest online platform for software development and code sharing. We analyze collaborations among open-source software engineers using the COVID-19 pandemic as a natural experiment, which forced GitHub teams (as many other knowledge teams worldwide) to work remotely, with limited opportunities for offline interactions.

We first show that despite the available digital infrastructure for remote work and the fact that both the production process and the output of GitHub users are immaterial, geographic proximity still plays a vital role for online collaborations. By applying a standard gravity model to a city-pair panel data of global activities on GitHub between 2012 and 2021, we find that once distance doubles, the flow of expected collaborations between a given city pair drops by about 50%. This highlights the importance of co-location for knowledge teams even when technology for virtual work is available. We note, however, that the role of geographic proximity has decreased with the start of the COVID-19 pandemic.

Our main analysis zooms in on the level of individual projects and focuses on their performance before and during the pandemic.[2] We benefit from rich data that allows us to observe teams and their spatial distribution before the pandemic. We can thus estimate how the productivity effect of COVID-19 varied between ex-ante *co-located* teams, in which all members shared the same location, ex-ante *fully distributed* teams, in which all members worked from distinct locations, and ex-ante *mixed* teams. While all types of teams had to work remotely during the pandemic, the prior reliance on face-to-face vs. remote collaboration and, hence, the actual exposure to the shock varied. It is plausible that co-located teams were hit the hardest as their team members had to adjust to the new work mode.

First, we estimate the effect of the COVID-19 pandemic on performance, separately for each team type. To that end, we compare teams' performance during the pandemic to their hypothetical performance in the absence of the shock. We find that co-located teams are significantly less active during COVID-19 when compared to an artificial control group of earlier projects at similar maturity that did not experience the pandemic. However, we do not observe any negative effects among the fully distributed and mixed teams. If anything, the productivity of mixed teams seems to increase during the pandemic relative to a control group of earlier projects. These results confirm our intuition that co-located teams were negatively

[2]We refer to a group of individuals working on the same project as a team.

affected by the pandemic and show an interesting heterogeneity of pandemic effects.

In order to investigate this heterogeneity more thoroughly, our second specification compares team performance *across*, rather than within team types. This allows us to better isolate the differential impact of the COVID-19 pandemic on different teams. Members of distributed teams made about 45% more contributions than members of co-located teams, conditional on a rich set of pre-shock team characteristics. These results are robust to matching and placebo tests. The findings are consistent with the idea that distributed teams were already used to online collaboration before COVID-19 and thus were better prepared for fully remote work during the pandemic.

Next, we explore potential mechanisms and address concerns that co-located teams could be inherently different from distributed ones. The differential impact of COVID-19 on teams is not driven by differences in experience[3] or a better ability of distributed teams to source workers from the open-source community. Online communication, as proxied by comments on code contributions, is more abundant in distributed teams and plays an increased role in team productivity after the onset of COVID-19, which supports our idea that prior experience with remote work was beneficial during the pandemic. Note, that even combined these additional factors cannot (fully) explain the performance differences between co-located and distributed teams during the pandemic.

At the individual member level, we find suggestive evidence that remote workers in mixed teams have become relatively more productive. This is consistent with the idea that remote workers benefited from better integration into their teams as all communication had to move online, such that they no longer missed out on offline exchanges.

Taken together our results show that organization matters for team productivity, highlighting, in particular, the crucial role of communication and of the ability to involve all team

[3]Teams with more experienced members are more resilient to the pandemic on average, but this does not explain differences between co-located and distributed teams.

members regardless of their location.

This project relates to several strands of literature. First, it contributes to the growing body of studies on the determinants of team productivity, in particular under remote and hybrid work arrangements. Many of these studies use the COVID-19 pandemic shock as a natural experiment. Gibbs et al. (2021) investigate the difference in productivity before and during the work-from-home period of COVID-19. Using personnel data from over 10,000 skilled professionals at a large Asian IT services company, the study finds that working hours increased by 18%, while the average output decreased by 8–19%. The authors suggest that one of the reasons for the decline in productivity are higher communication and coordination costs associated with remote work. Yang et al. (2022) analyze data from over 60,000 Microsoft employees in the US over the first six months of 2020. The study finds that remote work caused the collaboration network among workers to become more static and isolated with a decrease in synchronous communication, and an increase in asynchronous communication. According to the researchers, these changes could make it more difficult for employees to share new information across the network.

McDermott and Hansen (2021) examine the impact of COVID-19 on labor activity by analyzing data from GitHub. Their findings suggest that the pandemic led to a shift in the pattern of labor allocation, with a higher likelihood of users working on weekends and outside of regular working hours. Closely related to our research question, Lu et al. (2023) examine how remote GitHub teams were affected by the COVID-19 pandemic. The authors compare the productivity and sizes of the projects with what was predicted in the absence of a pandemic. The study finds that the productivity and number of active members of GitHub teams varied considerably during different phases of the pandemic. Also, the resilience of a team under shock is closely tied to specific team characteristics before the pandemic, such as the country diversity, multitasking level, member experience and prestige, as well as emotions. However, Lu et al. (2023) focus on teams that already relied on remote work before the pandemic.

In contrast, our study specifically investigates the differential impact of the pandemic on distributed teams and co-located teams, which allows us to analyze the productivity effects of moving from offline collaboration to an entirely virtual production process.

Similar to our conclusions, a number of studies highlight the importance of communication for productivity of remote teams. Emanuel et al. (2022) examine whether online interactions can replace face-to-face interactions or whether the two are complementary by focusing on software engineers at a Fortune 500 company. Their findings indicate that online technology complements face-to-face interaction with proximity being an important factor in how much knowledge workers gain from their colleagues. Bojinov et al. (2021) conduct a field experiment to determine the most effective way to onboard organizational newcomers who are working remotely. The results stress the importance of virtual water coolers, in particular when the participants share demographic characteristics, such as gender or ethnicity, which facilitate communication. DeFilippis et al. (2020) examine how the COVID-19 pandemic has affected employees' digital communication patterns by conducting an event study of lockdowns in 16 large metropolitan areas across North America, Europe, and the Middle East. Using de-identified and aggregated meeting and email metadata from over three million users the study reveals that compared to the pre-pandemic levels, the number of meetings per person and the number of attendees per meeting increased by 12.9% and 13.5%, respectively, while the average length of meetings decreased by 20.1%. Additionally, the study finds that the average workday increased by 48.5 minutes.

Second, by estimating a gravity model for online collaborations we contribute to the established literature that has helped identify the determinants of bilateral trade in goods and services or of migration flows between different geographical units. The gravity equation models bilateral interactions between geographic units, where economic size and distance effects enter multiplicatively. Such models have been used as a workhorse for understanding the determinants of bilateral trade flows for over 50 years since being introduced first by

Tinbergen (1962); see Head and Mayer (2014) for a recent survey. They have also been widely applied to study the determinants of migration flows, see Beine et al. (2016) and Ramos (2017) for reviews of modelling approaches and Mayda (2010) and Migali et al. (2018) for applications to international migration. By applying the gravity model to an online setting, we can identify the determinants of virtual collaborations and benchmark them against those established in the trade literature. We can also benefit from finer geographical data and identify drivers of collaboration at the city, rather than country level.

Third, studying cross-city and cross-country code contributions, our paper is not only related to trade, but also to the literature on knowledge flows and knowledge production. Knowledge has been shown to be more localized than what would be expected from agglomeration effects alone (Jaffe et al. 1993). Furthermore, knowledge spillovers to other countries have been shown to take time (Hu and Jaffe 2003; Jaffe and Trajtenberg 1999) and the effect of international localization has turned out to be more robust over time than within-country localization (Thompson and Fox-Kean 2005). While a large body of this literature draws on analyzing patent data and thus focuses on inventors, we provide new evidence on collaboration and knowledge flows among software engineers.

The rest of the paper is organized as follows. Section 2.2 discusses the GitHub context and provides details on the dataset used for the analysis. Section 2.3 presents estimations of the gravity model for city-level collaborations on GitHub and highlights the role of proximity on the platform. We focus on the performance of individual projects and remote workers in Section 2.4. Section 2.5 concludes.

2.2 Context and data

GitHub is a software development platform featuring a collaborative version control system and was launched in April 2008. As of early 2023, GitHub hosts the world's largest community of software developers comprising almost 100 million users and over 300 million repositories.[4,5]

GitHub projects cover a wide variety of (mostly) software applications, some of which provide tools for other developers, while some serve a wider audience. Projects can be started by both individual users and companies. GitHub allows its users to choose between private and public repositories for their projects. The latter are usually licensed under common open-source licenses such as the GNU General Public License, MIT License or Apache 2.0 License. Many open-source software repositories hosted on GitHub are used in thousands of other projects, including academic research, proprietary software, as well as projects in governmental and nonprofit organizations. Motivations of open source contributors have been the subject of economic research and include knowledge seeking and creation, career concerns (showcasing skills), paid work at software companies, as well as writing software for one's own needs or to help others (Belenzon and Schankerman 2008; Hergueux and Jacquemet 2015; Lerner and Tirole 2001, 2005).

Our data cover the activities in GitHub's public repositories for which we can observe all code contributions to open-source projects and other user interactions, such as networking, code borrowing, bug reporting or commenting. To contribute to a public project or to create a new one, users have to set up an account (unless they already have one) where they can share their real name, location (usually a current city) and additional biographical information. Each project has only one owner. The owner may invite other users to contribute and to become project members. Users can also initiate a collaboration and contribute to a project even before being invited (McDonald and Goggins 2013). Users who are not project members can suggest modifications to the code, which the project members can review and either reject

[4]Repository is GitHub lingo and denotes a project's virtual folder. Throughout this study we use the terms "repository" and "project" as synonyms.

[5]The platform is constantly growing. For instance, in 2022 alone, 20.5 million new developers joined GitHub. Source: https://octoverse.github.com

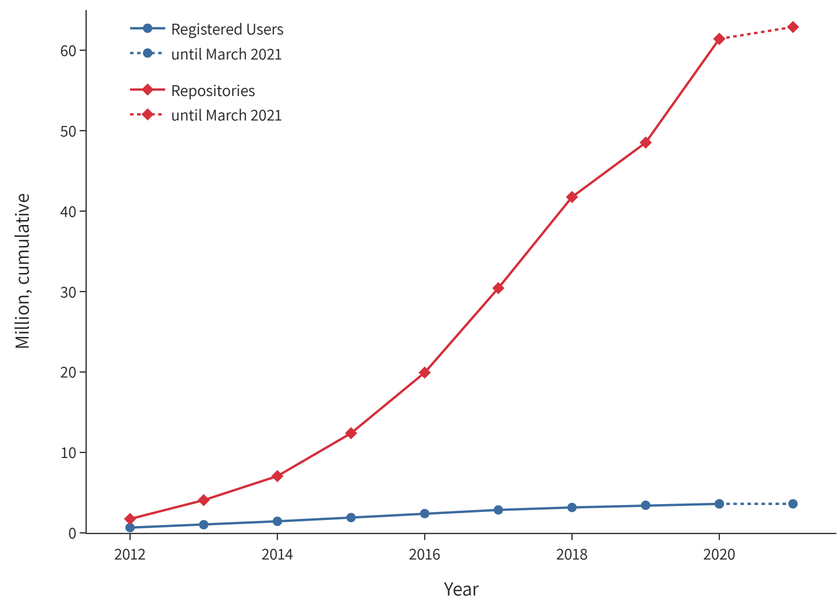
or accept. Users' profile pages on GitHub show their contributions to different public projects, while project pages reveal which users have contributed. Thanks to the version control system, the development history of a project is recorded down to each contribution. Along with tools for software development, GitHub shares features of social networks, allowing users to receive updates about each other's activities, follow projects and give "stars" to the ones they like.

We use two publicly available GitHub datasets for our analysis: a snapshot from GitHub Torrents (GHT) (Gousios 2013) and the GitHub Archive dataset (GHA).[6] Both datasets provide a mirror of the GitHub public event stream from 2012 on. We use the two datasets in a complementary way. We take the event stream data from GHA because it is updated in real time and allows us to incorporate the up-to-date activity data. Data on users (in particular, their reported geographic locations) is available in the GHT dataset. Therefore, we merge the latest available snapshot of the GHT dataset (March 2021) to our event data from GHA, which spans from 2012 to 2021. Given our research question, we have to limit the data to events where we can identify the geographic location of project owners and project committers. As Figure 2.1 shows, that leaves us with about 3.6 million registered users and about 62.6 million repositories.[7] Section B.0.1 in the Appendix provides technical details on merging the GHA and GHT datasets.

[6]https://www.gharchive.org

[7]In total, as of March 2021 there were about 36 million registered users working on GitHub public repositories and 183.5 million public repositories.

Figure 2.1 *Cumulative number of registered users and public repositories on GitHub*



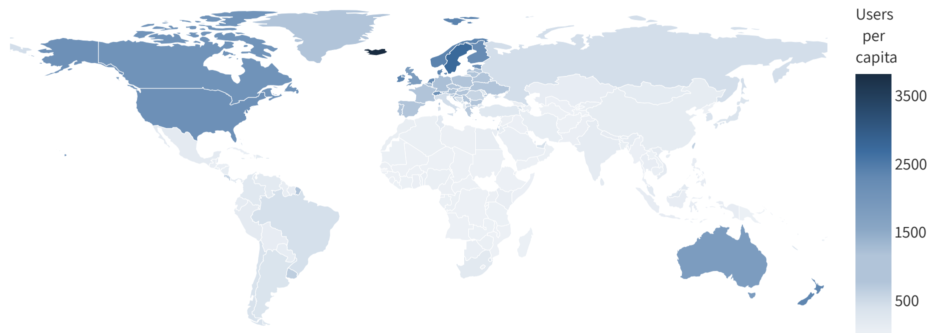
Notes: This graph shows the cumulative number of GitHub users with reported locations and the number of public repositories owned by users with reported locations.

2.3 Geography of collaborations on GitHub: importance of proximity

Since its start in 2008, GitHub has become popular with users worldwide. Figure 2.2 shows the number of GitHub users in our data relative to a country's population (in millions). Overall, more advanced countries have a higher share of registered users. It should be noted that even though per capita activity is highest in North America, Europe and Oceania, populous countries such as India and China have sizable user bases on GitHub, as well. As of March 2021, the top five countries in terms of the absolute number of GitHub users were the United States, India, China, Great Britain and Brazil.

The main purpose of GitHub has been to ensure smooth collaboration and knowledge exchange among users regardless of their location. Unlike many other settings in which remote work is a relatively new phenomenon, GitHub teams have had the necessary technological

Figure 2.2 *World map of GitHub users per capita*



Notes: This map shows the number of GitHub users per capita (population in millions, i.e. users per one million inhabitants). It is based on users with reported location and repositories owned by users with reported location.

infrastructure to function virtually ever since 2008. Therefore, the first question we ask in this paper is, to what extent geographic proximity still matters for collaboration in GitHub teams.

2.3.1 Gravity in online collaborations

To quantify the role of geographic barriers, we adapt the standard gravity model from the trade literature and estimate the gravity of collaborations on GitHub at the city-pair level. We aggregate the combined GHT-GHA dataset (2012–2021) at a city-pair and year level. We further restrict our data to about 700 of the most active cities on GitHub (as proxied by the number of registered users as of March 2021).[8] These cities together account for 76% of all users and for 82% of all commits by users with reported locations. We construct a strongly balanced annual panel dataset by forming all possible city pairs from our sample for a period between 2012 and 2021, which results in about 5.3 million observations. Only 194,630 of our observation cells, however, are greater than zero. Section B.0.2 in the Appendix provides details of the empirical specification.

Our results in Table 2.1 show that geographic proximity matters even in virtual environ-

[8]We set a cutoff of at least 400 registered users per city as of March 2021, resulting in 727 cities. Our results are robust to setting a lower cutoff.

ments on GitHub. The effect of geographic distance on online collaborations is negative and statistically significant with an estimated elasticity of 0.493: if distance doubles, the number of commits drops by about 50% (column 1); conditional on non-zero collaboration, the number of commits drops by about 40% (column 2). On the extensive margin (column 3), the linear probability results show that if distance doubles, the probability of having a cross-city collaboration drops by about 2.4 percentage points.

Column (4) in Table 2.1 uses distance bins instead of continuous distance measures to capture non-linear distance effects. The reference category corresponds to collaborations within the same city. The results highlight a non-linearity in the distance effect and suggest that interactions on GitHub are substantially more likely to happen within the same city, i.e. between people who know each other personally or who can collaborate in an offline setting. Beyond the distance of 300km, the effect stays at about the same level.

These effects are economically significant. As Table B.1 in the Appendix shows, the distance elasticity for GitHub collaborations is almost two-thirds compared to that for trade, even though the role of trade costs (transportation, legal costs, search costs for partners) should be negligible on the platform. Our results also show, that conditional on distance, state borders reduce virtual collaborations, while a common language mitigates this negative effect. Similar patterns have been observed in the literature on trade, migration and knowledge flows.

Table B.2 in the Appendix compares the baseline results with those obtained from running regressions within the same programming language. Hence, we check to what extent the effect of distance is driven by technological differences between cities. While the magnitude of the coefficient decreases (from -0.420 to -0.282), it remains statistically significant suggesting that technological differences cannot fully explain the gravity in online collaborations.

Table 2.1 *Gravity model for collaborations on GitHub*

Variables	Contributions (1)	Contributions >0 (2)	Contributions yes/no (3)	Contributions distance dummies (4)
Distance	−0.493***	−0.391***	−0.034***	
	(0.093)	(0.059)	(0.007)	
1–50km				−0.195
				(0.380)
50–100km				−1.845***
				(0.324)
100–300km				−1.781**
				(0.815)
300–700km				−3.151***
				(0.524)
>700km				−3.791***
				(0.474)
Foreign country	−2.839***	−1.281***	−0.053	−2.978***
	(0.613)	(0.359)	(0.040)	(0.315)
Common language	0.667***	0.177	0.010*	0.607***
	(0.174)	(0.147)	(0.006)	(0.162)
Users, owner	0.378***	0.212***	0.027***	0.439***
	(0.041)	(0.056)	(0.005)	(0.047)
Users, committer	1.094***	0.825***	0.032***	1.092***
	(0.068)	(0.049)	(0.005)	(0.074)
Remoteness, committer	2.233	2.271	0.104***	1.459
	(1.866)	(1.953)	(0.031)	(2.039)
Remoteness, owner	−1.027	−1.892	0.076***	−1.588
	(1.071)	(1.189)	(0.025)	(1.225)
Observations	5,347,967	194,630	5,372,810	5,347,967
Clusters	8464	3394	8464	8464
R-squared	0.659	0.520	0.109	0.669

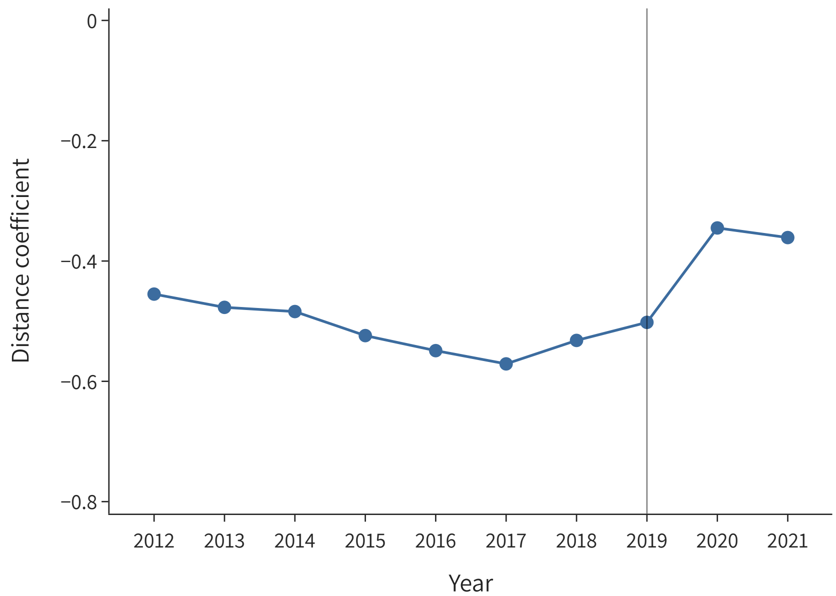
Notes: The dependent variable in columns (1), (2), and (4) is the number of contributions between a given city pair. Column (2) limits the sample to city pairs with non-zero contributions. The dependent variable in column (3) is a dummy equal to one if there are any contributions in a given city pairs. Distance, number of users, and remoteness are in natural logarithms. All specifications include country of committers and country of project owners time-specific fixed effects. Standard errors are clustered at a country pair level. Estimation method: PPML in columns (1), (2) and (4) and OLS in column (3). For specifications estimated with PPML, pseudo R-squared is calculated.

2.3.2 The effect of the COVID-19 pandemic on the role of proximity

Given the emerging evidence that the COVID-19 pandemic has transformed the organization of work, we are interested in tracing whether the role of proximity for online collaboration has changed between 2012 and 2021. In Figure 2.3 we show how the coefficient for distance has changed over time. We estimate the same regression model as in Table 2.1 column (1) but separately for each year in our sample.

We observe a drop in the role of proximity that coincides with the COVID-19 pandemic: in 2020, distance elasticity constituted -0.3 (while being below -0.4 in 2019). This suggests that the role of geographic distance for online collaborations has weakened during the pandemic. The pandemic and associated lockdowns likely represented a stronger shock for ex-ante co-located teams compared to already distributed ones, because the former had to adjust their work practices under time pressure. It could also align with the fact that the shift to remote work had geographically close projects lose their comparative advantage in terms of easier offline communication. In the next section, we explore the COVID-19 effects on GitHub collaborations in more detail, by using project-level data and comparing the performance of co-located and distributed teams.

Figure 2.3 *Distance elasticity of collaborations on GitHub in 2012–2021*



Notes: This figure shows the yearly distance elasticity of code contributions. The coefficients are obtained from estimating regressions equivalent to that of Table 2.1 column (1), separately for each year.

2.4 The effect of COVID-19: team-level analysis

The analysis in the previous section has shown that international online collaboration patterns have changed during the pandemic. This section moves from the aggregate cross-city perspective to the level of individual projects. How was the activity of existing projects affected by COVID-19, and how did this effect differ by (co-)location of team members?

Our analysis focuses on small projects, which have attracted exactly three members within the first year of existence. We select these small teams for mainly two reasons. First, it allows us to clearly distinguish team compositions into three categories: *Co-located* teams in which all members share the same location, *mixed* teams in which two members are from the same location and one member is remote, and *(fully) distributed* teams consisting of members from three distinct locations. Second, fixing the number of members ensures that we do not mechanically introduce systematic team-size differences between the categories, since larger teams are more likely to contain at least one non co-located member.[9] Note, however, that we only fix the number of members within the first year and allow additional members and contributions from non-members afterwards. We explicitly investigate this potential mechanism below.

We further choose projects which were started between 2015 and 2018 and thus analyze contributions to pre-existing projects but not to new projects. This way we exclude effects from systematic differences in team composition of newly formed teams during the pandemic. Using earlier years than 2015 would not add much power, given that only few projects last for more than five years. Instead, it might introduce projects that are less comparable if the type of projects publicly developed on GitHub has shifted over time. Including projects that were started after 2018 would limit our ability to observe how these projects performed before the onset of the coronavirus pandemic.[10]

We investigate the effect of the coronavirus pandemic by employing two complementary difference-in-differences (DiD) designs. First, in Section 2.4.1 we estimate how COVID-19 affected team productivity for the three types of teams individually. For this, we create artificial controls groups using earlier projects. Second, in Section 2.4.2 we more directly estimate the differential effects of COVID-19 on co-located, mixed and distributed teams. While the first approach allows us to estimate *actual* performance implications of the pandemic, the second approach estimates *relative* performance differences across teams, enabling us to net out general pandemic effects that are independent of team types.

[9] Furthermore, our requirement that all team members state their location would less likely be satisfied for larger teams.

[10] In addition, we select projects consisting of three members *within the first year*. For projects created after February 2019 this could partly include members who joined during COVID-19, leading to undesired selection effects.

2.4.1 The effect of COVID-19 on team performance

In our first analysis we use a difference-in-differences design to compare the performance of teams during the coronavirus pandemic to the performance during normal times, separately for the three types of teams.

Empirical setup

Since COVID-19 is a global pandemic affecting all teams at (roughly) the same time, we construct an artificial control group using the following approach: We define projects which were started in 2018 as the treatment group and compare their performance during the pandemic to the performance of earlier teams at similar maturity. For these earlier projects we mimic the COVID-19 timing of 2018 projects, i.e. we assume a hypothetical onset of the COVID-19 pandemic in the third calendar year of their existence. Thus, the hypothetical outbreak of the coronavirus is set to March 2019 for projects started in 2017, March 2018 for projects started in 2016 and March 2017 for projects started in 2015.

We restrict the observational period to 24 months, centered around the hypothetical coronavirus outbreak. This ensures that none of the control teams ever actually experience the COVID-19 pandemic, thus avoiding the empirical challenges through staggered treatment (e.g. Callaway and Sant'Anna 2021). Further, we only keep projects which received at least one contribution in the observational period, and we remove outliers which received more than 500 contributions within the first two years of existence.

In order to estimate the effect of the coronavirus pandemic we estimate the following equation using a Negative-Binomial regression to account for the fact that the number of commits is count data:

$$c_{it} = \beta_0 + \beta_1 \text{COVID}_t \times \text{Treated}_i + \beta_2 \text{COVID}_t + \beta_3 \text{Treated}_i + X_{it}\delta + \varepsilon_{it}, \qquad (2.1)$$

where c_{it} is the number of code contributions to project i in month t. COVID_t denotes the hypothetical COVID-19 pandemic and is set to one in March of the respective year (see above). Treated_i is set to one for projects created in 2018, those which actually experience the pandemic, and zero for projects started earlier. X_{it} is a vector of control variables and ε_{it} is the error term, which is clustered at the project level. The coefficient of interest is β_1 which measures the impact of the (actual) coronavirus pandemic on the activity of treated projects.

We split the sample by team composition and estimate Equation (2.1) separately for co-located, mixed and fully distributed teams. Because co-located teams had to adjust to new virtual work modes, we expect their performance to decrease during the pandemic (negative β_1). The effect on mixed and distributed teams is ex-ante not clear and therefore we let the data speak. In addition to the split-sample analysis, we estimate an extended version of Equation (2.1): We add triple interactions between COVID_t, Treated_i and indicators for mixed teams, D(Loc.=2)_i, and fully distributed teams, D(Loc.=3)_i in order to estimate COVID-19 effects on all team types in a single regression.

The key identifying assumption in our setting is that in absence of the coronavirus pandemic the activity of projects started in 2018 would have followed a parallel trend to the activity of those started earlier, conditional on a set of control variables. Note, that this design only compares projects within and not across team types and therefore identification does not hinge on parallel trends across co-located, mixed and distributed teams.

To ensure the robustness of our estimates we add a rich set of control variables. These controls include month fixed effects to capture seasonality in commits; project age (measured in months, linear and quadratic) to model declining activity as projects mature; the number of commits and the number of watchers within the first year of a project's existence to account for differences in project productivity and success; the country of the project owner[11] to control for differences in COVID-19 responses across countries. Except for project age and

[11]We distinguish between the five most represented countries in our sample (USA, Brazil, Great Britain, India and Canada) and an "other" category.

month fixed effects, we allow all of these controls to have time-varying effects before and after the onset of the (hypothetical) coronavirus pandemic. In the Appendix we estimate two-way fixed-effects models, dropping month fixed effects but adding time and project fixed effects while keeping interacted controls. Also in the Appendix, we use matching methods to further probe the robustness of our results.

Results

To get a first impression of the effect of the pandemic on team productivity, Figure 2.4 shows two event-study graphs of quarterly effects on the productivity (number of commits) in different types of teams. Figure 2.4a splits our sample into co-located teams and distributed teams (including both fully distributed and mixed teams). While the blue line for distributed teams is rather flat throughout, the red line for co-located teams clearly decreases after the start of the pandemic and becomes significant at the 5% level in Q4 of 2020.

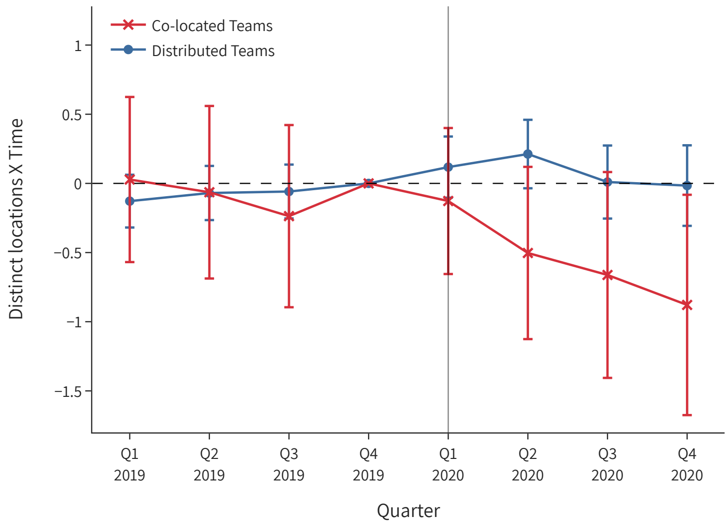
Figure 2.4b further subdivides distributed teams in those with three locations (fully distributed) and those with two locations (mixed teams). Although differences are not statistically significant, this graph shows an interesting pattern that we will explore further in the analysis on remote worker productivity in Section 2.4.3: The yellow line (mixed teams) appears to increase slightly with the start of the pandemic, whereas the blue line (fully distributed teams) is flat throughout. Overall, Figure 2.4 confirms the intuition that co-located teams, which had to adjust to online collaboration, were significantly negatively affected, while mixed teams show a small increase and fully distributed teams were hardly affected.

Table 2.2 summarizes the results of estimating Equation (2.1).[12] In columns (1), (2) and (3) we split the sample into co-located, mixed and fully distributed teams, respectively. For fully distributed teams the interaction effect in column (3) is close to zero and insignificant,

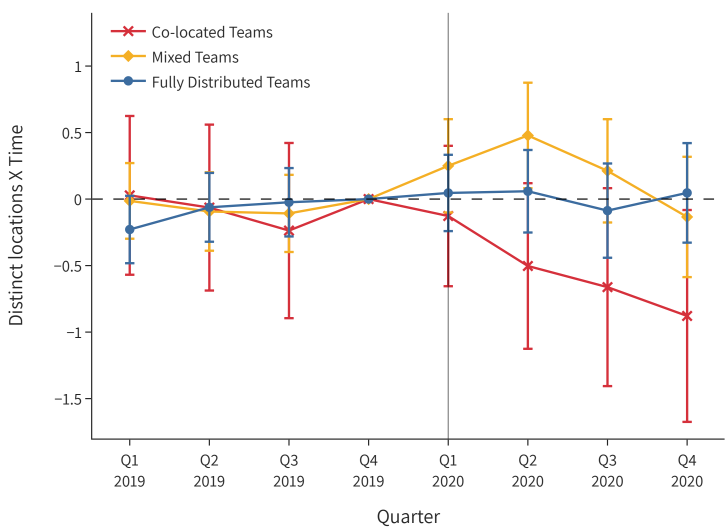
[12]We interact COVID_t with multiple control variables. Therefore, its main effect and the constant are omitted since they cannot be interpreted in a meaningful way.

Figure 2.4 *Quarterly COVID-19 effects on team performance*

(a) *Co-located vs. distributed*



(b) *Number of locations*



Notes: The figures show interaction coefficients and 95% confidence intervals of quarterly time dummies and indicators for projects started in 2018. Each line is from a separate regression with earlier projects as comparison category. The dependent variable is the number of commits and controls include project age (linear & quadratic), number of commits and watchers within a project's first year and country-of-owner fixed effects. Panel (a) splits the sample into co-located teams and distributed teams (two and three locations). Panel (b) splits the sample into co-located, mixed and fully distributed teams. Estimation method: Negative-Binomial ML.

Table 2.2 *COVID-19 effects on team performance*

	Dependent: Number of commits			
Team composition	Co-located (1)	Mixed (2)	Distributed (3)	Full sample (4)
Treated	–0.033 (0.115)	0.015 (0.069)	0.162*** (0.058)	–0.067 (0.119)
Treated × COVID	–0.424** (0.199)	0.150 (0.127)	0.031 (0.121)	–0.366* (0.204)
D(Loc.=2)				0.020 (0.075)
D(Loc.=3)				0.096 (0.068)
Treated × D(Loc.=2)				0.064 (0.140)
Treated × D(Loc.=3)				0.233* (0.131)
D(Loc.=2) × COVID				–0.084 (0.136)
D(Loc.=3) × COVID				0.088 (0.138)
Treated × D(Loc.=2) × COVID				0.520** (0.245)
Treated × D(Loc.=3) × COVID				0.457* (0.255)
Controls	X	X	X	X
Pseudo R^2	0.121	0.118	0.094	0.132
N	18,888	68,136	106,968	193,992
Clusters	787	2,839	4,457	8,083

Notes: Columns (1), (2) and (3) split the sample into co-located, mixed and distributed teams, respectively. *Treated* projects are founded in 2018 and are exposed to 12 months of COVID-19. The control group are earlier projects with the hypothetical onset of the pandemic shifted backwards. For example, the hypothetical onset of the pandemic is set to March 2019 for projects started in 2017. Controls include project age (linear & quadratic), number of commits and watchers within a project's first year, country-of-owner fixed effects and month fixed effects. Except for project age and month fixed effects all controls are interacted with the (hypothetical) COVID indicator. The sample covers 24 months centered around the hypothetical onset of the coronavirus pandemic. Estimation method: Negative-Binomial ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

confirming that these teams are barely affected by the coronavirus pandemic.[13] The coefficient for mixed teams in column (2) is statistically insignificant, as well, but the point estimate is slightly larger at 0.150. The effect on co-located teams in column (1) is negative, the largest in magnitude (-0.424) and significant at the 5% level. The interaction effect suggests that during the pandemic co-located projects were only about 65% ($e^{-0.424} = 0.654$) as productive as one would expect had COVID-19 not happened. In column (4) we estimate an alternative specification, adding triple interactions to measure differential treatment effects for the three team types in a single regression. This yields qualitatively very similar results. The coefficient of $\text{COVID}_t \times \text{Treated}_i$ measures COVID-19 effects for co-located teams (the left-out category) and is negative, slightly smaller than in column (1) and still significant at the 10% level. Both triple interactions for mixed and distributed teams are positive and significant, offsetting the general negative impact of the coronavirus pandemic. Table B.5 in the Appendix repeats all four regressions including project fixed and time fixed effects. This not only improves the amount of explained variance (Pseudo R^2), but also leads to even stronger results, especially for co-located teams.[14]

We additionally probe the robustness of our results using two matching methods. First, we use the inverse probability weighting (IPW) approach suggested in Abadie (2005) which keeps all projects but re-weights them to achieve balance between treatment and control group. Second, we use one-to-one propensity score matching applying the optimized algorithm of Hansen and Klopfer (2006). We estimate propensity scores using logistic regressions of Treated_i on the full set of control variables[15] and in two alternative approaches we add

[13]The significant effect of Treated in column (3) may indicate that fully distributed teams that were started in 2018 were somewhat more productive than earlier fully distributed teams before the pandemic.

[14]Negative-Binomial models do not converge if both project fixed and time fixed effects are included. Therefore, these estimates are obtained using Poisson ML. Negative-Binomial models using project fixed and month fixed effects, instead of time fixed effects, yield very similar results (not shown).

[15]We include project age (at the beginning of the observational period), the number of commits and watchers in the first year of a project's existence as well as country of the project owner.

pre-pandemic outcome levels or pre-pandemic outcome trends.[16]

Combining the different methods and propensity scores and applying them separately to co-located, mixed and fully distributed teams yields 18 individual regressions, which are summarized in Table B.6 in the Appendix.[17] Throughout, the results are closely comparable to those of unmatched regressions in Tables 2.2 and B.5, with two noteworthy (but minor) deviations. First, the estimated negative impact of COVID-19 on co-located teams appears to be even larger when applying one-to-one matching. Second, also when using one-to-one matching, the positive COVID-19 effect on mixed teams is statistically significant at the 10% level in columns (4) and (5). However, Figure B.2 suggests that one-to-one matching performs worse than IPW in balancing the covariates across treatment and control projects. Furthermore, we are aware of the general caveats of matching in DiD settings, such as regression to the mean (e.g. Daw and Hatfield 2018), which can increase bias especially when matching on pre-treatment outcomes. Therefore, we refrain from over-interpreting these observations and prefer unmatched results, but conclude that results from the matched regressions further bolster the robustness of our main findings.

2.4.2 Differential effects of COVID-19 across team types

While the design in the previous subsection allows us to quantify the effect of the coronavirus pandemic separately for co-located, mixed and fully distributed teams, the estimates capture both team-type specific effects and general effects of COVID-19, which are similar for all teams. The combined magnitude and direction of the latter are a priori not clear.[18] Although

[16]We include quarterly pre-pandemic outcome levels (number of commits), yielding four observations per project, and match on trends by including first-differences of these quarterly levels.

[17]Figure B.2 shows standardized mean differences of control variables across treated and control projects. It suggests that, in general, the IPW approach yields better balance than one-to-one matching. Adding pre-pandemic outcome levels or trends slightly improves the balancing of these factors.

[18]For example, COVID-19 lock-downs could have increased the time available to work on GitHub projects for some users, but could also have limited the available time for parents if they have to supervise kids who are in homeschooling.

the resilience of distributed teams supports the notion that the negative impact on co-located teams is driven by the fact that they are co-located, we cannot entirely rule out alternative channels. Therefore, we directly estimate relative performance differences across team types during the coronavirus pandemic in this subsection. This approach better isolates the *differential* impact of the coronavirus pandemic on the three types of teams from general disruptions which affect all teams. One further advantage of this approach is that it does not rely on comparing projects from 2018 to an artificially created control group with hypothetical pandemic periods, but allows us to estimate COVID-19 effects based on projects of all vintages. However, this comes with the caveat that we can only estimate relative performance differences between teams and thus have to rely on our previous results of Section 2.4.1 to quantify actual performance effects of COVID-19.[19]

Empirical setup

We estimate the following alternative specification, again using Negative-Binomial regression:

$$\begin{aligned} c_{it} = {} & \beta_0 + \beta_1 \text{COVID}_t \\ & + \beta_2 \text{D(Loc.=2)}_i + \beta_3 \text{D(Loc.=3)}_i \\ & + \beta_4 \text{D(Loc.=2)}_i \times \text{COVID}_t + \beta_5 \text{D(Loc.=3)}_i \times \text{COVID}_t \\ & + X_{it}\delta + \varepsilon_{it}. \end{aligned} \tag{2.2}$$

The dependent variable (c_{it}) is the number of commits to project i in month t. COVID_t is a dummy that is equal to one starting from March 2020 and zero before, and accounts for the difference in activity that affected all projects after the start of the pandemic. D(Loc.=2)_i and D(Loc.=3)_i indicate mixed and distributed teams, respectively, and their main effects (β_2 and β_3) measure general performance differences compared to co-located teams, the reference

[19] A relative performance gain of distributed teams over co-located teams can be driven either by distributed teams becoming more productive or co-located teams becoming less productive or a combination of both dynamics.

category.[20] X_{it} is a vector of control variables and ε_{it} is the error term, which we cluster at the project level. The coefficients of interest are β_4 and β_5, which estimate the differential effect of the pandemic on projects with two and three different locations, respectively. Based on our intuition and previous results we expect both coefficients to be positive, indicating relative performance gains for mixed and distributed teams compared to those that are co-located.

Equation (2.2) estimates the causal differential impact of COVID-19, if activity patterns of co-located, mixed and distributed teams would have followed parallel trends in the absence of the pandemic.[21] We probe the robustness of this assumption by adding a rich set of controls, documenting insignificant pre-pandemic trends, using matching methods and performing a placebo analysis. If we further assume that the idiosyncratic COVID-19 shocks of team members are independent of team types we isolate the effect of being geographically co-located or distributed. In addition to adding controls and applying matching methods, we explicitly probe the plausibility of this assumption in Section 2.4.2 which explores potential mechanisms and alternative explanations.

Control variables include month fixed effects, project age (measured in months, linear and quadratic), the number of commits and watchers within the first year, the country of the project owner and projects' starting year to control for potential systematic differences in team composition across project vintages. Again, we allow all of these controls to have time-varying effects before and after the onset of the coronavirus pandemic, except for project age and month fixed effects. In our most restrictive specification we add project and time fixed effects, which replace the main effects of our controls but not the interaction terms with COVID_t.

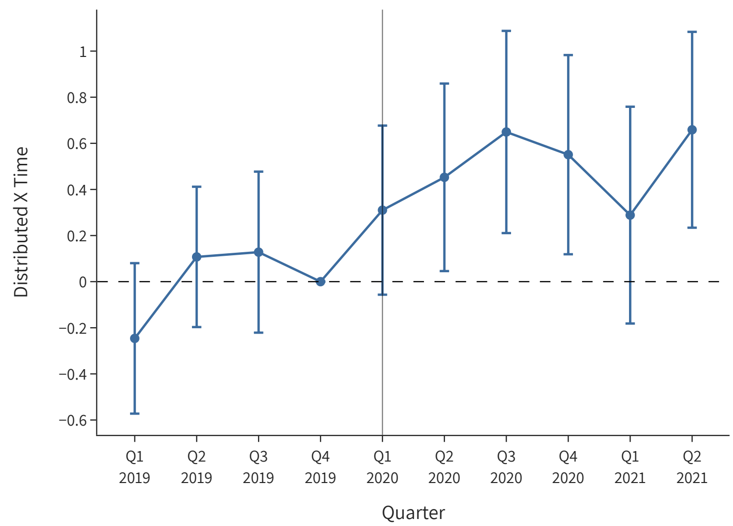
[20]The previous analysis has revealed that co-located teams are affected the most by the coronavirus. Therefore, we choose co-located as reference category since this allows us to compare their performance to both mixed and fully distributed teams.

[21]This assumption differs from the one of Equation (2.1) as it requires parallel trends *across team types* rather than within team types but *across project vintages*.

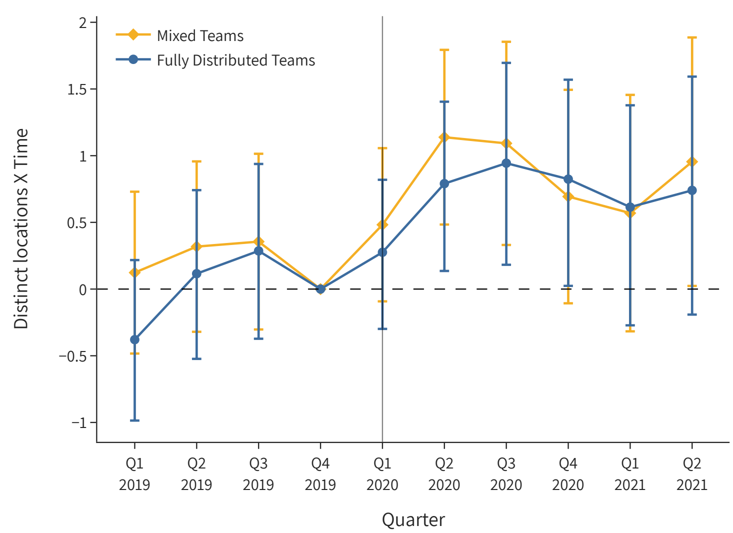
Results

Figure 2.5 *Quarterly effects of team distribution*

(a) *Co-located vs. distributed*



(b) *Number of locations*



Notes: The figures shows interaction coefficients and 95% confidence intervals of quarterly time dummies and indicators for distributed teams. In all regressions the dependent variable is the number of commits and controls include project age (linear & quadratic) and starting year, number of commits and watchers within a project's first year and country-of-owner fixed effects. Panel (a) compares co-located teams to distributed teams (two and three locations). Panel (b) compares co-located teams to mixed teams and fully distributed teams. Estimation method: Negative-Binomial ML.

Figure 2.5 shows the quarterly performance of distributed teams relative to co-located teams in two event study graphs and Figure 2.5a pools mixed and fully distributed teams together. While pre-pandemic coefficients are close to zero[22] and insignificant, we observe a clear positive shift in relative performance of distributed teams starting with the onset of the pandemic and turning significant from the second quarter of 2020. Figure 2.5b estimates relative performance of mixed and fully distributed teams separately. The patterns for both types of teams look very similar, but mixed teams seem to have slightly larger relative performance gains than fully distributed teams. However, this difference is not statistically significant and fades in the fourth quarter of 2020. Taken together, Figure 2.5 confirms our intuition that distributed teams are more resilient to the COVID-19 shock than co-located teams. Based on our results in the previous subsection, these relative performance differences during the pandemic are more likely driven by productivity losses of co-located teams rather than productivity gains of distributed teams.

Table 2.3 shows detailed regression results. In all specifications, co-located teams are the reference category and the dependent variable is the number of monthly commits.[23] The coefficients for mixed and fully distributed teams (with two and three distinct locations, respectively) show only small differences in (pre-pandemic) productivity compared to co-located teams. The interaction of COVID_t and the dummies for mixed and fully distributed teams confirm the patterns observed in Figure 2.5. Teams which are at least partially distributed did statistically significantly better during the pandemic than co-located teams. Across specifications, fully distributed teams did best, while mixed teams were in between. Note, that this may

[22]The point estimate of the coefficient for Q1 2019 is a negative insignificant outlier. It is driven by very "young" projects, which seem to be slightly more productive when team members are co-located. Removing projects started either after June 2018 or alternatively after October 2018, changes this estimate to nearly zero (June 2018: 0.007; October 2018: 0.036) while leaving the remaining coefficients largely untouched.

[23]Since we interact COVID_t with multiple control variables including project-founding-year fixed effects, we can only interpret relative differences between team types but not the main effect of COVID_t, which we therefore omit in Table 2.3. In column (1), without any controls, the coefficient of COVID_t captures both the negative effect of the pandemic on co-located teams and the general decline in activity as projects mature.

Table 2.3 *Collaborations by team distribution*

	Dependent: Number of commits					
	(1)	(2)	(3)	(4)	(5)	(6)
D(Loc.=2)	–0.027	–0.009	0.022	0.009	–0.004	
	(0.104)	(0.106)	(0.107)	(0.107)	(0.103)	
D(Loc.=3)	0.132	0.157	0.189*	0.177*	0.142	
	(0.095)	(0.097)	(0.099)	(0.099)	(0.098)	
D(Loc.=2) × COVID	0.377**	0.320**	0.296**	0.295**	0.298**	0.356**
	(0.150)	(0.145)	(0.144)	(0.145)	(0.142)	(0.147)
D(Loc.=3) × COVID	0.587***	0.514***	0.492***	0.472***	0.418***	0.486***
	(0.151)	(0.149)	(0.141)	(0.137)	(0.135)	(0.143)
Project age		X	X	X	X	X
Project year FE		X	X	X	X	X
N. Commits (first year)			X	X	X	X
N. Watchers (first year)				X	X	X
Owner country					X	X
Month FE		X	X	X	X	
Project FE						X
Time FE						X
Pseudo R^2	0.008	0.016	0.057	0.059	0.063	0.518
N	102,720	102,720	102,720	102,720	102,720	102,720
Clusters	4,280	4,280	4,280	4,280	4,280	4,280

Notes: Co-located teams are the left-out category. Project age is included in linear and quadratic form. Except for project age and month fixed effects, all controls are interacted with the COVID indicator. Column (6) includes project and time fixed effects, project age (linear & quadratic) and the interacted remaining controls. The sample covers 24 months centered around the onset of the coronavirus pandemic in March 2020. Estimation methods: columns (1) to (5): Negative-Binomial ML; column (6): Poisson ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

seem to contradict the order of effects seen in Figure 2.5b at first. However, the regressions estimate an average effect over the entire post-pandemic period such that the initial relative increase in productivity for mixed teams can be too small to determine the overall effect.

Starting from the simple regression in the first column, the specifications in Table 2.3 add more and more controls until column (5) which includes the full specification of interacted controls. Column (6) replaces most of their main effects with project fixed effects[24] and includes time fixed effects instead of month fixed effects, which improves the explained

[24]The interaction terms of controls and COVID_t are still included.

variance (Pseudo R^2). Across all specifications our main results are stable, both in magnitude and significance. Our most restrictive specifications in columns (5) and (6) suggest that during the pandemic mixed teams on average received 34% to 43% more code contributions than co-located teams, while fully distributed teams received even 52% to 65% more contributions.[25]

In the Appendix we probe the robustness of our main findings with two additional analyses. First, we re-estimate Equation (2.2) after applying inverse probability weighting and one-to-one propensity score matching. Analog to the approach in Section 2.4.1 we use logistic regressions to estimate three propensity scores including just controls, controls and pre-pandemic outcome levels, or controls and pre-pandemic trends.[26] We combine mixed and fully distributed teams into one category ($Distributed_i = 1$). This simplifies the matching process and the differences between these projects are not significant anyway. Figure B.3 in the Appendix shows the balance of control variables after applying the individual approaches. All methods perform well and significantly improve the balance between co-located and distributed teams. Table B.7 summarizes the regression results and column (1) shows a baseline specification without any matching applied. The interaction coefficient is highly significant and the estimate of 0.372 lies in between the respective coefficients for mixed (0.298) and fully distributed (0.418) teams from Table 2.3 column (5). Across all matching methods the interaction coefficients stay highly robust and similar in magnitude to the baseline specification in column (1). Applying one-to-one matching on controls and pre-pandemic outcome levels or trends reduces the statistical significance to the 5% level. However, the sample size is also roughly cut in half, which lowers estimation power. Overall, Table B.7 shows that our results are robust to matching and suggests a relative performance gain of about 45% for (at least partially) distributed teams over co-located teams during the pandemic.

[25]These percentages are calculated using the following formula: $(e^{\beta} - 1) * 100$.

[26]Controls include project age (at the beginning of the observational period), project starting-year, the number of commits and watchers in the first year as well as country of the project owner. Pre-pandemic outcome levels are measures as quarterly number of commits and pre-pandemic outcome trends are measured as first-differences of quarterly levels.

Second, one might be concerned that distributed teams are systematically more likely to work on long-term projects than co-located teams and that our estimates capture this dynamic rather than exceptional differences caused by COVID-19.[27] If this was the case, we should already observe similar effects in the years before the coronavirus pandemic. We address this concern in Table B.8, which summarizes results of placebo tests around alternative pandemic dates. In columns (1) and (2) we consider treatment effects if COVID-19 had happened one year earlier. We remove all projects created after 2017 and set our COVID_t indicator to one from March 2019 on.[28] Otherwise, we re-estimate our main specification without controls in column (1) and with the full set of interacted controls in column (2). Columns (3) and (4) repeat this exercise, shifting the onset of the pandemic backwards an additional year to March 2018. Throughout, the interaction coefficients for mixed and fully distributed teams are insignificant, indicating no systematic differences between co-located and distributed teams. These results are therefore reassuring that our main specification captures COVID-19 induced differences in productivity.

Mechanisms and additional results

So far we have consistently documented a differential impact of the coronavirus pandemic on co-located, mixed and distributed teams, resulting from a negative productivity shock of co-located teams. We argue that this is largely driven by co-located teams having lost their comparative advantage through offline collaboration, while distributed teams already had a functioning online production process in place before the pandemic. In order to explore underlying mechanisms and dismiss alternative explanations, Table 2.4 adds additional control variables to the baseline specification in column (1). For each additional variable, an

[27]This concern does not apply to our design in Section 2.4.1, which estimates COVID-19 effects *within* and not *across* team types (co-located, mixed and fully distributed).

[28]We again consider an observational period of 24 months centered around the hypothetical onset of the coronavirus pandemic and thus exclude the true COVID-19 period by keeping only observations until February 2020.

interaction term with COVID_t is added as well, to understand its role during the pandemic.

Table 2.4 *Collaborations by team distribution: mechanisms*

	Dependent: Number of commits					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.698	−0.085	1.109**	−1.170*	−0.453	
	(0.577)	(0.539)	(0.554)	(0.614)	(0.556)	
COVID	−0.402	−1.085***	−0.252	−1.616***	−1.598***	
	(0.264)	(0.276)	(0.257)	(0.424)	(0.412)	
D(Loc.=2)	−0.004	−0.044	−0.032	0.015	−0.060	
	(0.103)	(0.101)	(0.101)	(0.099)	(0.098)	
D(Loc.=3)	0.142	0.071	0.003	0.102	−0.053	
	(0.098)	(0.096)	(0.097)	(0.093)	(0.094)	
D(Loc.=2) × COVID	0.298**	0.294**	0.250*	0.339**	0.266*	0.333**
	(0.142)	(0.149)	(0.145)	(0.149)	(0.154)	(0.147)
D(Loc.=3) × COVID	0.418***	0.412***	0.337**	0.399***	0.324**	0.408***
	(0.135)	(0.143)	(0.139)	(0.139)	(0.150)	(0.145)
Experience		0.143***			0.114***	
		(0.013)			(0.013)	
Experience × COVID		0.088***			0.087***	0.075***
		(0.016)			(0.017)	(0.017)
N. contributors			0.066***		0.056***	
			(0.006)		(0.005)	
N. contributors × COVID			0.011		0.005	0.007
			(0.007)		(0.006)	(0.004)
Message length				0.583***	0.250***	
				(0.056)	(0.059)	
Message length × COVID				0.319***	0.171**	0.144
				(0.073)	(0.075)	(0.081)
Project FE						X
Time FE						X
Pseudo R^2	0.063	0.095	0.105	0.088	0.138	0.520
N	102,720	102,720	102,720	102,696	102,696	102,696
Clusters	4,280	4,280	4,280	4,279	4,279	4,279
Means		Experience	N. contributors	Message length		
D(Loc.=1)		7.152	4.460	3.774		
D(Loc.=2)		7.300	5.179	3.815		
D(Loc.=3)		7.582	6.594	3.932		

Notes: *Experience* is the log average contributor's experience, measured as the number of commits to any project before the coronavirus pandemic. *N. contributors* indicates the total number of pre-COVID contributors to a given project. *Message length* indicates the log average number of characters of pre-COVID commit messages. Co-located teams are the left-out category. All regression control for project age (linear & quadratic) and starting year, number of commits and watchers within a project's first year, country-of-owner fixed effects and month fixed effects. Except for project age and month fixed effects all controls are interacted with the COVID indicator. Column (6) includes project and time fixed effects, project age (linear & quadratic) and the interacted remaining controls. The sample covers 24 months centered around the onset of the coronavirus pandemic in March 2020. Estimation methods: columns (1) to (5): Negative-Binomial ML; column (6): Poisson ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

One possible alternative channel is that experienced developers are more likely to join distributed teams and at the same time are more resilient to pandemic shocks than less experienced developers. We address this in column (2) adding a measure for a team's pre-pandemic experience: the log average number of commits by team members to any project. First, the averages at the bottom of column (2) reveal that co-located teams have the least experienced members, members of mixed teams are more experienced and fully distributed teams' members are the most experienced. Second, experience has a highly significant positive effect on a team's productivity. Furthermore, the pandemic increased the importance of experience a lot, as the coefficient of the interaction with COVID_t indicates, which is more than half the size of the main effect of experience. However, the interaction coefficients of mixed and fully distributed teams are virtually unaffected by the inclusion of experience.

A second alternative channel is that distributed teams, having an established online production process, are better in sourcing remote contributors from the international open-source community. This could give them an additional advantage over co-located teams in dealing with the pandemic. Column (3) includes the total number of pre-pandemic contributors to a given project. Distributed teams indeed receive contributions from a larger number of users than co-located teams and the regression results show that projects with more contributors are more active before the pandemic. However, the insignificant interaction suggests that this relationship was not altered by the pandemic, i.e. the "additional" existing contributors of more distributed teams are not the driving force behind their better performance during the pandemic.

Column (4) includes the length of commit messages (the log number of characters). With each code contribution, Git allows the contributor to write a message explaining the changes. We consider this variable a proxy for written (online) communication and documentation on GitHub, which should be more extensive in distributed teams. Mean values at the bottom of column (4) suggest that commit messages of mixed and distributed teams are 4% and 17%

longer than those of co-located teams, respectively.[29] The regression results show that longer messages are associated with higher productivity in general.[30] Further, the coefficient of the interaction term with COVID_t is more than half the size of the main effect of message length. This supports the idea that written communication has become more important for productivity during the pandemic.

Column (5) adds all additional explanatory variables at once. The combined regression shows that their individual contributions remain important. At the same time, the interactions between COVID_t and the dummies for the number of distinct locations remain statistically significant at (at least) the 10% level, with similar magnitudes to the baseline specification in column (1). Column (6) adds project and time fixed effects, which leaves the interaction coefficients of all three additional variables mostly unchanged, but the effect of message length loses its significance. In contrast, the interaction coefficients for mixed and distributed teams increase in magnitude and become statistically more significant than in column (5).

Overall, Table 2.4 shows that the additional factors are important determinants of team performance before and after the onset of the coronavirus pandemic. Nevertheless, they cannot explain the differential impact of the COVID-19 shock on co-located, mixed and distributed teams. Although we cannot entirely rule out other unobserved systematic differences between team types, our results strongly suggest that the proximity of team members plays a key role for the post-pandemic productivity divergence of co-located and distributed teams.

The results in Table 2.4 suggest that the number of contributors is a key factor for performance before the pandemic. We turn to the number of post-pandemic contributors next. If during the pandemic co-located teams fail to effectively move their production process into the virtual world, it is likely that some team member become entirely inactive due to the

[29]These numbers are calculated as follows: $e^{3.815-3.744} \approx 1.04$ for mixed teams and $e^{3.932-3.744} \approx 1.17$ for fully distributed teams.

[30]Note that if some teams simply had a habit of committing more often (i.e. smaller changes per commit), the relationship may be biased in the negative direction, leading to an underestimate of the importance of communication.

increased communication costs. As this problem is less likely to occur in (at least partially) distributed teams, we may observe a higher number of active contributors in these teams compared to co-located teams. In order to explore this channel, we change the dependent variable to *the number of active contributors* in Table B.9.[31]

Column (1) shows results for the full sample without any restrictions. Before the coronavirus pandemic, there was no size difference between co-located and mixed teams, but distributed teams received contributions from a statistically significantly greater number of users. During the pandemic mixed and distributed teams both were statistically significantly larger than co-located teams, which is consistent with our intuition. In column (2) we re-estimate the regression on a sample in which we only keep projects that were active, i.e. received at least one contribution, during the coronavirus pandemic. Excluding teams which became entirely inactive, which was more likely for co-located teams, these estimates resemble an intensive margin. In this sample the interaction coefficient for mixed teams is much smaller and statistically insignificant. Distributed teams are still statistically significantly larger than co-located teams, but the interaction coefficient is roughly cut in half. The main effect of distributed teams is virtually unchanged compared to column (1).[32] Taken together Table B.9 suggests that part of the reason for the better performance of distributed teams is that they were able to maintain a higher number of active contributors.

[31]The observational period still covers 24 months centered around the onset of the pandemic, but we sum the total number of pre-pandemic and post-pandemic contributors, respectively. Therefore, the data contains two observations per project, one before and one after the onset of COVID-19. We include the same configuration of interacted controls as in Equation (2.2) except for month fixed effects. Project age is measured at the beginning of the observational period.

[32]Evaluating the coefficients at the pre-pandemic (3) and post-pandemic (2) median number of contributors in co-located teams, suggests that before COVID-19 fully distributed teams had received commits from $(e^{0.252}-1)*3 = 0.86$ more contributors and this difference increased to $(e^{0.252+0.235}-1)*2 = 1.25$ during the pandemic.

2.4.3 The effect of COVID-19 on individual performance

The previous subsection has shown that teams with members in multiple locations were more resilient during the pandemic. This raises the question of why remote teams are better able to stay productive and, in particular, which team members drive this result. Therefore, we focus next on the effects of the pandemic on individual productivity, comparing remote workers to those who shared the same location with at least one other team member.

We analyze code contributions of project members who joined a project within the first year of its existence. In order to focus on active collaborations, we only include projects to which at least two members have made contributions during the observational period, i.e. 12 months before and after the start of the pandemic.[33] Further, we consider the consistency of a team's pre-pandemic activity for an additional sample split. We measure consistency by the number of months with contributions to a project in the 12-month period preceding the pandemic, i.e. from March 2019 to February 2020. In addition to analyzing the full sample we focus on the most consistently active projects, which we define as those projects that received contributions in at least nine months before the onset of COVID-19.[34]

[33]As before, we further exclude projects with more than 500 contributions within their first two years to exclude outliers and, in particular, commits by bots.

[34]The choice of which percentage of the most active projects to include is a trade-off between a high activity level, such that changes are more measurable than for (initially) rather inactive projects, and sample size for statistical power. Requiring 10 or more months of activity would keep 9.2% of repositories in the sample, whereas our choice of requiring at least 9 months keeps 11.9%. The results are qualitatively similar if we further restrict the sample to only include projects that received at least one commit each month prior to the pandemic. By contrast, if we include less active projects by only requiring at least 6 months of activity out of the preceding 12 months, the results are no longer significant and more similar to the unrestricted sample.

Empirical setup

To identify the differential effect of the pandemic on remote workers, we estimate the following equation:

$$c_{ijt} = \beta_0 + \beta_1 \text{COVID}_t + \beta_2 \text{Remote}_{ij} + \beta_3 \text{Remote}_{ij} \times \text{COVID}_t + X_{it}\delta + \varepsilon_{ijt}. \tag{2.3}$$

The dependent variable is the number of commits to a project i by user j in month t. Remote_{ij} indicates that the worker is in a different location than the other first-year project members and COVID_t is an indicator that is equal to one from March 2020 on and zero before. The interaction of Remote_{ij} and COVID_t thus estimates the differential impact of the pandemic on remote workers compared to co-located workers, the left-out category. Therefore, β_3 is the coefficient of interest. Note, that the effect identified in this analysis can be seen as a combination of individual and team-level (differential) effects. The regressions include project fixed effects and monthly time fixed effects, and we control for linear and quadratic project age. We further control for projects' starting year, the number of watchers and commits in the first year, as well as their interactions with COVID_t.

Results

We again start by showing quarterly effects in an event-study graph. Figure B.4 in the Appendix shows the quarterly coefficients separately for all projects in the sample for this analysis and only for those that were consistently active before the pandemic, as described above. The fourth quarter of 2019 is the reference category.

The blue line, indicating results for all teams, does not reveal a strong effect. The coefficients are mostly positive, but close to zero and insignificant. Furthermore, the coefficients during the pandemic are comparable to the small spike in Q3 of 2019, i.e. similar to the usual variation. The red line, by contrast, shows positive and significant effects for those projects that had consistently received contributions before the pandemic. Again, we observe

Table 2.5 *Individual productivity for co-located and remote workers*

	Full sample		Active pre-COVID	
	Num. commits (1)	Commits (yes/no) (2)	Num. commits (3)	Commits (yes/no) (4)
Remote	–0.062	–0.014	0.140	0.020
	(0.279)	(0.017)	(0.400)	(0.091)
Remote × COVID	0.076	0.002	0.551***	0.045*
	(0.130)	(0.006)	(0.157)	(0.027)
Controls	X	X	X	X
Project FE	X	X	X	X
Time FE	X	X	X	X
Model	Poisson	OLS	Poisson	OLS
(Pseudo) R^2	0.349	0.409	0.205	0.110
N	92,016	108,312	10,872	10,872

Notes: The dependent variable in columns (1) and (3) is the number of commits. In columns (2) and (4), the dependent variable is an indicator whether there was at least one commit. The first two columns include the full sample, while the last two columns restrict the sample to consistently active projects (which had commits in at least 9 out of the 12 months preceding the pandemic). Remote indicates that the worker is in a different location from the two other project members and COVID is an indicator that is equal to one in March 2020 and later. The sample includes 24 months centered around the start of the pandemic. The regressions include project fixed and time fixed effects. All regression additionally control for project age (linear & quadratic) and starting year, number of commits and watchers within a project's first year. Except for project age all controls are interacted with the COVID indicator. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

a small spike in Q3 of 2019, but the remaining pre-pandemic coefficients are very close to zero, strongly suggesting an absence of pre-trends. We cautiously interpret these results to suggest that remote workers became relatively more productive as a result of the pandemic compared to co-located members.

The corresponding Poisson regression results are shown in Table 2.5, in column (1) for the full sample and in column (3) for projects which were consistently active before the pandemic. Columns (2) and (4) focus on the extensive margin for the full and restricted samples, respectively. We replace the dependent variable with an indicator that is one if there was a contribution by the member to the repository in a given month and zero otherwise, estimating linear probability models (using OLS). In all four columns, the coefficient of the remote indicator is insignificant. While the sign is negative for the full sample and positive for the restricted sample, overall remote workers seem to be similarly active compared to co-located members before the coronavirus pandemic.

The coefficient of interest on the interaction of Remote_i and COVID_t is positive in columns (3) and (4). The effect in the full sample is insignificant, with a magnitude of slightly less than 8%. Column (3), by contrast, shows a sizable and significant increase of more than 70% in the productivity of remote workers for projects that were among the most consistently active before the pandemic.[35] In this restricted sample, the effect is thus somewhat larger than the general effect on the productivity of distributed teams estimated in Section 2.4.2. Column (2) shows no significant effect on the intensive margin for the full sample. Again, the corresponding effect for the most active projects in column (4) is larger, implying that remote members were 4.5 percentage points more likely to make at least one contribution to their projects in a given month in the 12 months after the start of the pandemic. The effect in this linear probability model is smaller than the one in column (3) and significant at the 10% level.

Overall, the results show no significant change for the full sample, but a positive effect on remote worker productivity among the most active projects, positively affecting both intensive and extensive margins. Due to the limited sample size, we interpret the results of this section cautiously as suggestive evidence that remote workers may be important drivers of the higher productivity of (partly) distributed teams. This is in line with the idea that communication plays a central role in all of our results and the forced move online may have led to better integration of remote workers into teams' communication.

[35] This percentage change is calculated from the estimation coefficient β as in the previous section: $(e^{\beta}-1)*100 = (e^{0.551}-1)*100 \approx 73$ percent.

2.5 Conclusion

The COVID-19 pandemic has led to disruptive changes in the labor market by forcing a large share of the workforce to work from home. This is expected to entail a persistent increase in remote work of various forms, such as hybrid work or fully distributed teams. We show that even if digitally skilled knowledge workers have a virtual production environment (GitHub) already at their disposal, geographic distance still matters. Hence, when organizations allow for work from home or tap the global talent pool by hiring remote workers, it is crucial to understand what makes distributed teams productive to exploit the full potential of remote work models. We use the exogenous shift in exposure to remote work modes that was caused by the coronavirus pandemic to investigate the determinants of successfully organizing team work across different locations.

Analyzing small teams of open-source programmers, we find that the productivity of co-located teams suffered as a result of the pandemic, while fully distributed teams' performance was remarkably stable. Mixed teams may even have become slightly more productive. Directly comparing post-pandemic productivity patterns of teams with different geographic distributions, suggests that remote work experience made distributed teams more resilient to the pandemic than co-located teams. These results are robust to including a rich set of controls, matching and placebo tests. Exploring mechanisms, we find that while experience and talent sourcing are important determinants of team performance and systematically vary among (unmatched) types of teams, they do not drive our main results. Digital communication, proxied by commit messages on the platform, is crucial for team productivity and is likely one channel of our main effects. Nevertheless, the entirety of our results suggests that the performance differences between co-located and distributed teams are not (largely) driven by pre-existing differences or selection. Finally, focusing on individual productivity, we find suggestive evidence that remote workers have become relatively more productive when everyone had to work online. This is in line with the idea that this work mode helped them become better integrated into communication, instead of missing out on water cooler talks and other informal discussions offline. Nevertheless, our suggestive results on remote workers can only provide a starting point. It remains for future research to identify the best ways to ensure that remote workers are not excluded from communication.

Taking advantage of fine-grained, real-time collaboration data, our study sheds light on which aspects matter for successful remote collaboration in teams. While co-located teams suffered during the pandemic, an optimistic view of our results implies that learning from teams with remote work experience and setting up effective online communication can help mitigate negative effects of the shift to remote work. Future research should investigate further tools and organizational practices to enhance the productivity of remote teams. A better understanding of how to keep both co-located and remote employees productive will help organizations thrive in a future of increasingly flexible work modes.

Chapter 3

It Might Be Time to Fix It: Management Styles Standing Still[1]

Abstract

I study how firms adjust the bundles of management practices they adopt over time, using repeated survey data collected in Germany from 2012 to 2018. Employing unsupervised machine learning, I leverage high-dimensional data on human resource policies to describe clusters of management practices (management styles). My results suggest that two management styles exist, one of which employs many and highly structured practices, while the other lacks these practices but retains training measures. I document sizeable differences in styles across German firms, which can (only) partially be explained by firm characteristics. Further, I show that management is highly persistent over time, in part because newly adopted practices are discontinued after a short time. I discuss two potential hindrances to the adoption of structured management, miscalculations of cost-benefit trade-offs and non-fitting corporate culture, which should be further investigated. In light of previous findings that structured management increases firm performance, my findings have important policy implications since they show that firms which are managed in an unstructured way will continue to underperform.

[1]The work on this chapter originated from earlier work together with Florian Englmaier (LMU) and Stefanie Wolter (German Institute for Employment Research).

3.1 Introduction

Good management matters for firm performance. This is a well established result in the economics literature (Bloom et al. 2014; Bloom and Van Reenen 2011). However, determining what constitutes good management remains a challenging task. The *design perspective* of management argues that, synergy effects between individual practices (Ichniowski and Shaw 2003) as well as contingency on the environment (Englmaier et al. 2022; Gibbons and Roberts 2013) make studying management a highly complex problem. Others highlight the *management as technology* aspect and argue that some practices are superior for all firms (Bloom et al. 2016). Regardless the perspective, reducing the dimensionality of management data is inevitable to go beyond assessing individual practices and analyze management as a whole. For this reason, recent literature has started to use machine learning (ML) in order to detect bundles of individual practices that firms employ, which can be interpreted as *management styles* (Bandiera et al. 2020; Englmaier et al. 2022). However, to the best of my knowledge, this approach has not yet been applied to panel data of management practices and thus little is known about the dynamics of such management styles.

This study analyzes dynamic developments of management. Using machine learning I describe which styles (bundles of management practices) are employed and how the adoption of these styles has developed over recent years.

I address my research questions by utilizing data from the *Linked-Personnel-Panel* of the German Institute for Employment Research (IAB). It is administered to German establishments and asks detailed questions about human-resource (HR) management instruments. The survey has been conducted four times from 2012 to 2018, a time without major economic crises but of increasing workplace digitalization.

In a first step I identify two management styles using Latent Dirichlet Allocation (LDA)[2],

[2]This machine learning algorithm was initially developed to identify topics in text data, but can also be applied to survey data.

which allows me to detect potentially complex correlation patterns and identify those practices that distinguish management styles the most. My results reveal that firms are mainly distinguished by the adoption of highly structured practices, such as development plans, employee surveys and target agreements. While one management style is characterized by the adoption of these practices, the other style lacks structured practices but retains employee training measures.

Second, I show that the adoption of these styles varies largely across German firms and describe how styles are distributed, based on firms' characteristics. I find that larger, non-manager-owned and multiplant firms yield the most structured management styles, which is in line with previous findings made with international data (Bloom and Van Reenen 2011; Englmaier et al. 2022).[3] In a second paper, I and coauthors, further show that management positively correlates with technology adoption and promotion of diversity, suggesting that the structured style is one part of modern corporate governance (Englmaier et al. 2023).

Third, exploiting the panel structure of my data, I analyze how firms adjust management styles over time. Overall, I report a striking absence of trends toward either of the two styles and the average number of practices stays remarkably constant, as well. Analyzing differentiated trends across firms I find suggestive evidence that the smallest firms slowly move toward more structured management styles. However, small firms are not able to fully catch up, leaving the gap to bigger firms sizeable. I further show that single-plant and owner-managed firms, both starting with very unstructured management styles, are unable to catch up to other firms. Even changes of ownership structure or managers do not systematically affect management styles. With an absence of trends across self-reported market competition categories, I find no evidence that competition increases the adoption of structured management practices. If anything, firms facing no competition have moved the

[3]My results differ in one regard. While the literature has identified market competition as one of the key drivers for the adoption of structured management (Bloom et al. 2015, 2016; Bloom and Van Reenen 2007), I cannot confirm this in my data.

farthest toward a more structured management style. I further show that although many firms adopt structured practices, they drop them again shortly after their introduction.

The observed rigidity in management styles has clear policy implications. Since structured management has been shown to positively affect firm performance (see literature review below), it is striking that firms which lag behind in this respect fail to catch up. Backed by additional results, I discuss potential obstacles of adopting structured management styles. I suggest that miscalculations of cost-benefit trade-offs and a mismatch of corporate culture could play a key role and propose directions for future research.

This study contributes to a comprehensive and still growing literature on management. I limit my review to a recent strand of this literature which empirically analyzes management at large scale and is most closely related to my analysis.[4]

While earlier studies on management were focussed on few firms and often single practices, researchers have started to collect more comprehensive data in the mid 2000s. The most influential studies are based on the World Management Survey (WMS), which systematically collects management data around the world (Bloom and Van Reenen 2007). This line of work has shaped the technology perspective of management, introducing a uni-dimensional measure called *management score*, that measures the degree to which structured management practices are in place. Cross-sectional evidence highlights that management scores vary considerably across but also within countries. Market competition, separation of ownership and control as well as multinational presence are associated with high levels of management scores (Bloom et al. 2014; Bloom and Van Reenen 2007). While cross-sectional differences are well documented the time-series dimension of management is only scarcely investigated. One exception is Bloom et al. (2016) who show that product market competition accelerates the adoption of structured management practices, widening the already existing gap in

[4] See Gibbons and Roberts (2013) or Bloom and Van Reenen (2011) for more comprehensive surveys of the literature.

management scores. Bloom et al. (2019) find that changes in the business environment (introduction of right-to-work laws) can affect management scores. They also show that the presence of large multinationals can lead to positive spillovers in management quality.

Further, the literature shows that differences across management have implications on productivity and firm performance. Bloom and Van Reenen (2007) show that management scores are positively associated with profitability and firm survival. Bloom et al. (2016) find that differences in management scores account for 30% of cross-country total factor productivity differences. Causal evidence from an RCT in India can be found in Bloom et al. (2012a), who show that adopting management practices leads to increased productivity, decentralization and better use of information technologies. Also, conducting an RCT with Mexican enterprises, Bruhn et al. (2018) document a positive causal effect of management consulting on total factor productivity and return on assets. These results highlight the importance of management for firm performance and thus the need to better understand differences in management across firms and how these differences evolve over time.

Most of the above-mentioned studies rely on ex-ante assumptions on whether practices are "good" or "bad". Recent contributions to the literature, as well as mine, loosen this assumption and employ machine learning to add back elements of the design perspective. Bandiera et al. (2020) are the first to apply LDA to management data, more specifically diaries of CEO activities. They show that CEO behavior differs considerably and that CEOs can be characterized either as "leaders" or as "managers". Regarding firm performance, neither CEO type is clearly superior but rather the matching of CEOs to firms matters, which aligns with the contingency perspective of management. Englmaier et al. (2022), which is most closely related to my work, apply LDA to Spanish survey data, explicitly allowing for complementarities between management instruments. In their analysis — similar to mine — two distinct management styles emerge, a highly structured and a less structured style. Analyzing the impact of the financial crisis in 2008, their results suggest that "good" management is contingent to the environment. The structured style performs well in times of an economic boom, but it makes firms less flexible to adjust to economic crises. Both studies demonstrate the value of using machine learning to measure and analyze management using cross-sectional data. I contribute to this by applying ML to panel data in order to study how management styles evolve over time.

The remainder of this paper is structured as follows. Section 3.2 describes the data source and preprocessing. Section 3.3 introduces my management measures and the machine learning algorithm I use to estimate management styles. In the second part of this section I describe the results and characterize management styles. In Section 3.4 I first correlate my management measures with firm characteristics and then analyze how management evolves over time. I discuss my findings in Section 3.5 and conclude in Section 3.6.

3.2 Data

In order to study management styles I use the Linked Personnel Panel (LPP) provided by the Institute of Employment Research (IAB). The LPP consists of matched employer and employee surveys which were conducted in four waves from 2012 to 2018. It covers between 765 and 1,219 German establishments per wave and is representative of German private sector firms with more than 50 employees. Establishment managers are asked to provide information on human resource (HR) practices covering four broad categories: (i) "HR planning and recruitment", (ii) "HR development", (iii) "Remuneration structure" and (iv) "Commitment, values and corporate culture". In the second part of the LPP, which I shortly cover in Appendix C.2[5], a random sample of employees working at the establishments is interviewed. Between 6,500 and 7,500 employees per wave provide information about experienced quality of work, work attitude and behavior and personal characteristics. Further,

[5]In an accompanying paper I, together with coauthors, use the employee survey to investigate the relationship between management and employee satisfaction (Englmaier et al. 2023).

the data is complemented by rich socio-economic indicators.[6]

I estimate latent management styles using data from the employer survey. I employ an unsupervised machine-learning algorithm that requires categorical data at a single common scale. Since the vast majority of the data is in binary form I transform the remaining data into binary indicators as well. All non-binary categorical variables are of a five-point agreement scale type, which I convert into two binary indicators: (i) an indicator for being to the "left" of the neutral position (disagreement) (ii) and an indicator for being to the right of the neutral position (agreement).[7] I split numerical indicators at the median-value and add two binary indicators for being above and below the median, respectively.

For estimating latent management styles I strictly stick to questions regarding actual management practices and disregard firm-level or employee-level outcome variables. This ensures that I do not force the algorithm to explain any of these outcomes, but solely detect latent management styles (bundles of practices). I restrict the data to practices that are featured (and unchanged) in all four survey waves. This way I analyze a constant set of management practices across time and my findings are not driven by changes in the survey design. Further, I calculate TF-IDF-like scores, which penalize frequent and infrequent practices, and exclude five practices with the lowest scores. These practices are not informative for detecting differences in management bundles across firms and should thus be excluded.[8] The algorithm requires an input matrix of complete cases. To deal with missing values, I first remove all firms with more than 10% missing values in any given wave and then remove all practices which are missing for more than 10% of firms. Remaining missing values are set to zero.[9]

[6]For a more detailed discussion of the survey and data refer to Kampkötter et al. (2016).

[7]I exclude indicators for the neutral position, because this position is of little informative value. Firms choosing the neutral position are covered by setting both of the remaining indicators to zero.

[8]For robustness, I re-estimate the LDA model using the full set 46 practices without removing frequent and infrequent occurrences. The results are not shown but are very similar.

[9]Only 0.3% of the answers are missing and set to zero.

The final input data to estimate latent management styles contains 41 binary variables.[10]

3.3 Management styles

This section briefly introduces my approach to construct management styles using machine learning. I then present and analyze the results in order to characterize management styles.

3.3.1 Estimating management styles

To reduce the dimensionality of the survey data I employ an unsupervised machine learning algorithm: Latent Dirichlet Allocation (LDA) (Blei et al. 2003). LDA is a hierarchical Bayesian factor model that was originally developed to discover topics in text data. However, the algorithm is also applicable to survey data and was initially introduced to the economics literature by Bandiera et al. (2020).

In the context of this study, I argue that a firm's management is a mixture of a small number of *latent management styles* which determine the adoption of *individual practices*. More specifically, the core idea of LDA is based on two distributions: First, a latent management style is a mixture distribution over individual practices, the *style-over-practices distribution*. Thereby practices carry *loadings* that determine which practices are the most prevalent and therefore most characteristic of each latent style. Second, the *firm-over-styles distribution* describes a firm's actual configuration of management practices as a weighted combination of latent management styles. I call these style weights *style intensities*. Intuitively, LDA estimates both distributions by detecting bundles of practices (management styles) which tend to appear together and at the same time discriminate across firms. Being unsupervised ML, an important advantage of LDA is that it detects patterns of co-occurrence without forcing practices or latent styles to explain any firm outcomes. Further, as Bandiera et al. (2020)

[10] An overview of these practices and related questions is provided in Table C.4 in the Appendix.

argue, LDA was developed to naturally handle high-dimensional data which enables me to detect potentially complex correlation patterns.

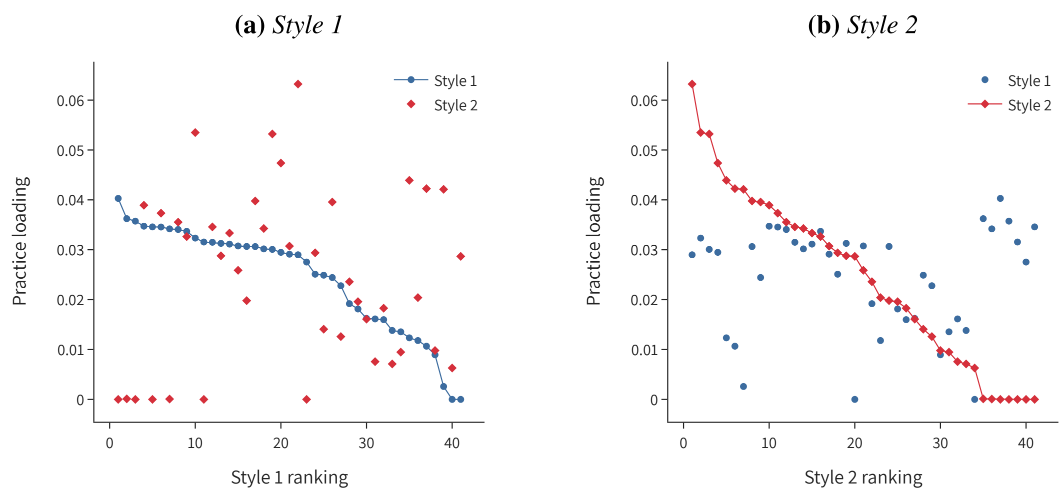
LDA requires the researcher to specify the number of latent factors (styles) to be estimated, and I set this number to two. I choose two styles for the following reasons: First, these latent styles are complex data objects which are not straightforward to understand. A low number of two latent management styles therefore facilitates the interpretability of my results, which according to Blei et al. (2003) should be taken into account. Second, LDA is a probabilistic classifier, which *does not* deterministically label firms but assigns each firm a linear combination of the two "pure" management styles. Therefore, the model retains a high degree of flexibility despite limiting the number of latent factors. Third, cross-validation shows no significant improvement of the model's fit when increasing the number of management styles.

Further, LDA requires priors on both of the Dirichlet distributions. I follow Bandiera et al. (2020) and Englmaier et al. (2022) in setting these. Similar to Englmaier et al. (2022) I assume that only few rather than many practices are characteristic of latent styles. To incorporate this concept in the model I choose a low prior of 0.1 for the style-over-practices distribution, which promotes sparsity. I am agnostic about the firm-over-styles distribution and thus choose a symmetric uniform distribution by setting the prior to 1.0. This initially distributes firms uniformly across the linear combination between the two latent styles. I estimate posteriors using the Gibbs sampling method based on 41 individual management practices and the pooled sample of 3,453 firm-year observations.

The analysis of management dynamics in Section 3.4 will be based primarily on the above described style estimates. However, I additionally construct a much simpler measure, which calculates the share of adopted practices (hereafter: SAP). While this measure provides a simple way to study *how many practices* firms adopt, the advantage of my main approach is that it additionally determines *which practices* distinguish firms from each other.

3.3.2 Describing management styles

Figure 3.1 *Practice loadings*



Notes: This figure shows differences in practice loadings across both styles. Each style is a distribution across 41 individual practices, all of them having a strictly positive weight, and with the sum of weights equal to one. In panel (a) practices are shown in a decreasing order of Style 1 loadings. In panel (b) practices are shown in a decreasing order of Style 2 loadings. The vertical axis shows the respective practice loadings.

Figure 3.1 describes the style-over-practices distribution. It plots practice loadings of both styles and panel (a) orders practices from highest to lowest according to their Style 1 loading. The relatively flat line reveals that many practices carry a similarly high loading. This means that a high Style 1 intensity can only result from the adoption of *many* of these high-loading practices. In contrast, Style 2 loads high on just a few practices as reflected by the comparatively steep practice-loadings curve in panel (b) of Figure 3.1. Therefore, the adoption of these *few* high-loading practices will lead to high Style 2 intensity. Table C.5 in the Appendix shows correlations of different management measures, and confirms this observation. A correlation of 0.6 between Style 1 intensity and the SAP indicates that firms with high Style 1 intensity tend to adopt more practices than those with high Style 2 intensity.[11]

[11]By constructions style intensities sum to one. Therefore, the correlation between Style 2 intensity and SAP equals -0.6.

Table 3.1 takes a closer look at which practices are the most characteristic of both styles. The top panel shows the five individual practices which carry the highest loadings in each style.[12] In the bottom panel practices are ordered by practice scores as suggested in Blei and Lafferty (2009), highlighting those practices with the largest difference in loadings across styles.[13] Style 1 practices include development plans, employee surveys and appraisal interviews, which all reflect a highly structured approach to management. Practices related to development plans and employee surveys are also those that carry the highest Style 1 scores compared to Style 2 scores. Figure 3.1 reveals that loadings of these practices are relatively high in Style 1 and at the same time almost zero in Style 2, meaning that these practices are highly differentiating between management styles. The top five list of Style 2 lacks structured practices but contains practices related to employee training. Practices with the highest Style 2 scores are mainly related to dealing with inefficient employees, but point toward a lack of structured ways to deal with these employees.[14]

Taken together Figure 3.1 and Table 3.1 suggest that Style 1 is characterized by the adoption of *many* management practices that lead to a *highly structured* approach to people management. Style 2, on the other hand, leads to the adoption of *fewer* and *less structured* management practices, but retains employee training measures.[15] Therefore, Style 1 seems to be more closely related not only to the SAP but also to the management score of the WMS than Style 2. However, at this stage I am agnostic about quality differences across these styles since my results simply reflect patterns in the data and are not forced to explain any differences in firm outcomes.

Now I turn to the firms-over-styles distribution. By construction of the LDA algorithm

[12]Table C.3 in the Appendix reports the full list of practices and their loadings in both styles.

[13]The disadvantage of this approach is that practices with high scores might still have relatively low loadings in both styles.

[14]For example, these firms dismiss inefficient people rather than reallocating them to better fitting jobs within the firm or taking other HR development measures.

[15]Since latent management styles are not ordinal, these interpretations are necessarily subjective.

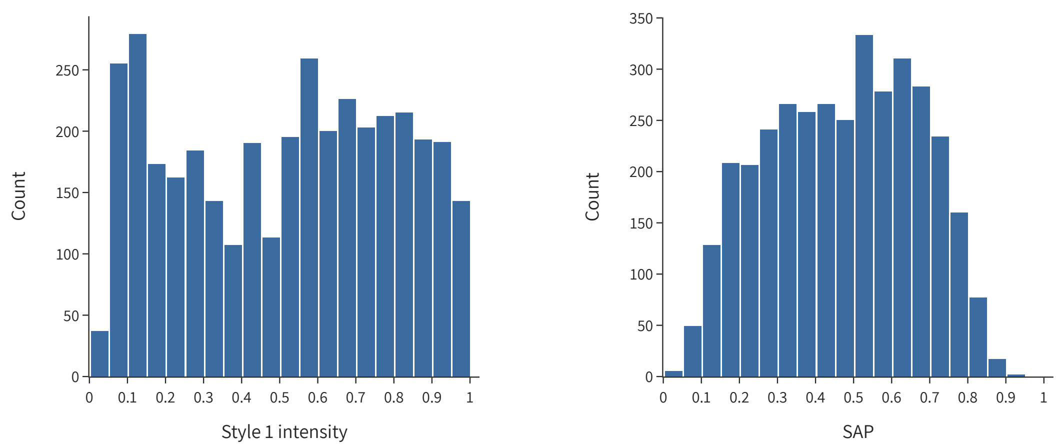
Table 3.1 *Most characteristic practices of both styles*

Rank	Style 1	Style 2
Ranked by practice loadings		
1	Development plans	Inefficiency: Discussions (high)
2	Employee surveys	Internal training
3	Development plans: Implementation	On-the-job training
4	Appraisal interviews	Attending lectures
5	Development plans: Management	Inefficiency: Dismissal (high)
Ranked by practice-loading scores		
1	Development plans	Inefficiency: HR development measures (low)
2	Development plans: Implementation	Inefficiency: Another position (low)
3	Development plans: Management	HR at highest management level
4	Employee surveys: Communicated to employees	Inefficiency: Dismissal (high)
5	Employee surveys: Develop solutions	Inefficiency: Discussions (low)

Notes: This table shows the most characteristic practices for both styles. The top panel ranks practices from highest to lowest according to loadings in Style 1 and Style 2. The bottom panel ranks practices according to TF-IDF inspired practice scores as suggested in Blei and Lafferty (2009).

Figure 3.2 *Distributions of management measures*

(a) *Style 1 intensity* **(b)** *Share of adopted practices (SAP)*



Notes: This figure shows histograms of both management measures using bins of size 0.05. Panel (a) shows the distribution of Style 1 intensity and panel (b) the distribution of the share of adopted practices. Both measures range from zero to one. The counts are based on the pooled sample of $N = 3,508$ firm-year observations.

style intensities are always positive and sum to one, which allows me to fully describe the distribution by focusing on Style 1 intensities. Figure 3.2 panel (a) plots the distribution of Style 1 intensities across firms. The distribution spreads across the whole range, indicating a good amount of variation of management styles across firms. Most firms employ a combination of both styles, but one can observe a slight tilt toward Style 1 with more mass to the right of 0.5. However, there is also a bunching region at very low Style 1 intensity levels, around 0.1. The average Style 1 intensity is 0.51 with a standard deviation of 0.29.[16] Panel (b) of Figure 3.2 shows the distribution of the SAP, which is more centered, i.e. few firms adopt a very small or a very large number of practices. The average firm has adopted 47% (about 19 out of 41) of practices and the standard deviation of the SAP is 0.19. Figure C.1 and Table C.2 repeat the exercise for the subset of panel firms and show similar patterns. However, the means of Style 1 intensity and the SAP are both slightly lower.

3.4 Results

This section describes the main results of the paper. First I document how management styles and the SAP correlate with firm characteristics. Then I show how firms adjust management over time.

3.4.1 Correlates of management styles

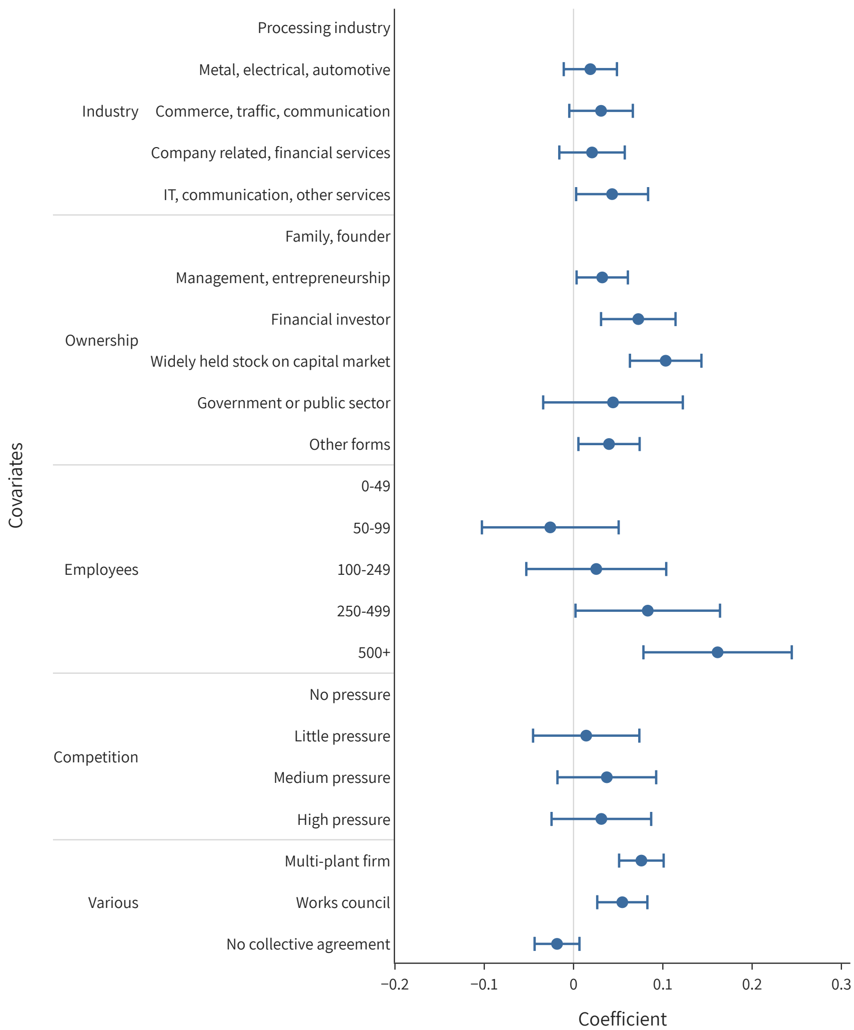
I explore correlates of firm characteristics with management styles to get an overview of the management landscape in Germany. For this, I estimate regressions of the form:

$$\theta_{it} = \alpha + X_{it}\beta + \varepsilon_{it}, \tag{3.1}$$

[16]Table C.1 in the Appendix shows summary statistics of management measures.

where θ_{it} refers to Style 1 intensity of firm i at time t, and X_{it} is a vector of firm characteristics.[17] Figure 3.3 shows coefficients and 95% confidence intervals of a multivariate pooled OLS regression. Table C.6 in the Appendix summarizes the corresponding univariate and multivariate regression results.

Figure 3.3 *Management Style 1 correlates*



Notes: This figure shows coefficients and 95% confidence intervals from an OLS regression of Style 1 intensity on firm characteristics. All regressors are either dummies or categorical variables. Reference categories are those without coefficient indicators. The regression is estimated on the pooled sample including all firm-year observations. The number of observations is $N = 3,508$ and standard errors are clustered at the firm-level.

[17]Summary statistics of firm characteristics are reported in Tables C.1 and C.2 in the Appendix.

I find relatively little variation in Style 1 intensity across industries, although the processing industry (the left-out category) seems to have the least structured management approach, indicated by positive and significant (at the 10% level) coefficients for all other industries. Table C.6 confirms this observation for the univariate case. Therefore, styles do not just reflect potential industry-specific management requirements. Instead, I observe considerable variation of management within industries.

Style 1 intensities differ across principal ownership and I can confirm earlier findings that family-owned firms tend to be managed in comparably unstructured ways (Bloom and Van Reenen 2007). The data offers additional ownership categories and results show that firms owned by financial investors or listed on the stock market have the highest Style 1 intensities. Bloom and Van Reenen (2007) argue that in theory the effect of a separation of management and control is ambiguous, since it allows selecting (potentially) more skilled managers but also introduces principal-agent problems. My results suggest that the positive selection effects predominate and a separation of ownership and control leads to more structured management styles.

Figure 3.3 shows that self-reported competition intensity does not affect style intensities, which is contrary to previous findings from international data (Bloom and Van Reenen 2007). This could indicate that selection effects or variations in incentives to provide (managerial) effort through competition play a less significant role in Germany than in other countries. However, in contrast to the measure in Bloom and Van Reenen (2007), competition in my data is self-reported and could thus be subject to heterogeneous reporting. If structured management leads to better performing firms, then their managers might systematically underestimate the pressure from competition. This could lead to the observed differences in results between self-reported and non-self-reported competition.

Larger firms — as measured by workforce size — lean toward Style 1, which reflects that these firms naturally require structured management to cope with the challenges of size. The same observation can be made for multi-plant firms which also show higher levels of Style 1 intensity. Again, given the increased organizational effort that multi-plant firms require, it is natural that these firms employ a more structured approach to management (Bloom et al. 2012b,c).[18]

Not surprisingly, firms with works councils or collective agreements have higher Style 1 intensities, since both reflect a structured approach to corporate governance in general. Note that the effect size in the univariate case in Table C.6 is much larger since both indicators correlate strongly with firm size.

In Figure C.2 and Table C.8 in the Appendix I re-estimate Equation (3.1) using the SAP as the dependent variable. Given the high correlation between Style 1 intensity and the SAP, the patterns are very similar: Non owner-managed, larger and multiplant firms have adopted the most practices leading to high SAP levels. There are two noteworthy differences. First, the gap between "IT, communications, other services" and the remaining industries is more pronounced than with management styles. One potential explanation is that this industry is very knowledge intensive and regular employee training measures, which load high in management Style 2, are required. This would lead to a larger number of adopted practices but at the same time keep Style 1 intensity comparably low. Second, medium and high market competition (self-reported) leads to statistically significantly higher SAP, which contrary to my results for Style 1 intensity is in line with previous findings. Taking these results at face value, they suggest that although firms which operate in competitive markets tend to employ a greater number of management practices, these are not necessarily structured practices.

Tables C.7 and C.9 repeat the regressions for the subset of panel firms, which I observe in every survey wave and use to estimate dynamics below. The observed patterns are qualitatively identical, however less significant due to the reduction in sample size. Overall, I find

[18]An alternative interpretation of these observations is that firms which employ more structured management styles grow faster and thus are larger and more likely to have multiple establishments.

significant and systematic differences in management styles (and the SAP) across firms which are largely in line with previous findings. These correlations corroborate my interpretation of Style 1 as being highly structured, since I observe that firms whose management I can describe as “naturally” structured are those with high Style 1 intensities. However, firm characteristics cannot fully explain the variation in management styles and thus the LDA model is able to capture systematic differences beyond those that I can readily explain with observables. Further, in the light of the positive effect of structured management on firm performance, theoretically all firms should benefit from a structured management style. Thus, I next analyze how firms change their management styles over time and whether a secular trend toward more structured management exists.

3.4.2 Dynamics of management styles

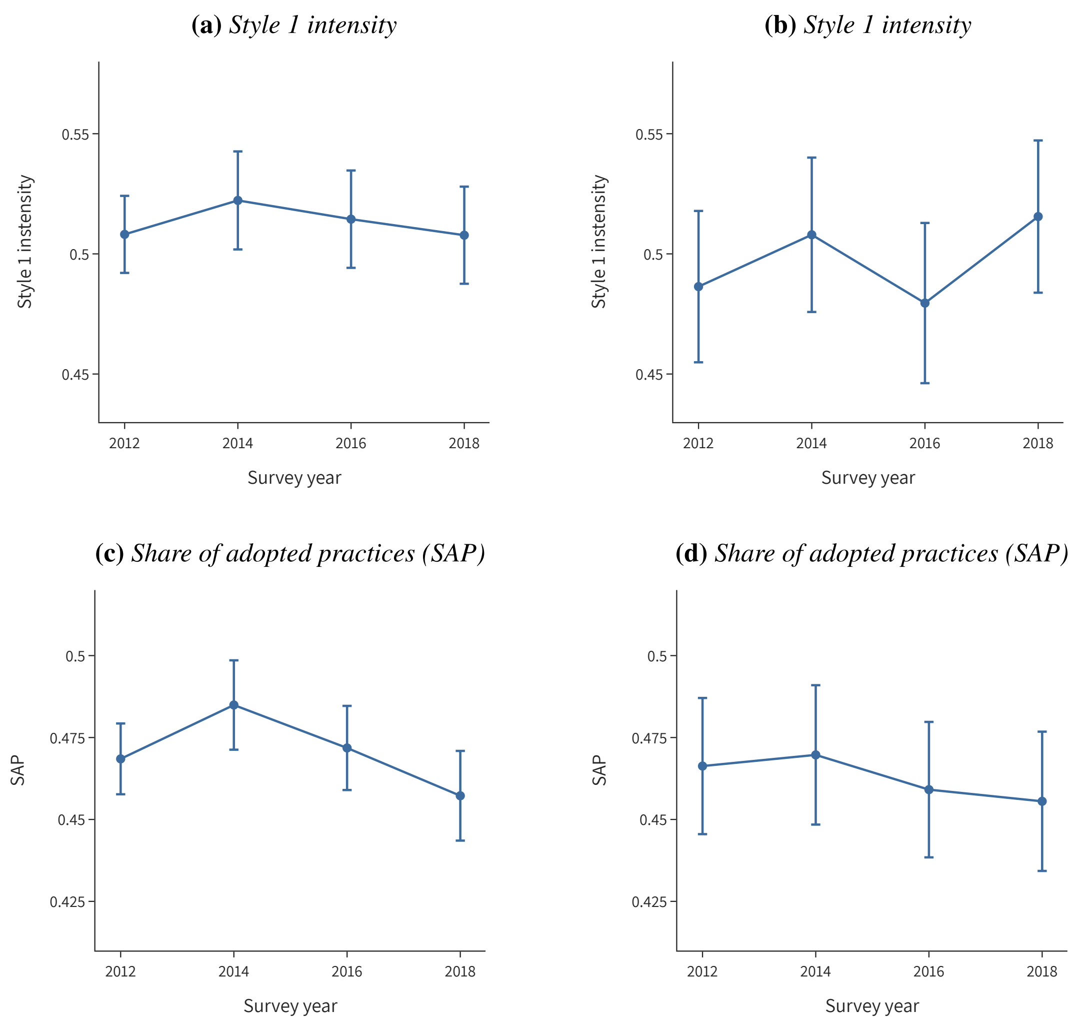
This section analyzes dynamics of management practices. First I describe overall trends and then investigate how subgroups of firms adjust management styles over time.

I begin by observing overall trends in management. Figure 3.4 shows mean Style 1 and SAP levels in each survey wave. Panels (a) and (c) employ the full sample and show trends in management across German firms. The average Style 1 intensity is 0.51 in the first survey wave and remains virtually unchanged across all other waves. A similar picture emerges with the SAP, however, after a small (statistically insignificant) increase in 2014 to 0.48 SAP levels seem to decline slightly until 2018 to 0.46.[19]

Analyzing the full dataset has the advantage that I observe a representative sample of German firms in each survey wave, allowing me to detect potential country wide trends. However, the sample composition changes in each survey year, which could dilute within-company developments. To address this, panel (b) and (d) restrict the sample to firms I observe in every wave, thus holding the sample constant across all years. Again Style 1 intensities

[19]This amounts to a reduction of on average 0.82 individual practices.

Figure 3.4 *Change of management*



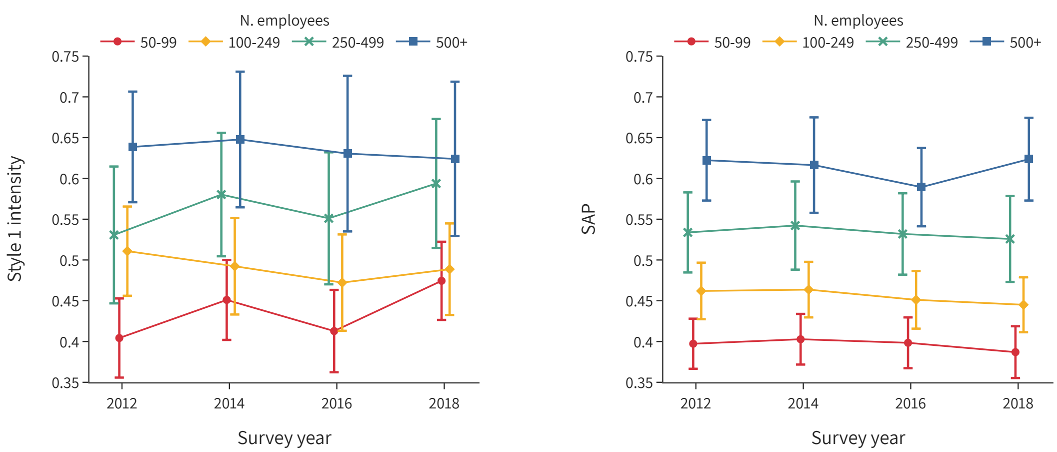
Notes: This figure shows Style 1 intensities (SAPs) across survey-years. Panel (a) and (b) show mean values and 95% confidence intervals of Style 1 intensity. Panel (c) and (d) show mean values and 95% confidence intervals of SAP. Panel (a) and (c) are based on the full sample, panel (b) and (d) contain only firms which I observe in every survey wave.

remain virtually unchanged over the whole period, although the point estimates increase from 0.49 to 0.52. The SAP remains constant throughout, indicating no adoption of additional practices.

Although I find a striking absence of country-wide trends in management styles, there might be differentiated developments in management styles across subgroups of firms, especially those for which I document large differences in levels. In this section I mainly analyze univariate relationships but provide multivariate regression estimations in Table C.12 in the Appendix. I estimate two versions of Equation (3.1): First, I set first-differences of Style 1 intensity and the SAP as dependent variable and include all firms which I observe for at least two consecutive survey waves.[20] Second, in order to capture long-term developments, I restrict the sample to panel firms and regress total differences (from first to last observation) in management measures on firm characteristics.[21]

Figure 3.5 *Change of management by number of employees*

(a) *Style 1 intensity* **(b)** *Share of adopted practices (SAP)*



Notes: This figure shows Style 1 intensities (SAPs) across survey-years split by initial number of employees category. The markers are slightly shifted to enhance the readability of the figure. Panel (a) shows mean values and 95% confidence intervals of Style 1 intensity. Panel (b) shows mean values and 95% confidence intervals of SAP. The figure is based on the panel sample of $N = 1,288$ firm-year observations of firms which I observe in every survey wave.

In Section 3.4.1 I have established that firms with a larger workforce employ a more structured management style than smaller firms. Figure 3.5 investigates whether this gap narrows over time. Similar to Figure 3.4 it shows average Style 1 intensities and SAPs across years, but separately for each workforce-size category. First, management of the largest firms

[20]This maximizes the number of observations but can only capture short-term developments.

[21]I use firm characteristics from the first observation of each firm.

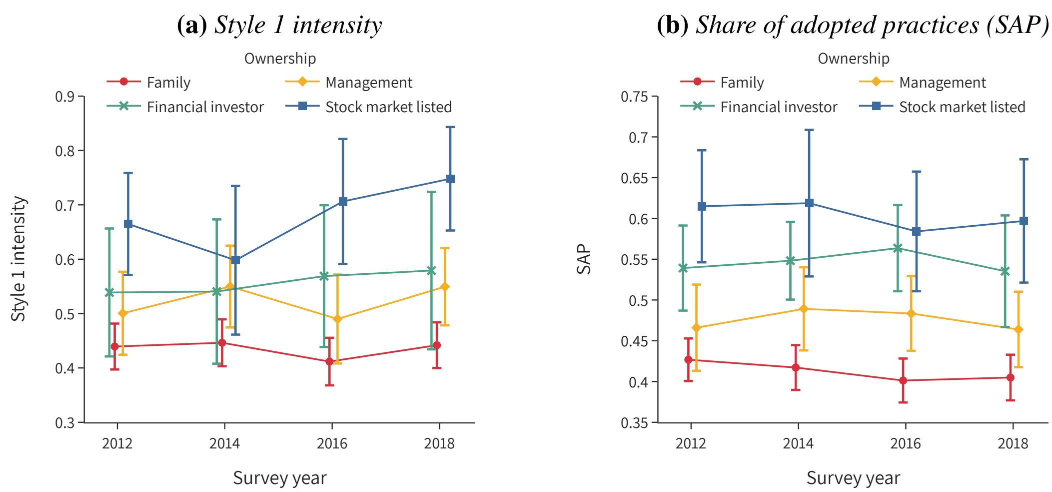
(squares) remains very stable at a high level of Style 1 intensity. Second, the smallest firms (circles) shift their management toward Style 1 indicated by a statistically significant increase in Style 1 intensity from 0.40 in 2012 to 0.47 in 2018. Third, the two medium size-groups remain fairly stable over time, however, one can observe a diverging pattern. While in the first survey wave Style 1 intensities of both size-categories are almost identical, the gap in point estimates widens from 0.02 to 0.10 (statistically significant) in 2018. This is mostly driven by medium-large firms (crosses) which slowly narrow the gap in Style 1 intensity to the largest firms in the sample. Medium-small firms do not change their management style and are caught up by the smallest firms in the last survey wave.

Panel (b) of Figure 3.5 depicts yearly SAPs. All four size categories show virtually no change in SAPs across years, i.e. the total number of adopted practices stays constant throughout. This means that the observed dynamics of Style 1 intensities are not driven by adopting additional practices or dropping practices which are already in place. Instead, firms seem to discard Style 2 practices in favor of more structured Style 1 practices.

One reason for firms to adopt structured practices could be that their workforce grows and thus requires a more structured management style. I investigate this in Figure C.3 in the Appendix. Panel (a) splits the sample into firms with an increasing workforce, i.e. those that move into a higher size category, and firms that either shrink or stay constant. Although the point estimates for growing firms lie slightly above others, there is no clear difference in trends. Panel (b) repeats the exercise for firms that start out in the smallest size category and again shows that trends are similar between growing and non-growing firms. Although I cannot entirely rule out that firms grow within categories, these patterns suggest that smaller firms adopt a structured management style deliberately rather than out of necessity as they grow.

Figure 3.6 investigates how firms have changed their management based on their ownership model. However, there are no notable dynamics since both management indicators remain

Figure 3.6 *Change of management by ownership*

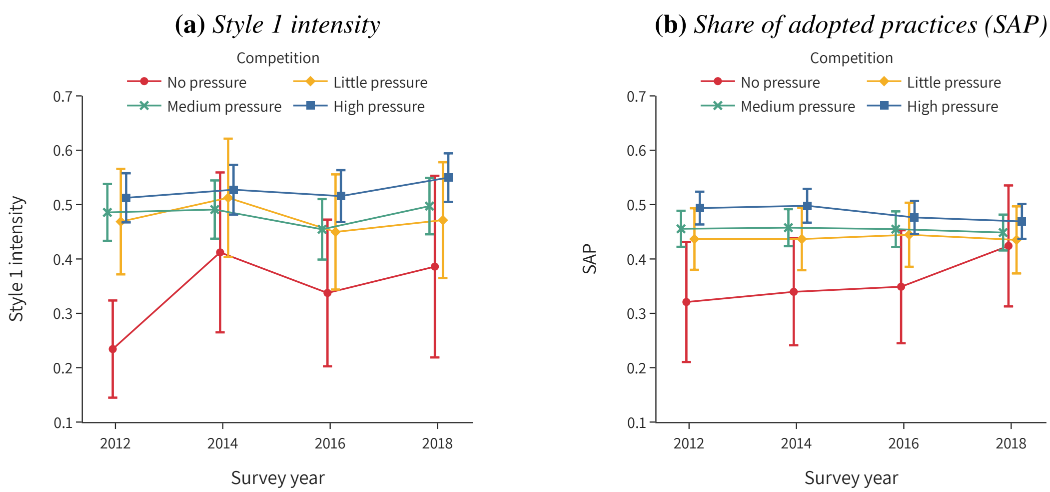


Notes: This figure shows Style 1 intensities (SAPs) across survey-years split by initial number of employees category. The markers are slightly shifted to enhance the readability of the figure. Panel (a) shows mean values and 95% confidence intervals of Style 1 intensity. Panel (b) shows mean values and 95% confidence intervals of SAP. The figure is based on the panel sample of $N = 1,288$ firm-year observations of firms which I observe in every survey wave.

fairly constant in all groups. If anything, stock-market-listed firms, which already start at a high Style 1 intensity, slightly increase the gap to all other ownership categories (not statistically significant).

A potential trigger of larger adjustments of management styles could be a change in ownership or managers. In each wave the LPP survey asks whether management or ownership has changed over the previous two years. Of the 322 firms which I observe in every year, ownership has changed in 75 (23%) firms and managers have changed in 171 (53%) firms. To investigate whether these changes affect management styles I estimate a regression of the total change in Style 1 intensity (SAP) on an indicator whether ownership or management has changed at least once during the observational period. Tables C.10 and C.11 summarize the results. Contrary to my expectations I find no significant effect on Style 1 intensity and changes of ownership seem to slightly reduce the number of adopted practices. I use the absolute change of the respective management measure in column (5) of both tables to estimate whether either event triggers adjustments but in varying directions. Again the coefficients are close to zero and statistically insignificant. One potential explanation for my results could be that out of the 75 firms in which ownership changes 43 remain within the same ownership category and only very few firms switch from any category to "Financial investor" or "Stock market listed". For the latter cases I would expect the largest changes in management style, since these categories show the highest Style 1 intensities in my cross-sectional analysis. However, it still remains puzzling why adopted management styles appear so rigid, even if new managers take over.

Figure 3.7 *Change of management by competition*



Notes: This figure shows Style 1 intensities (SAPs) across survey-years split by competition category. The markers are slightly shifted to enhance the readability of the figure. Panel (a) shows mean values and 95% confidence intervals of Style 1 intensity. Panel (b) shows mean values and 95% confidence intervals of SAP. The figure is based on the panel sample of $N = 1,288$ firm-year observations of firms which I observe in every survey wave.

Although, my cross-sectional results do not indicate management-style differences across self-reported competition levels, previous literature has identified competition as a key driver for improvements in management quality (Bloom et al. 2016; Bloom and Van Reenen 2007). Bloom et al. (2016) show that average management quality at the industry-country level increases over time in markets with high product competition and attribute this to a reallocation of market share from badly-managed to well-managed firms. Different to their setting, my data allows for analyzing within-firm dynamics. Therefore, I am able to investigate whether self-reported market competition induces firms to invest in "improving" their management isolated from distributive effects. Figure 3.7 shows the results.[22] Panel (a) reveals not only an absence of level differences, but also shows that facing stronger competition does not lead to the adoption of more structured practices. Firms facing no pressure from competition, appear to be an outlier as they start from lower levels but subsequently adopt structured practices. However, these results are based on very few firms and therefore should be taken with caution. Panel (b) of Figure 3.7 shows a very similar picture for the SAP. In sum, my results do not indicate that stronger competition causes firms to introduce a greater number of practices or more structured practices, but if anything indicate the opposite. Note again, that the competition measure in my data is self reported and could thus be biased if some managers systematically underestimate or overestimate the pressure from competition. Nevertheless, taking my results at face value, they suggest that previously found differences in average management quality could be mostly driven by redistributive effects rather than actual changes of firms' management styles.

Table C.12 in the Appendix includes differences in dynamics with respect to the remaining firm characteristics. Similar to above, I find no significant differences across groups and again management appears very stable over time. One exception is the processing industry which seems to slightly catch up to the "IT, communication, other services" industry.

In a final analysis I aim to better understand the origin of the observed management rigidity and turn my focus to dynamics of individual structured management practices. Are firms adopting structured practices but discontinuing them again quickly or are they not implementing these practices in the first place? Figure C.4 in the Appendix summarizes how firms adopt or drop the most characteristic practices of management Style 1. Panel (a)

[22] I divide the sample by competition in 2012.

is based on firms which had not adopted a given practice in the first survey wave (2012). For each of the practices listed along the y-axis, it shows the number of firms which never adopt the practice (gray), the number of firms which introduce the practice but drop it again (yellow) and the number of firms which introduce and keep the practice until the last survey wave (blue). Each of the practices is introduced at some point by more than 35% of the firms. However, between 40% and 50% of the adopting firms do not keep those practices until the end of the observational period. Panel (b) is based on firms which had adopted a given practice in the first survey wave and for each practice shows the number of firms which kept it throughout the whole period (gray), dropped but reintroduced it (yellow) or permanently (until the last survey wave) dropped the practice. Each of the practices is permanently dropped by about 30% of the firms, while a smaller share drops but reintroduces the practices. These patterns suggest that many managers try out introducing structured practices, but for some reason (see the discussion below) a lot of those managers decide not to stick with the practices. Revisiting Indian weaving firms nine years after their field-experiment Bloom et al. (2020) report similar patterns. Most firms had dropped a considerable amount of management practices that were introduced in the initial experiment. My results show that this also happens in non-experimental settings when management practices are not imposed by an outside party.

3.5 Discussion

The previous section documents sizeable differences in management styles, which can only to a minor part be explained by firm characteristics. Moreover, due to a striking absence of management dynamics, these differences are persistent. Part of the absence of trends can be explained by firms dropping recently adopted structured practices after only a short time. In light of previous research which consistently documents the positive impact of structured management on performance, across all types of firms, it is puzzling why no trend toward more structured management styles can be observed. It is particularly unclear why industries and firms that lag behind in terms of practice adoption fail to catch up. In this section I start to disentangle this puzzle by discussing potential mechanisms and offering direction for future research. To support my discussion I refer to additional results from Appendix C.2.[23]

A first factor which may play a role is that I observe management dynamics in times of high economic stability, 2012–2018, a period after Europe had largely recovered from the Great Financial Crisis and before the onset of the COVID-19 pandemic. If firms are successful, managers could be reluctant to initiate costly changes toward more structured management styles. However, considering recent findings by Englmaier et al. (2022), who show that structured management practices are especially valuable in economically thriving times, this constitutes a particularly harmful lost opportunity. Table C.15 in the Appendix further supports this notion by documenting that structured management styles are correlated with a higher likelihood of making profits. Extending my analysis to the most recent and future waves of the LPP will cover management data before, during and after the COVID-19 crisis. This offers an opportunity to analyze whether the challenge of a global pandemic, with its push toward more flexible and mobile work models, triggered larger adjustments of management styles.

Another potential hindrance to adopting structured management styles is that managers miscalculate the cost-benefit trade-off of introducing structured practices. While the costs of these practices — e.g. effort and time — are immediately noticeable, the benefits are likely not immediate, potentially indirect,[24] and hard to measure. This is consistent with both firms abstaining from adopting practices in the first place and firms abandoning management practices shortly after they have introduced them. As a first step to better understand the

[23] A description of the data and methods used to obtain these results is provided in the Appendix.

[24] For example, Table C.17 in the Appendix shows that more structured management is associated with higher levels of employee satisfaction and lower turnover intention, which is likely beneficial for firm performance in the long run (e.g. Böckerman and Ilmakunnas 2012; Halkos and Bousinakis 2010).

role of this mechanism, future surveys should include questions, which specifically ask why managers choose not to adopt structured practices and whether the associated cost-benefit trade-off plays a role. A second step is to further investigate why some firms adopt structured practices and others fail to do so, with a particular focus on who bears the costs of adopting these practices. Consider a situation in which these costs are not directly borne by those who decide whether to introduce structured practices. Although this carries the danger of adopting counterproductive practices[25], it could also help overcome the reluctance of adopting management practices which carry high up-front costs but are beneficial in the long run. Such a situation would create an advantage for bigger firms if they had multiple hierarchy levels and management decisions were made at the top management level but implemented mostly by lower management. My finding that larger and non-owner-managed firms employ relatively structured management styles, is consistent with the discussion above.

Lastly, I return to the design perspective on management. If management is contingent on firm characteristics and the environment, I may (in an extreme case) observe a steady state in which all firms have already chosen their optimal configuration of management practices, subject to an unobserved factor to which the observable configuration is maximally complementary. While this would explain the observed persistent level-differences in management styles, the combined findings in the literature raise doubts that this is the case. First, studies consistently document a positive relationship between structured management and firm performance, including both descriptive (Bloom et al. 2012b, 2016; Bloom and Van Reenen 2011; Englmaier et al. 2022, 2023) and causally identified findings (Bloom et al. 2012a, 2020). Table C.15 in the Appendix confirms that in my setting structured management positively correlates with a self-reported profit measure, as well. Second, my study and others show that firm characteristics and the environment can explain differences in management styles only to some degree. This suggests the existence of other, less researched, factors

[25] A negative example are the excessive documentation requirements in public institutions.

which are complementary to structured practices and thus make their introduction profitable for some firms but not for others.

One limiting factor for structured management can be a lack of digital tools and data skills to design and implement management practices as well as analyze their value. This especially concerns practices which aim to analyze a firm's workforce in a structured way, such as employee surveys or performance tracking. Table C.16 documents a strong positive correlation between structured management practices and the usage of digital tools, suggesting potential synergy effects.[26]

Another, and likely even more significant, potential precondition for the success of structured management is a suitable corporate culture. Good working relationships between employees and their (direct) supervisors could be especially important. The employee survey of the LPP, which I shortly introduce in the Appendix C.2, features several questions related to corporate culture and specifically asks about supervisor-employee relationships. Table C.18 shows conditional correlations of these corporate culture variables and my management measures. Structured management is positively associated with supervisors being perceived as fair, understanding, confident in their employees and offering good guidance. All of these factors are likely beneficial for the success of structured management practices. For instance, they could help employees to speak up and raise relevant issues in appraisal interviews or employee surveys. To this end, Castro et al. (2022) show that regular meetings (structured practice) with a particular focus on employees' individual needs and aspirations (corporate culture) lead to an increase in psychological safety and ultimately to higher team performance. Further, managers who are fair and offer good guidance should succeed in forming encouraging yet attainable target agreements and development plans. To the best of my knowledge, research specifically investigating synergies between corporate culture and structured management

[26]Although the related questions do not specifically ask whether these tools are used for HR management, they suggest a general digital and data-analytics competence of an organization.

practices is very scarce. One exception is Blader et al. (2019) who show that corporate culture can play an important role for the success of introducing performance tracking. However, their study only considers the introduction of this particular management practice in a single company. Therefore, analyzing synergies of corporate culture and other factors with structured management styles more broadly and at a larger scale constitutes a promising endeavor for future research.

3.6 Conclusion

In this paper I paint a picture of the dynamics of management styles in Germany. I analyze survey data on management practices provided by the German Institute of Employment Research, which offers a rare opportunity to repeatedly observe firms' management over a longer period of time. To reduce the dimensionality of the data I employ machine learning (LDA) and a simpler method of counting management practices. The advantage of LDA is that it is able to detect (potentially complex) underlying patterns and helps me identify practices which distinguish firms the most. My results show that firms differ the most in practices related to structured management such as development plans, employee surveys and appraisal interviews.

Using the low-dimensional representation of adopted management practices, I first analyze the cross-sectional dimension of the data to describe the German management landscape. My empirical findings show that investor-owned, large and multiplant firms tend to have the most structured management styles, confirming previous results in the literature of management (Bloom et al. 2019, 2014; Englmaier et al. 2022). Second, I exploit the panel structure of my data to investigate how firms adjust their management over time. Overall, I find that management is fairly rigid without any major trends in management styles. In particular, there is no secular trend toward more structured management. Analyzing these dynamics in more detail, I find suggestive evidence that small firms are able to slightly catch up, however the gap to larger firms remains big. My results further document some degree of experimentation with adopting practices, where in many firms recently introduced practices are not kept permanently.

My results have practical implications for other empirical studies of management, particularly for those that are based on cross-sectional data (e.g. Bandiera et al. 2020; Englmaier et al. 2022, 2023). Englmaier et al. (2022) rely on cross-sectional management data from 2006 to estimate effects on productivity over a time frame from 2001 to 2016. This temporal mismatch of management and outcome data will be less of a concern if management styles remain unchanged over time. Possible explanations as to why management styles remain rigid are offered in the discussion section of this paper. Future research should give a special focus to cost-benefit trade-offs of practices and to synergy effects of corporate culture. This study carries important implications for managers. Laying the groundwork for a successful and permanent implementation of structured management styles, managers can lead firms on the path of realizing their full potential.

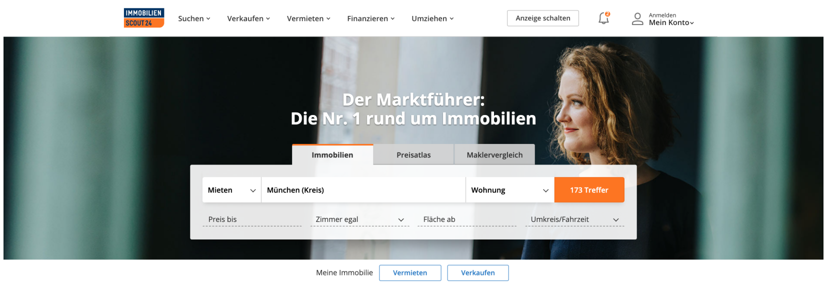
Appendices

Appendix A

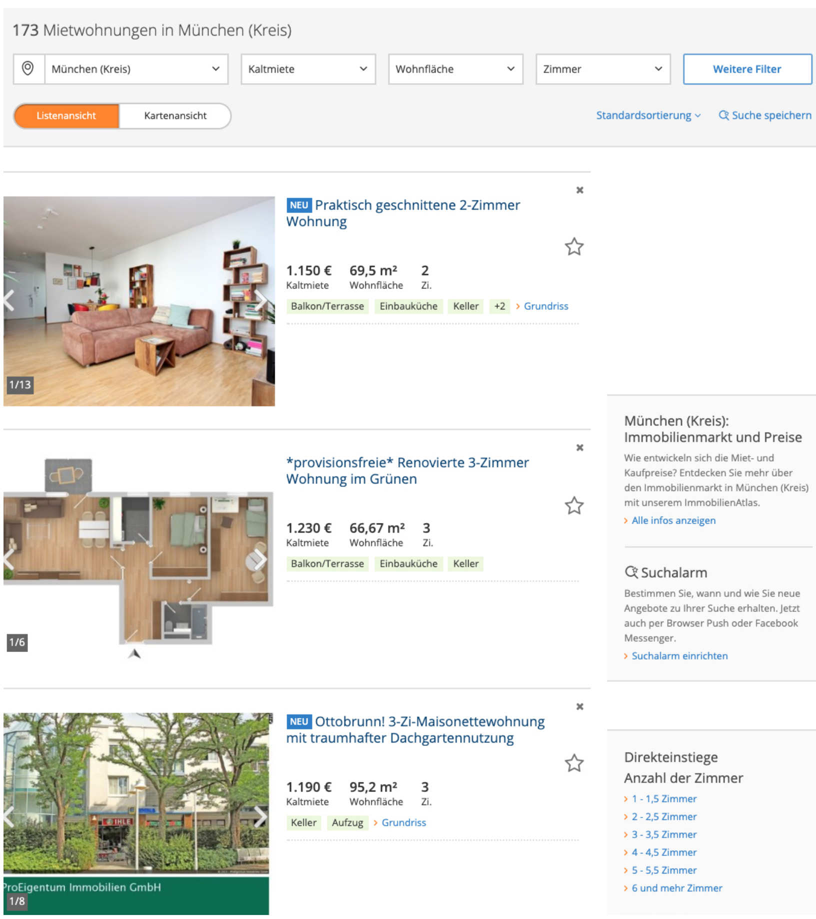
Appendix to Chapter 1

Figure A.1 *Immobilienscout24.de: website samples*

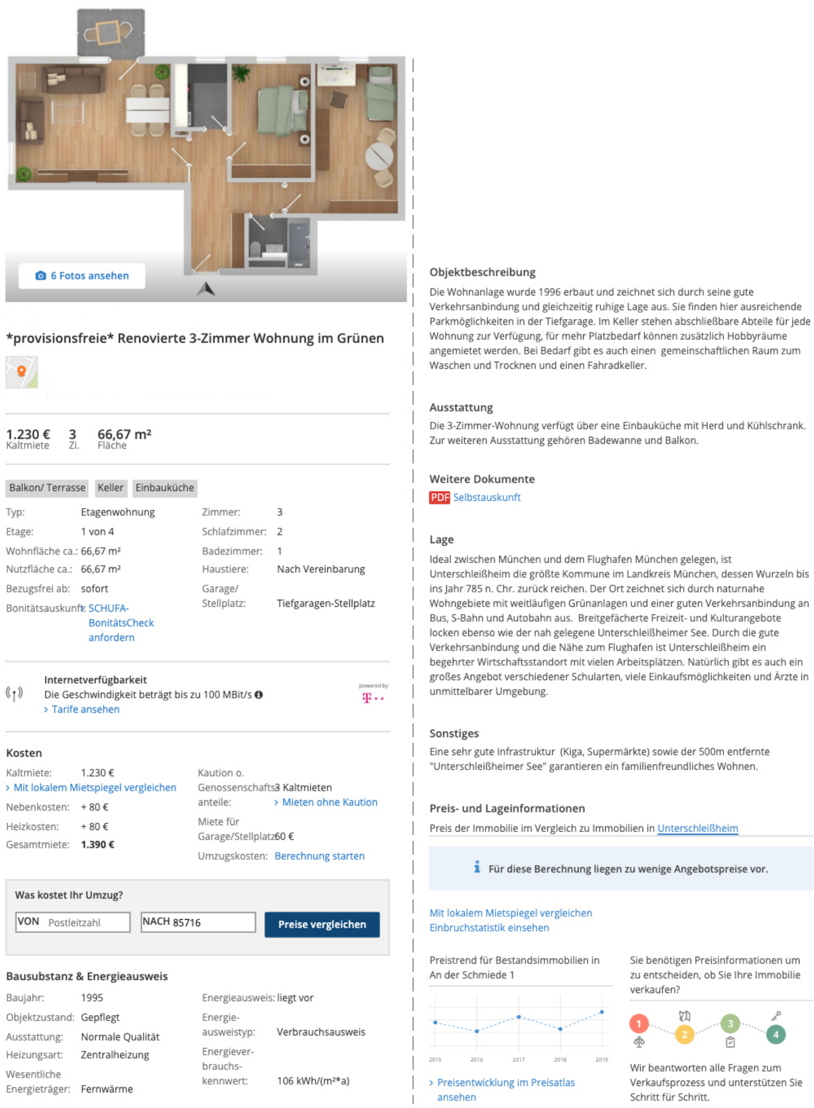
(a) *Search mask*



(b) *Result list*

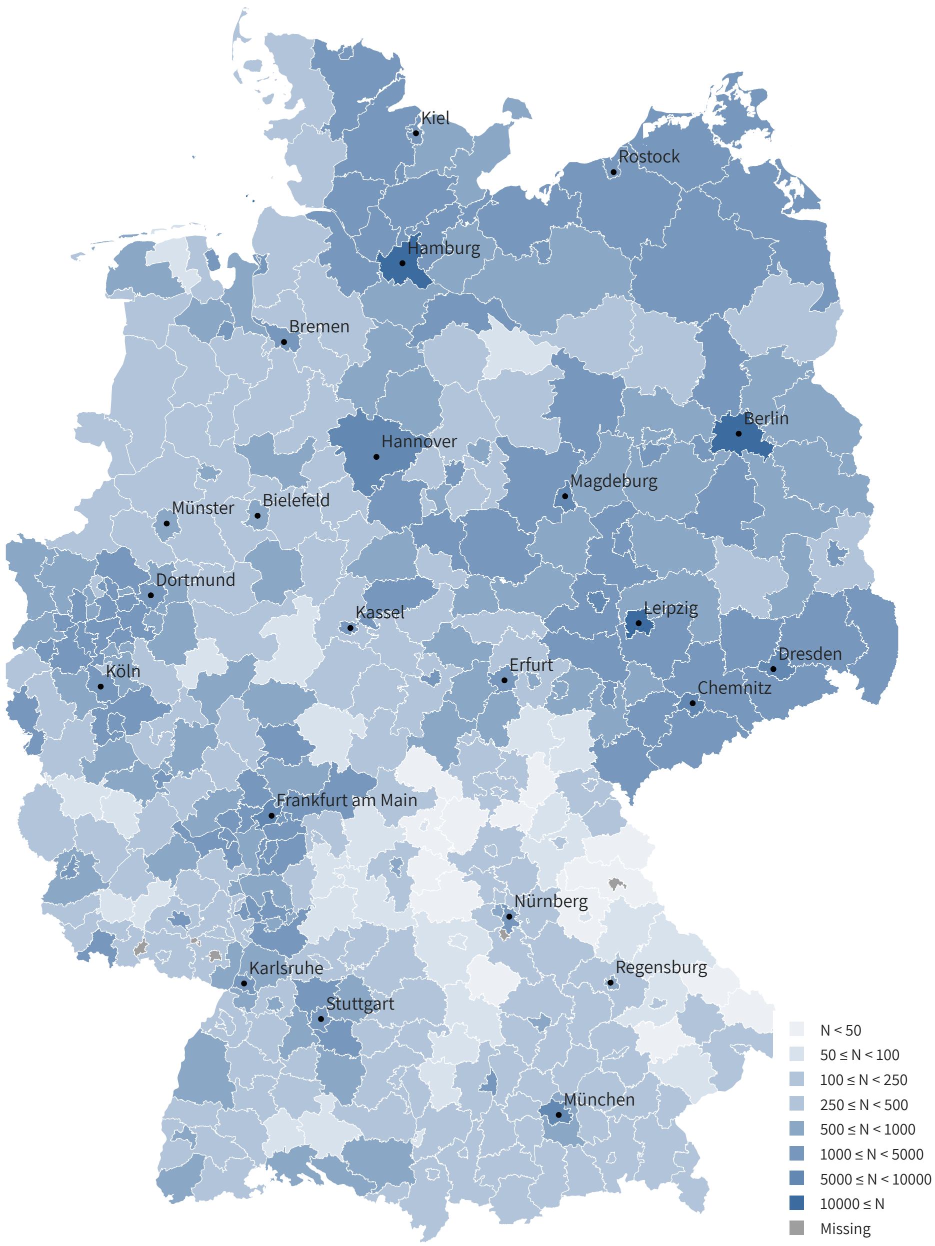


(c) *Detail page*



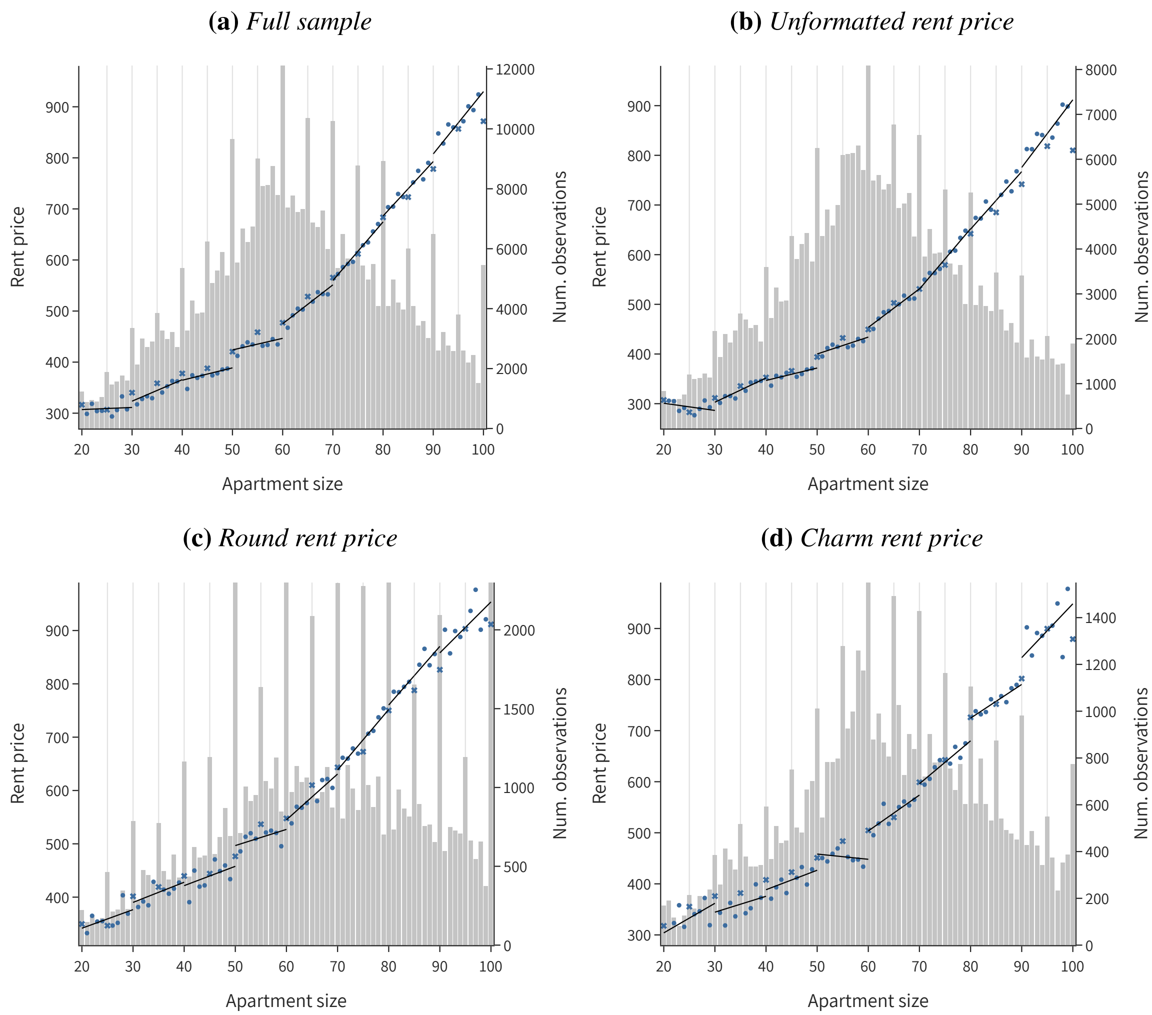
Notes: Shown are screenshots from immobilienscout24.de, illustrating the search process. Some (private) information, such as exact addresses or contacts to landlords or realtors, have been removed.
Source: immobilienscout24.de (accessed 12 Aug 2019).

Figure A.2 *Map of Germany: geographic distribution*



Notes: This map shows the geographic distribution of apartment ads across Germany, aggregated at the district level. Darker shades indicate a higher number of apartments in the respective district.

Figure A.3 *Average raw rent-price data*



Notes: Plotted are average rent prices within $1m^2$ bins. Solid lines represent linear fits between $10m^2$ multiples. Gray columns indicate the total amount of apartments per $1m^2$ bin in the respective subsample.

Table A.1 *Rent-price discontinuities: total and logged rent*

	Dependent variable: See below					
	Total rent			Log basic net		
	(1)	(2)	(3)	(4)	(5)	(6)
TightMkt × Left Digit		9.591***	7.653***		−0.001	−0.004**
		(1.034)	(1.451)		(0.001)	(0.002)
TightMkt × Round		23.154***	9.595***		0.014***	0.005*
		(1.502)	(2.135)		(0.002)	(0.003)
LL Type × Left Digit	22.381***		22.759***	0.029***		0.028***
	(5.135)		(5.112)	(0.006)		(0.006)
LL Type × Round	−16.736***		−18.776***	−0.016**		−0.019***
	(6.260)		(6.229)	(0.008)		(0.007)
Left Digit	5.900***	−2.147*	2.216	0.005**	0.001	0.008***
	(1.750)	(1.194)	(1.764)	(0.003)	(0.002)	(0.003)
Round	−1.035	−9.436***	−7.202***	0.006***	0.005***	0.001
	(1.504)	(1.045)	(1.647)	(0.002)	(0.002)	(0.003)
TightMkt		43.800***	36.204***		0.158***	0.141***
		(0.930)	(1.341)		(0.001)	(0.002)
LL Type	−55.244***		−3.478	−0.083***		−0.055***
	(4.100)		(4.403)	(0.005)		(0.006)
LL Type × TightMkt			−81.135***			−0.041***
			(4.838)			(0.006)
Adj. R^2	0.782	0.754	0.786	0.811	0.800	0.823
Observations	127,965	310,162	127,965	127,965	310,162	127,965

Notes: This table provides estimates using two alternative dependent variables: Total rent price and log basic rent price. Shown are coefficients from local-quadratic estimations using windows of size $10m^2$ centered at multiples of $10m^2$. *Left Digit* indicates whether the size measure includes the larger leftmost digit within each window. *Round* indicates whether the size measure is a multiple of $5m^2$. *LL Type* is the landlord type measure as constructed in Equation (1.3). *TightMkt* indicates tight local markets. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Table A.2 *Rent-price discontinuities: functional form of $f(Size_i)$*

	Dependent variable: Basic rent				
	(1)	(2)	(3)	(4)	(5)
TightMkt × Left Digit	5.951***	5.946***	6.068***	5.949***	5.929***
	(1.337)	(1.338)	(1.336)	(1.337)	(1.338)
TightMkt × Round	10.294***	10.307***	10.452***	10.305***	10.289***
	(1.970)	(1.970)	(1.972)	(1.970)	(1.970)
LL Type × Left Digit	20.229***	19.948***	19.593***	19.946***	19.864***
	(4.959)	(4.963)	(4.975)	(4.962)	(4.964)
LL Type × Round	−22.816***	−22.839***	−22.971***	−22.840***	−22.712***
	(6.111)	(6.112)	(6.119)	(6.112)	(6.112)
Left Digit	2.407	2.600	−2.050	2.598	2.619*
	(1.584)	(1.585)	(5.293)	(1.586)	(1.590)
Round	−4.400***	−4.328***	−4.396***	−4.573***	−4.629***
	(1.479)	(1.482)	(1.492)	(1.485)	(1.486)
TightMkt	30.994***	30.992***	30.904***	30.993***	30.997***
	(1.252)	(1.252)	(1.251)	(1.252)	(1.252)
LL Type	−6.062	−5.889	−5.698	−5.889	−5.859
	(4.119)	(4.120)	(4.128)	(4.120)	(4.121)
LL Type × TightMkt	−82.088***	−82.114***	−82.041***	−82.109***	−82.100***
	(4.583)	(4.584)	(4.584)	(4.584)	(4.584)
Functional form	See table notes				
AIC	1,599,948	1,599,949	1,599,929	1,599,949	1,599,929
Adj. R^2	0.768	0.768	0.768	0.768	0.768
Observations	127,965	127,965	127,965	127,965	127,965

Notes: This table summarizes regressions results using different functional forms of $f(Size_i)$. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%. The functional forms in columns (1) to (5) are:

(1) $f_1(Size_i) = \gamma_0 Size_i + \sum_{j=3}^{9} \left[\gamma_{1j} w_j + \gamma_{2j}(Size_i * w_j)\right]$

(2) $f_2(Size_i) = f_1(Size_i) + \gamma_3(Size_i * LD_i)$

(3) $f_3(Size_i) = f_1(Size_i) + \gamma_3(Size_i * LD_i) + \sum_{j=3}^{9} [\gamma_4 (LD_i * w_j) + \gamma_5 (Size_i * LD_i * w_j)]$

(4) $f_4(Size_i) = f_1(Size_i) + \gamma_3 Size_i^2$

(5) $f_5(Size_i) = f_1(Size_i) + \gamma_3 Size_i^2 + \sum_{j=3}^{9} \left[\gamma_{4j} \left(Size_i^2 * w_j\right)\right]$

Table A.3 *Rent-price discontinuities: alternative polynomial definitions*

	Dependent variable: Basic rent					
	(1)	(2)	(3)	(4)	(5)	(6)
$\geq 30m^2$	−15.478***	3.841*	−7.090***	−10.427***	−11.388***	−10.211***
	(1.713)	(2.177)	(2.592)	(2.715)	(2.715)	(2.849)
$\geq 40m^2$	−2.116	12.874***	9.315***	11.378***	1.056	1.697
	(1.324)	(1.594)	(1.619)	(1.720)	(1.909)	(1.928)
$\geq 50m^2$	−7.671***	1.039	3.006**	5.638***	4.587***	3.660**
	(1.155)	(1.161)	(1.217)	(1.335)	(1.321)	(1.446)
$\geq 60m^2$	−5.029***	−1.995*	2.299**	2.853**	8.060***	7.853***
	(1.085)	(1.022)	(1.157)	(1.134)	(1.245)	(1.223)
$\geq 70m^2$	7.328***	4.897***	9.378***	7.329***	9.756***	10.628***
	(1.219)	(1.307)	(1.318)	(1.518)	(1.482)	(1.634)
$\geq 80m^2$	21.993***	13.899***	15.596***	12.403***	5.330**	5.291**
	(1.458)	(1.809)	(1.744)	(1.948)	(2.244)	(2.250)
$\geq 90m^2$	27.296***	11.863***	7.347***	6.956**	0.076	−1.281
	(1.823)	(2.517)	(2.840)	(2.817)	(2.889)	(3.190)
Mult. $5m^2$	5.142***	4.557***	4.160***	4.412***	5.006***	5.079***
	(0.859)	(0.855)	(0.852)	(0.851)	(0.850)	(0.849)
Size	8.098***	4.255***	9.596***	16.871***	−35.366***	−14.369
	(0.097)	(0.329)	(0.878)	(2.335)	(6.188)	(17.498)
Size^2		0.029***	−0.062***	−0.267***	1.804***	0.732
		(0.003)	(0.015)	(0.064)	(0.240)	(0.880)
Size^3			0.000***	0.003***	−0.035***	−0.008
			(0.000)	(0.001)	(0.004)	(0.022)
Size^4				0.000***	0.000***	0.000
				(0.000)	(0.000)	(0.000)
Size^5					0.000***	0.000
					(0.000)	(0.000)
Size^6						0.000
						(0.000)
AIC	4,198,553	4,198,387	4,198,345	4,198,335	4,198,255	4,198,251
Adj. R^2	0.741	0.741	0.741	0.741	0.741	0.741
Observations	329,049	329,049	329,049	329,049	329,049	329,049

Notes: This table summarizes regression results using different size polynomials. The dependent variable is basic rent and all regression include apartment and location specific controls. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Table A.4 *Rent-price discontinuities: placebo thresholds (local-quadratic)*

	Dependent variable: Basic rent					
	(1)	(2)	(3)	(4)	(5)	(6)
Placebo thresholds (yd^2)						
TightMkt × Left Digit	5.990***		5.855***	6.821***		6.631***
	(1.372)		(1.366)	(1.387)		(1.381)
TightMkt × Round	−4.785***		−4.880***	−4.575***		−4.661***
	(1.689)		(1.683)	(1.694)		(1.687)
LL Type × Left Digit		−8.975*	−8.462*		−12.440**	−11.799**
		(5.025)	(5.018)		(5.027)	(5.016)
LL Type × Round		−3.572	−4.165		−3.742	−4.424
		(6.324)	(6.321)		(6.321)	(6.315)
Left Digit (LD)	−5.995***	−3.125	−6.303***	−3.265***	−0.031	−3.615***
	(1.928)	(1.907)	(1.930)	(1.031)	(0.994)	(1.038)
Round (R)	3.714***	0.960	3.627***	2.668***	0.005	2.522***
	(1.312)	(1.243)	(1.312)	(0.965)	(0.904)	(0.968)
Actual thresholds (m^2)						
TightMkt × Left Digit				−1.978		−1.734
				(1.384)		(1.378)
TightMkt × Round				15.381***		14.844***
				(2.034)		(2.018)
LL Type × Left Digit					19.319***	19.436***
					(5.064)	(5.054)
LL Type × Round					−29.203***	−27.729***
					(6.209)	(6.179)
Left Digit (LD)				6.139***	5.636***	6.640***
				(1.655)	(1.649)	(1.659)
Round (R)				−4.982***	0.962	−7.139***
				(1.484)	(1.430)	(1.543)
TightMkt	33.259***	35.240***	33.338***	31.163***	35.212***	31.211***
	(1.287)	(1.032)	(1.284)	(1.465)	(1.031)	(1.457)
LL Type	2.425	7.616*	7.512*	−0.792	6.505	3.598
	(2.977)	(4.016)	(4.011)	(2.986)	(5.047)	(5.016)
LL Type × TightMkt	−85.732***	−85.648***	−85.728***	−81.456***	−85.781***	−82.098***
	(4.656)	(4.664)	(4.662)	(4.670)	(4.657)	(4.670)
Adj. R^2	0.767	0.767	0.767	0.768	0.768	0.768
Observations	126,847	126,847	126,847	126,847	126,847	126,847

Notes: This table summarizes results from regressions using placebo-thresholds after converting apartment size measures to square-yards ($1m^2 = 1.196yd^2$). Shown are coefficients from local-quadratic estimations using windows of size $10yd^2$ centered at multiples of $10yd^2$. *Left Digit* indicates whether the size measure includes the larger leftmost digit within each window. *Round* indicates whether the size measure is a multiple of $5m^2$. *LL Type* is the landlord-type measure as defined in Equation (1.3). *TightMkt* indicates tight local markets. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Table A.5 *Rent-price discontinuities: placebo thresholds*

	Dependent variable: Basic rent			
	Full sample	Unformatted	Round	Charm
	(1)	(2)	(3)	(4)
$\geq 40yd^2$	–3.411	–6.720***	0.445	3.863
	(2.159)	(2.257)	(6.783)	(6.343)
$\geq 50yd^2$	5.852***	4.657**	8.935	5.122
	(1.879)	(1.921)	(6.051)	(6.016)
$\geq 60yd^2$	–2.019	–0.582	0.230	0.195
	(1.387)	(1.446)	(4.411)	(4.259)
$\geq 70yd^2$	2.936**	5.496***	–2.647	–0.516
	(1.313)	(1.384)	(4.310)	(3.719)
$\geq 80yd^2$	–4.193***	–2.590*	–2.003	–15.403***
	(1.437)	(1.536)	(4.589)	(4.122)
$\geq 90yd^2$	–0.590	2.723	1.365	–22.122***
	(1.948)	(2.167)	(5.184)	(5.436)
$\geq 100yd^2$	–10.236***	–6.664**	–2.040	–36.088***
	(2.608)	(2.944)	(6.725)	(6.854)
$\geq 110yd^2$	3.952	18.475***	–1.451	–16.608*
	(3.411)	(4.166)	(7.188)	(9.272)
Mult. $5m^2$	0.006	–0.994	–0.015	4.498**
	(0.657)	(0.688)	(2.030)	(1.962)
Adj. R^2	0.741	0.752	0.719	0.739
Observations	329,049	227,301	58,529	43,219
% Integer	55.61	50.07	73.83	58.52
% Mult. $5m^2$	20.07	16.23	33.28	21.27

Notes: This table summarizes results from regressions using placebo-thresholds after converting apartment size measures to square-yards ($1m^2 = 1.196yd^2$). Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%. The two bottom rows indicate the shares of apartments with integer and multiple-of-$5m^2$ size measures.

Table A.6 *Charmers & rounders: correlations*

	Landlord type			Rent-price format	
	LL Type (1)	LL %Charm (2)	LL %Round (3)	Charm price (4)	Round price (5)
Basic rent	0.000	0.000**	0.000	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Basic rent / m^2	−0.004***	0.001***	0.005***	0.005***	0.003***
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Rounded rent price	−0.181***	−0.014***	0.167***		
	(0.002)	(0.001)	(0.002)		
Strategic rent price	0.089***	0.090***	0.001		
	(0.002)	(0.002)	(0.001)		
Living space (m^2)	0.000***	0.000	0.000***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Integer living space	−0.028***	0.014***	0.042***	0.025***	0.035***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)
Living space mult. 5m^2	−0.019***	0.001	0.020***	0.011***	0.068***
	(0.002)	(0.001)	(0.002)	(0.004)	(0.004)
Number rooms	0.006***	−0.002***	−0.008***	−0.012***	−0.004
	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)
Equipment quality	0.003*	0.000	−0.003**	0.002	−0.013***
	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)
Age	0.000**	0.000	0.000*	0.000	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number photos	0.000	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number words	0.000***	0.000***	0.000***	0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LL Type				0.340***	−0.636***
				(0.006)	(0.006)
Landlord size*	0.000***	0.000***	0.000***	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GDP p.c.	−0.001***	0.000**	0.000***	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH inc p.c.	−0.003***	0.001**	0.004***	−0.001*	0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Population / area	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Apartments / population	0.000***	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
TightMkt	0.003	0.003***	0.000	−0.001	−0.007*
	(0.002)	(0.001)	(0.002)	(0.003)	(0.004)
East	0.041***	0.001	−0.040***	−0.024***	−0.012***
	(0.002)	(0.001)	(0.002)	(0.004)	(0.004)
Intercept	0.068***	0.077***	0.008	0.061***	0.027
	(0.012)	(0.006)	(0.009)	(0.018)	(0.021)
Adj. R^2	0.227	0.121	0.316	0.050	0.176
Observations	73,302	73,302	73,302	73,302	73,302

Notes: Shown are estimate from linear probability models and dependent variables are indicated in the column headers. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.
* We define landlord size as the number of apartment offerings we observe from the same landlord.

Figure A.4 *Observable heterogeneity at salient size measures*

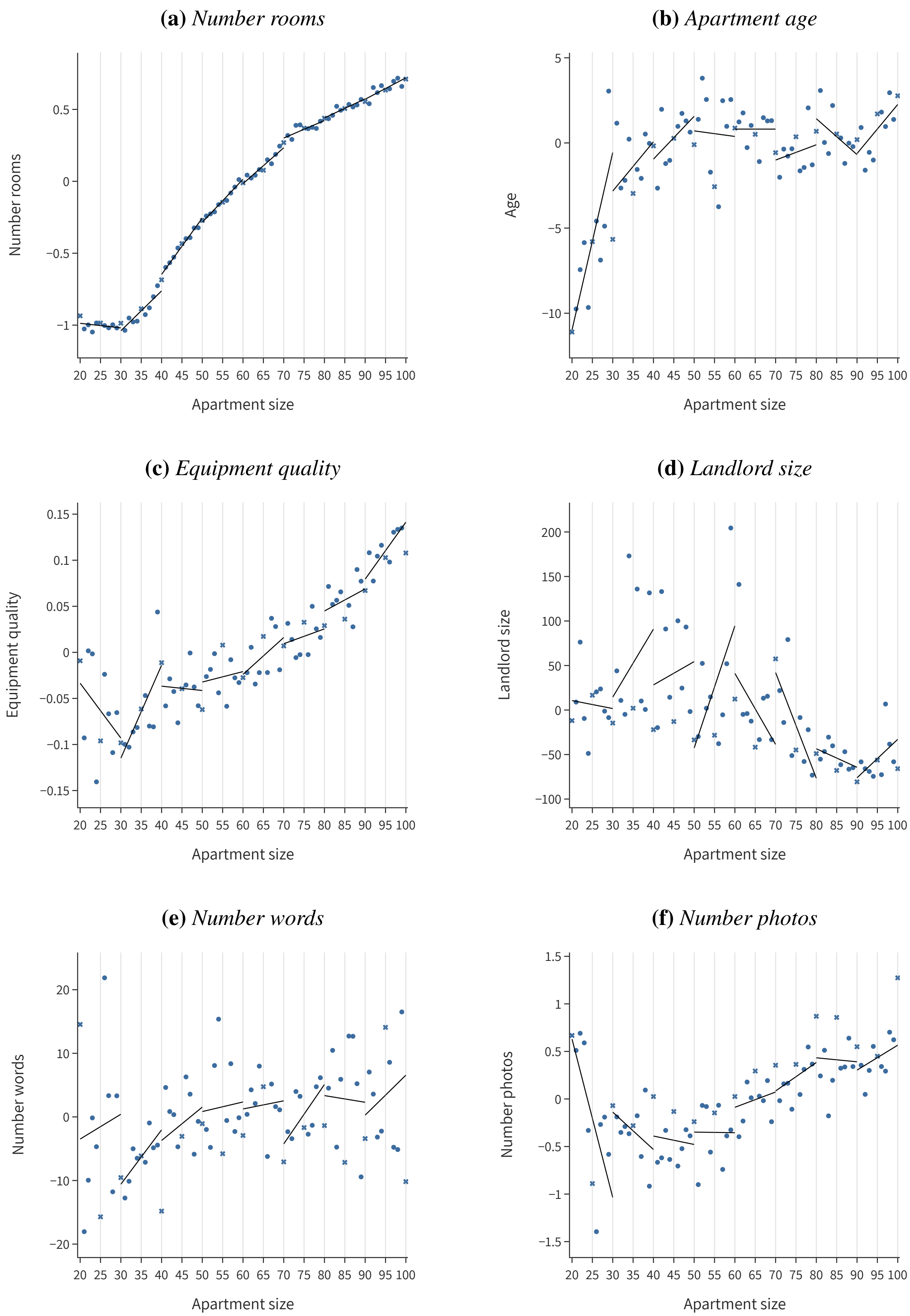
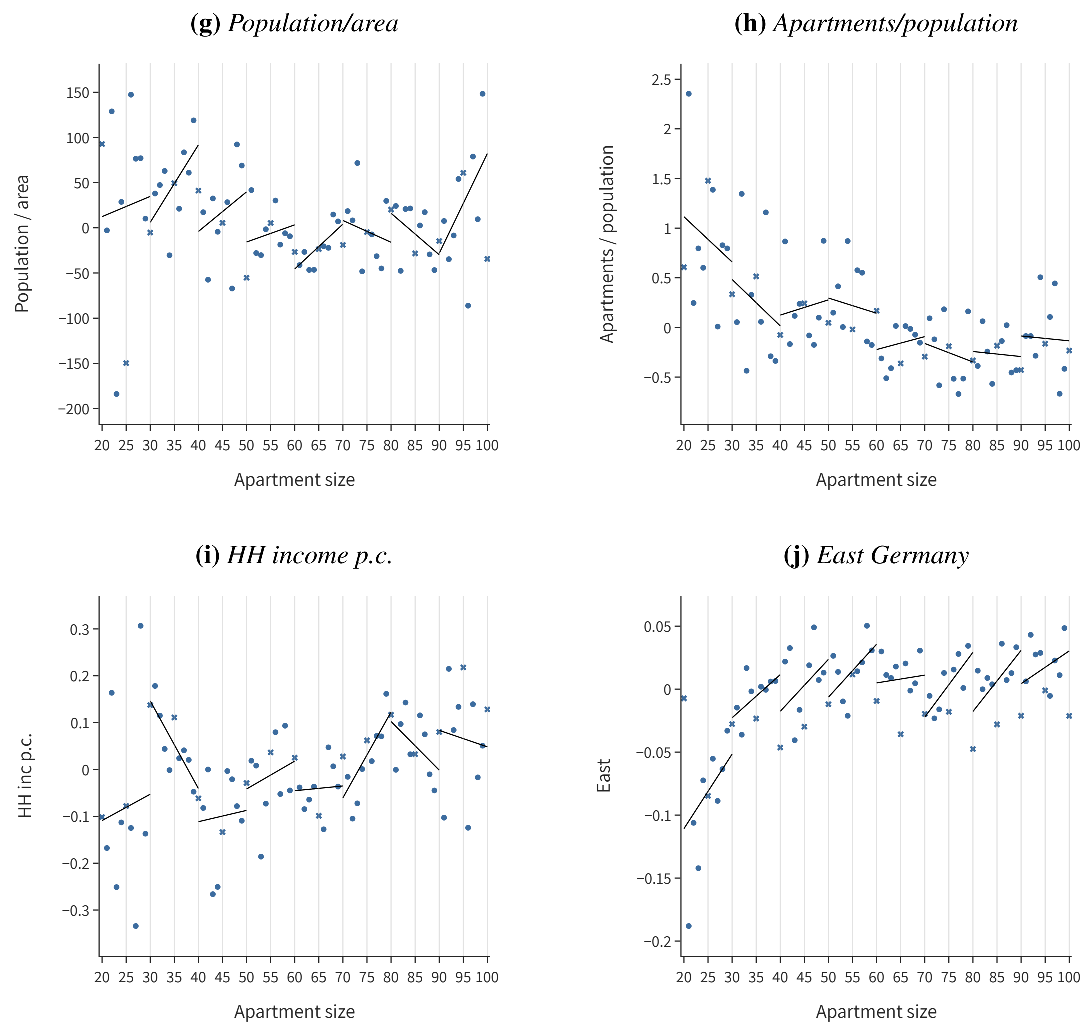


Figure A.4 *Observable heterogeneity at salient size measures (cont'd)*



Notes: These figures show average residuals of the respective covariate as a function of apartment size. The residuals are taken from regressions of the covariates on all other controls used in the main regressions. The residuals are averaged within $1m^2$ bins. Solid lines represent linear fits between $10m^2$ multiples.

Table A.7 *Hedonic pricing model: apartment & location features*

	Dependent variable: Basic rent			
	(1)	(2)	(3)	(4)
Living space (m^2)	8.079***	4^{th}-order Polynomial		
	(0.055)			
Number rooms	−17.507***	−14.966***	−15.151***	−22.621***
	(0.908)	(0.917)	(0.919)	(0.611)
Floor	2.999***	3.054***	3.064***	4.232***
	(0.372)	(0.371)	(0.371)	(0.253)
Age	−1.782***	−1.769***	−1.766***	−2.078***
	(0.045)	(0.045)	(0.045)	(0.031)
Age^2	0.015***	0.014***	0.014***	0.016***
	(0.000)	(0.000)	(0.000)	(0.000)
Number photos	1.806***	1.810***	1.803***	
	(0.098)	(0.097)	(0.097)	
Number words	0.056***	0.056***	0.056***	
	(0.004)	(0.004)	(0.004)	
Landlord size	−0.002***	−0.002***	−0.002***	
	(0.000)	(0.000)	(0.000)	
Equipment quality				
-Einfach (Simple)	−48.348***	−49.270***	−49.457***	−56.288***
	(3.516)	(3.507)	(3.503)	(2.370)
-Normal (Normal)	−26.676***	−26.612***	−26.693***	−22.104***
	(0.945)	(0.943)	(0.942)	(0.599)
-Gehoben (High)	34.620***	34.471***	34.381***	50.882***
	(1.541)	(1.538)	(1.537)	(1.003)
-Luxus (Luxury)	153.411***	153.085***	152.863***	192.264***
	(6.515)	(6.481)	(6.480)	(4.366)
Apartment features				
-Basement	3.059***	4.004***	4.117***	
	(0.938)	(0.939)	(0.938)	
-Balcony	6.165***	8.085***	8.308***	
	(0.883)	(0.884)	(0.885)	
-Garden	−2.001*	−2.021*	−2.127*	
	(1.208)	(1.204)	(1.204)	
-Lift	36.680***	35.541***	35.685***	
	(1.302)	(1.299)	(1.299)	
-Kitchen	50.152***	49.710***	49.559***	
	(0.995)	(0.992)	(0.992)	
-Guest toilet	39.481***	31.616***	31.691***	
	(2.151)	(2.199)	(2.199)	
-Step-free access	6.113***	6.489***	6.483***	
	(2.094)	(2.086)	(2.085)	
-Shared apartment	9.424***	10.405***	10.452***	
	(1.963)	(1.958)	(1.958)	
-Needs entitlement	−136.201***	−135.301***	−135.486***	
	(2.320)	(2.305)	(2.305)	
HH inc p.c.	22.968***	22.994***	23.021***	26.542***
	(0.297)	(0.296)	(0.296)	(0.205)
Population/Area	0.092***	0.092***	0.092***	0.097***
	(0.001)	(0.001)	(0.001)	(0.001)
Apartments/population	−2.707***	−2.712***	−2.701***	−2.552***
	(0.052)	(0.052)	(0.052)	(0.033)
Rounding dummies			X	X
Lef-digit dummies			X	X
Adj. R^2	0.768	0.769	0.770	0.739
Observations	103,476	103,476	103,476	254,156

Notes: This table summarizes regression results of basic rent on apartment and location characteristics. The reference category for all categorical indicators is “not reported”. Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Table A.7 *Hedonic pricing model: apartment & location features (cont'd)*

	Dependent variable: Basic rent			
	(1)	(2)	(3)	(4)
Apartment condition				
-Abbruchreif (Dilapidated)	6.482**	−7.105**	−10.813***	−33.129
	(2.573)	(3.076)	(3.342)	(33.233)
-Renovierungsbedürftig (Requires renovation)	−44.227***	−44.336***	−44.235***	−39.301***
	(4.236)	(4.241)	(4.233)	(2.787)
-Gepflegt (Maintained)	−11.799***	−11.621***	−11.582***	−9.383***
	(1.230)	(1.227)	(1.227)	(0.837)
-Modernisiert (Modernized)	−22.075***	−21.725***	−21.644***	−23.788***
	(1.758)	(1.753)	(1.753)	(1.138)
-Vollständig renoviert (Renovated)	−11.027***	−10.979***	−10.975***	−10.491***
	(1.638)	(1.636)	(1.637)	(1.080)
-Saniert (Refurbished)	−5.521***	−5.356***	−5.401***	−7.322***
	(1.484)	(1.478)	(1.478)	(0.987)
-Erstbezug nach Sanierung (Refurbished, First-time occupancy)	16.577***	16.864***	16.894***	13.661***
	(2.154)	(2.149)	(2.149)	(1.442)
-Neuwertig (As new)	34.075***	34.011***	33.979***	50.044***
	(2.253)	(2.244)	(2.244)	(1.533)
-Erstbezug (First-time occupancy)	102.241***	102.099***	102.305***	128.817***
	(2.827)	(2.812)	(2.813)	(1.864)
-Nach Vereinbarung (By arrangement)	−28.861***	−28.940***	−29.008***	−38.016***
	(3.411)	(3.410)	(3.405)	(2.192)
Apartment type				
-Souterrain (Basement)	−24.293***	−24.830***	−25.257***	−41.617***
	(6.295)	(6.255)	(6.265)	(4.131)
-Erdgeschosswohnung (Ground floor)	13.256***	13.140***	13.152***	−1.397
	(1.827)	(1.823)	(1.822)	(1.214)
-Hochparterre (Mezzanine)	19.126***	18.988***	18.991***	8.561***
	(3.003)	(2.991)	(2.990)	(2.013)
-Etagenwohnung (Standard)	21.256***	21.102***	21.141***	6.499***
	(1.440)	(1.437)	(1.436)	(0.963)
-Dachgeschoss (Top floor)	12.067***	12.481***	12.292***	−4.286***
	(1.807)	(1.802)	(1.802)	(1.226)
-Loft (Loft)	32.296*	33.585*	32.928*	15.843
	(18.113)	(17.994)	(18.036)	(11.137)
-Penthouse (Penthouse)	81.731***	80.089***	79.801***	97.522***
	(9.215)	(9.149)	(9.148)	(6.499)
-Terrassenwohnung (Incl. terrace)	32.053***	31.864***	31.851***	28.017***
	(4.697)	(4.710)	(4.703)	(3.132)
-Maisonette (Maisonette)	23.858***	21.619***	21.480***	23.858***
	(3.641)	(3.632)	(3.632)	(2.423)
-Sonstige (Other)	−24.226***	−25.176***	−25.149***	−30.811***
	(2.591)	(2.576)	(2.574)	(1.792)
Rounding dummies			X	X
Lef-digit dummies			X	X
Adj. R^2	0.768	0.769	0.770	0.739
Observations	103,476	103,476	103,476	254,156

Notes: This table summarizes regression results of basic rent on apartment and location characteristics. The reference category for all categorical indicators is "not reported". Robust standard errors are reported in parentheses. ***, **, * indicate statistical significance at 1%, 5% and 10%.

Appendix B

Appendix to Chapter 2

B.0.1 Combining GitHub Torrent and GitHub Archive

We use the latest (March 2021) snapshot from GitHub Torrent (GHT) (Gousios 2013) and the GitHub Archive (GHA) dataset. Both GHT and GHA provide a mirror of the GitHub public event stream from 2012 on. Our main data source is GHT, from which we derive information about projects (repositories) and project members, including their location data. GHT also features data on code contributions (commits), however, commit data is missing for the period between June 2019 and December 2019 as well as after March 2021 (see Figure B.1). Therefore, we utilize GHA to get commit data with full coverage for the period used in our analysis.

Matching project data is straightforward and can be achieved using unique project-owner-name combinations. Matching user data and their contributions is more arduous. A straightforward way to match contributions is to match based on push events.[1] However, there are two major caveats with this method. First, push events can contain multiple commits, potentially from multiple users, and thus commits could be attributed to the wrong user.[2] Second, a single commit can appear in multiple push events, e.g. if code is merged from one project or forked to another.[3] We want to ensure that we count each commit only once and associate it with the user and project it originates from. Consequently, we prefer matching at the commit level rather than the push level. This, however, is more complicated because GHT associates code commits with GitHub user accounts while GHA associates contributions with Git usernames, which are distinct from each other.[4]

[1]A commit records changes to the (local) code repository. It is made using the Git version control system. A push event then sends one or several commits to the remote repository on GitHub.

[2]We do not consider this a problem in Section 2.3 because we want to allow for commits to be ”exported” to multiple destinations. In Section 2.4, however, we measure teams' (team members') productivity, where it is more crucial to attribute code contributions to their original author correctly.

[3]Forks are new repositories that share the code of an original "upstream" repository. They are created by contributors who do not have direct access to (i.e. are not members of) the original project.

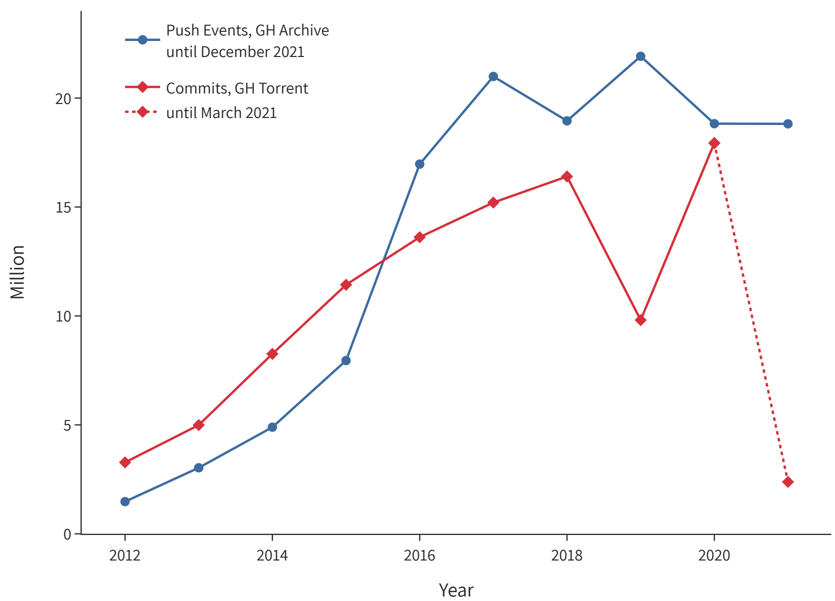
[4]Git is a locally installed version control system that helps manage source code and keep track of code history. It can be used independently of GitHub. GitHub is a cloud-based platform that is used to host, manage and collaborate on Git repositories.

In order to identify individual code contributions Git assigns a unique hash value to every commit. These hash values are featured in both datasets and thus allow us to directly match commits in the overlapping period from January 2015 to May 2019 and January 2020 to March 2021. Using this method we are able to obtain Git-GitHub user pairs. If a GitHub account is matched to multiple Git usernames we attribute all contributions from these usernames to the GitHub account. If a Git username is matched with multiple GitHub accounts, we choose the most frequent combination and attribute all contributions to the respective GitHub account. Using these matches, we replace missing GitHub account information of commits recorded in the second half of 2019. To further improve our data coverage, we fill in the remaining missing values based on push events as described above.[5] Overall, we are able to replace missing GHT user data for 97% of the GHA commits between June 2019 and Dec 2019.[6]

[5]This improves data coverage by nine percentage points.

[6]The remaining 3% of commits are from users who only contributed in the second half of 2019, but neither before nor afterward. Therefore, these users are of little relevance to our analysis.

Figure B.1 *Number of events on GitHub*



Notes: This figure shows commits from users with reported locations and to repositories owned by users with reported locations. The drop in the number of commits in the GHT dataset in 2019 is due to missing data from June 2019 until December 2019.

B.0.2 Estimation of the gravity equation with GitHub data

Our baseline specification for the gravity model is the following:

$$c_{ijt} = \beta_0 D_{ij}^{\beta_1} \{X\}^{\{\gamma\}} \tau_{c_it} \tau_{c_jt} \varepsilon_{ijt} \tag{B.1}$$

where

c_{ijt} — number of collaborations between a city pair ij in a year t. It is measured by the number of contributions by users from a city i submitted to a project owned by users from a city j. In our setting, direction matters: collaborations between city pairs ij and ji are treated as two different observations. To make an analogy to the trade and migration literature, we think of the city of a committer as the origin (e.g. origin of a service provider — an exporter) and the city of a project owner as the destination (e.g. destination of services — an importer). Another analogy to make is a citation of a patent or a scientific article, where a cited contribution could be considered as an origin of knowledge and a citing patent (article) — a destination of the knowledge transfer.

D_{ij} — geographic distance between two cities. We calculate it as the shortest path (in km) between two cities, using their geographic coordinates.

$\{X\}$ — a vector of controls. Economic size (mass analogy in the gravity equation) is proxied by the number of registered GitHub users in an "origin" city (city of a committer) and a "destination" city (city of a project's owner). In addition, we add a dummy for foreign country and a dummy for common language (in case of cross-border collaborations). Conditional on geographic distance, these dummies capture the effects of state borders and language barriers. In line with the trade literature, we also control for remoteness of committers and users, which we measure as the user-weighted average distance to other cities.

τ_{c_it} and τ_{c_jt} account for country of committers and country of owners time-specific fixed effects.[7] We conservatively cluster standard errors at the country-pair level.

For most regressions, we use Poisson pseudo-maximum likelihood estimation with multi-way fixed effects, which is consistent for models with count data in the presence of many zero outcomes. We estimate the effects first jointly and then separately at the intensive and extensive margins. For the latter, we use a linear probability model.

To benchmark the effect of geographic barriers on GitHub to their role in shaping modern trade flows, we use 2012–19 cross-country export data from the CEPII Gravity dataset, Version 202102, and aggregate GitHub data at the country-pair level for the same period.

[7]The estimates of the main parameters of interest are robust to including city of committers and city of owners fixed effects.

Table B.1 *Country-level gravity regressions: benchmark to trade*

Variables	Tradeflows (1)	Tradeflows > 0 (2)	Contributions (3)	Contributions > 0 (4)
Distance	–0.803***	–0.807***	–0.488***	–0.449***
	(0.037)	(0.037)	(0.078)	(0.081)
GitHub users, origin			1.381***	1.154***
			(0.318)	(0.326)
GitHub users, destination			–0.172	–0.421
			(0.241)	(0.257)
GDP, origin	0.543***	0.535***	–0.363	–0.322
	(0.077)	(0.078)	(0.320)	(0.323)
GDP, destination	0.259***	0.271***	0.479*	0.504*
	(0.064)	(0.063)	(0.272)	(0.276)
Remoteness, origin	–0.073	–0.098	–1.036*	–0.563
	(0.498)	(0.498)	(0.558)	(0.593)
Remoteness, destination	–0.912*	–0.928**	–0.073	0.066
	(0.466)	(0.466)	(0.724)	(0.755)
Contiguity	0.485***	0.485***	–0.129	–0.089
	(0.069)	(0.069)	(0.171)	(0.171)
Common language	0.148**	0.150**	0.262**	0.214*
	(0.066)	(0.066)	(0.113)	(0.115)
Regional trade aggrement	0.314***	0.309***	–0.228**	–0.263**
	(0.054)	(0.054)	(0.107)	(0.110)
Observations	268,590	197,779	225,208	20,057
Clusters	37,419	31,922	30,797	5283

Notes: The dependent variables are trade flows (exports from origin to destination) in columns (1) and (2) as well as GitHub contributions (from origin to destination) in columns (3) and (4) between a given country pair. Columns (2) and (4) limit the sample to pairs with non-zero trade flows/contributions. All specifications include year, country of origin, and country of destination fixed effects. Standard errors are clustered at a country pair level. Estimation method: PPML.

Table B.2 *City-level gravity regressions accounting for programming language*

Variables	Contributions (1)	Contributions >0 (2)	Contributions yes/no (3)	Contributions distance dummies (4)
Distance	−0.420***		−0.282***	
	(0.037)		(0.027)	
1–50km		−0.267		−0.009
		(0.357)		(0.270)
50–100km		−1.963***		−1.325***
		(0.334)		(0.279)
100–300km		−2.399***		−1.645***
		(0.252)		(0.191)
300–700km		−2.655***		−1.776***
		(0.284)		(0.186)
>700km		−3.038***		−2.035***
		(0.310)		(0.227)
Foreign country	−1.353***	−1.692***	−0.864***	−1.069***
	(0.233)	(0.170)	(0.157)	(0.108)
Common language	0.322***	0.280***	0.176**	0.127*
	(0.102)	(0.076)	(0.080)	(0.070)
Users, owner	0.161***	0.176***	0.093**	0.107***
	(0.031)	(0.031)	(0.043)	(0.040)
Users, committer	0.733***	0.744***	0.551***	0.548***
	(0.030)	(0.031)	(0.031)	(0.031)
Remoteness, committer	−0.011	−0.083	0.023**	0.023**
	(0.126)	(0.115)	(0.011)	(0.010)
Remoteness, owner	2.666***	2.622***	0.393***	0.378***
	(0.156)	(0.179)	(0.072)	(0.077)
Observations	307,718	307,718	610,522	610,522
Clusters	4487	4487	4075	4075
Same prog. language	No	No	Yes	Yes

Notes: The dependent variable is contributions between a city pair (the sample includes observations with non-zero contributions). All specifications include country of committers and country of project owners time-specific fixed effects. The specification in columns (3) and (4) adds programming language time-specific effects. Standard errors are clustered at a country pair level. Estimation method: PPML.

B.0.3 Team-level analysis: additional figures and tables

Table B.3 *Team-level data: summary statistics*

	Control			Treatment		
	Co-located (1)	Mixed (2)	Distributed (3)	Co-located (4)	Mixed (5)	Distributed (6)
Avg. mtly. commits						
pre-pandemic (hyp.)	2.52	2.44	2.59	2.25	2.28	3.02
	(4.10)	(4.06)	(4.34)	(3.71)	(3.81)	(4.68)
post-pandemic (hyp.)	1.18	1.08	1.39	0.74	1.21	1.82
	(3.41)	(3.54)	(5.47)	(2.36)	(3.92)	(7.68)
Num. commits (first year)	90.36	80.59	78.75	85.46	73.83	82.03
	(91.00)	(87.89)	(86.51)	(84.27)	(82.71)	(84.44)
Num. watchers (frist year)	8.32	24.50	42.21	7.95	23.55	37.15
	(52.28)	(189.49)	(374.22)	(69.07)	(272.72)	(269.65)
Project age	7.42	7.82	7.73	7.28	7.68	8.14
	(3.45)	(3.46)	(3.45)	(3.42)	(3.56)	(3.55)
Project year						
2015	0.29	0.29	0.32	0.00	0.00	0.00
2016	0.33	0.38	0.37	0.00	0.00	0.00
2017	0.38	0.33	0.31	0.00	0.00	0.00
2018	0.00	0.00	0.00	1.00	1.00	1.00
Owner country						
USA	0.23	0.30	0.44	0.21	0.31	0.42
Brazil	0.03	0.04	0.04	0.03	0.05	0.04
India	0.05	0.04	0.03	0.07	0.04	0.03
Great Britain	0.08	0.08	0.06	0.07	0.07	0.07
Canada	0.02	0.04	0.04	0.04	0.03	0.03
Other	0.60	0.51	0.40	0.58	0.50	0.41
Num. projects	533	2212	3695	254	627	762

Notes: This table shows mean values and standard errors in parentheses of control and dependent variables. Treatment group projects are those started in 2018. The control group are earlier projects with the hypothetical onset of the pandemic shifted backwards. For example, the hypothetical onset of the pandemic is set to March 2019 for projects started in 2017. Columns (1) to (3) show control group summary statistics for co-located, mixed and distributed teams, respectively. Columns (4) to (6) show treatment group summary statistics for co-located, mixed and distributed teams, respectively. The number of commits is first averaged within teams and then across teams. Project age is measured at the start of the observational period, i.e. 12 months before the (hypothetical) onset of the coronavirus pandemic. The sample includes projects which received at least on contribution in the observational period of 24 months centered around the (hypothetical) onset of the coronavirus pandemic.

Table B.4 *Team-level data: summary statistics*

	Co-located (1)	Mixed (2)	Distributed (3)
Avg. mtly. commits			
pre-pandemic	2.30	2.24	2.63
	(4.15)	(5.39)	(5.63)
post-pandemic	1.11	1.57	2.27
	(3.41)	(5.53)	(10.42)
Num. contributors			
pre-pandemic	3.01	3.16	3.48
	(4.00)	(6.36)	(8.66)
post-pandemic	0.18	0.28	0.54
	(2.28)	(3.93)	(8.14)
Num. commits (first year)	88.29	80.35	83.18
	(86.56)	(87.84)	(86.80)
Num. watchers (frist year)	11.50	32.88	45.32
	(74.77)	(260.46)	(260.47)
Log experience (pre-pandemic)	7.15	7.30	7.58
	(2.47)	(2.49)	(2.30)
Log message length (pre-pandemic)	3.75	3.79	3.91
	(0.60)	(0.58)	(0.60)
Project age	17.06	20.47	22.68
	(12.90)	(13.25)	(13.10)
Project year			
2015	0.09	0.11	0.15
2016	0.16	0.22	0.24
2017	0.21	0.25	0.28
2018	0.54	0.42	0.33
Owner country			
USA	0.22	0.32	0.44
Brazil	0.02	0.04	0.03
India	0.06	0.03	0.02
Great Britain	0.07	0.08	0.06
Canada	0.03	0.03	0.04
Other	0.60	0.50	0.41
Num. projects	467	1499	2313

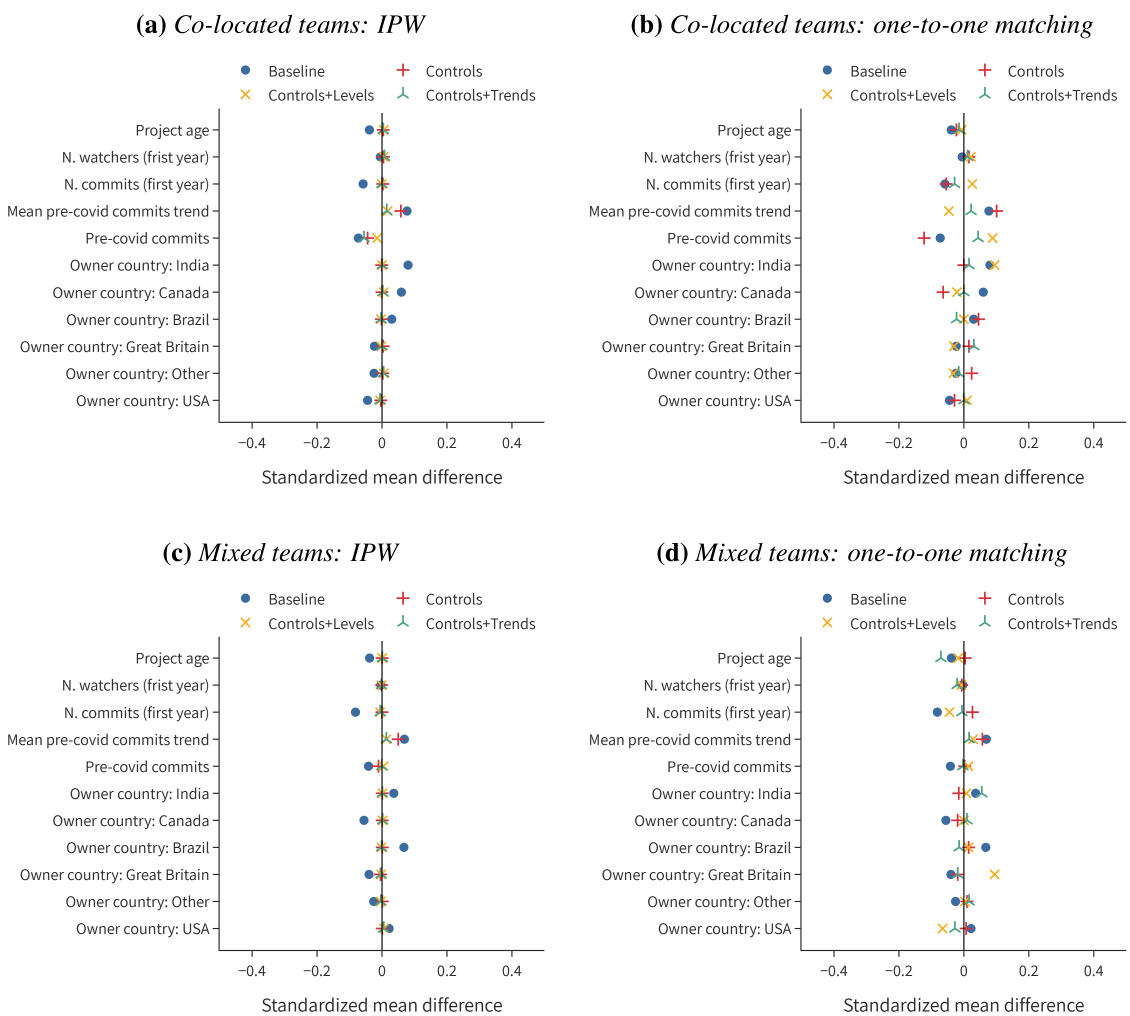
Notes: This table shows mean values and standard errors in parentheses of control and dependent variables. Columns (1), (2) and (3) show summary statistics for co-located, mixed and distributed teams, respectively. The number of commits is first averaged within teams and then across teams. Project age is measured at the start of the observational period, i.e. 12 months before the onset of the coronavirus pandemic. The sample includes projects which received at least on contribution in the observational period of 24 months centered around the onset of the coronavirus pandemic in March 2020.

Table B.5 *COVID-19 effects on team performance: project & month fixed effects*

	Dependent: Number of commits			
Team composition	Co-located (1)	Mixed (2)	Distributed (3)	Full sample (4)
COVID × Treated	−0.440**	0.175	0.108	−0.426**
	(0.215)	(0.132)	(0.155)	(0.213)
COVID × D(Loc.=2)				−0.082
				(0.135)
COVID × D(Loc.=3)				0.092
				(0.138)
COVID × Treated × D(Loc.=2)				0.585**
				(0.251)
COVID × Treated × D(Loc.=3)				0.531**
				(0.267)
Controls	X	X	X	X
Project FE	X	X	X	X
Time FE	X	X	X	X
Pseudo R^2	0.434	0.440	0.448	0.443
N	18,888	68,136	106,968	193,992
Clusters	787	2,839	4,457	8,083

Notes: Columns (1), (2) and (3) split the sample into co-located, mixed and distributed teams, respectively. *Treated* projects are founded in 2018 and are exposed to 12 months of COVID-19. The control group are earlier projects with the hypothetical onset of the pandemic shifted backwards. For example, the hypothetical onset of the pandemic is set to March 2019 for projects started in 2017. All regressions include project and time fixed effects. Further controls include project age (linear & quadratic) as well as interactions of the number of commits and watchers within a project's first year and country-of-owner fixed effects with the (hypothetical) COVID indicator. The sample covers 24 months centered around the hypothetical onset of the coronavirus pandemic. Estimation method: Poisson ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

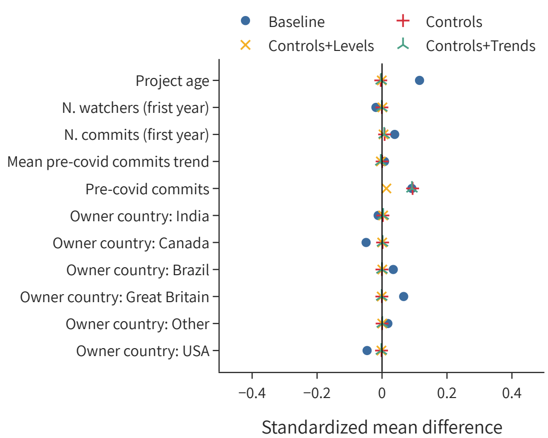
Figure B.2 *Artificial control teams: balance of covariates*



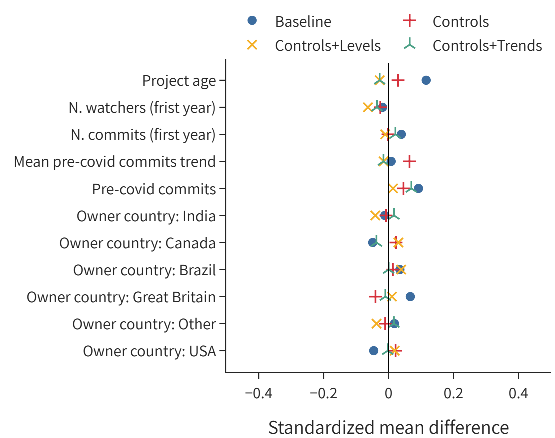
Notes: The figures shows standardized mean differences of control variables between projects started in 2018 (treatment) and earlier projects (control). The differences are calculated by subtracting the means of treated teams from those of control teams and dividing by the standard deviation of treated teams. In panels (a) and (c) teams are balanced using the inverse probability weighting approach suggested in Abadie (2005) and in panels (b) and (d) we apply one-to-one propensity score matching (Hansen and Klopfer 2006). Propensity scores are estimated using logistic regressions of a treatment indicator on either just controls, quarterly pre-pandemic commit levels and controls or quarterly pre-pandemic commit trends and controls. Controls include project age, number of commits and watchers within a project's first year and project-owner-country fixed effects. Pre-covid commits is the sum of commits a project has received in the 12 months before the onset of the (hypothetical) coronavirus pandemic. Mean pre-covid commits trend measures the project-mean of quarterly first-differences in the number of commits before the onset of the (hypothetical) coronavirus pandemic.

Figure B.2 *Artificial control teams: balance of covariates (cont'd)*

(e) *Fully distributed teams: IPW*



(f) *Fully distributed teams: one-to-one matching*



Notes: The figures shows standardized mean differences of control variables between projects started in 2018 (treatment) and earlier projects (control). The differences are calculated by subtracting the means of treated teams from those of control teams and dividing by the standard deviation of treated teams. In panel (e) teams are balanced using the inverse probability weighting approach suggested in Abadie (2005) and in panel (f) we apply one-to-one propensity score matching (Hansen and Klopfer 2006). Propensity scores are estimated using logistic regressions of a treatment indicator on either just controls, quarterly pre-pandemic commit levels and controls or quarterly pre-pandemic commit trends and controls. Controls include project age, number of commits and watchers within a project's first year and project-owner-country fixed effects. Pre-covid commits is the sum of commits a project has received in the 12 months before the onset of the (hypothetical) coronavirus pandemic. Mean pre-covid commits trend measures the project-mean of quarterly first-differences in the number of commits before the onset of the (hypothetical) coronavirus pandemic.

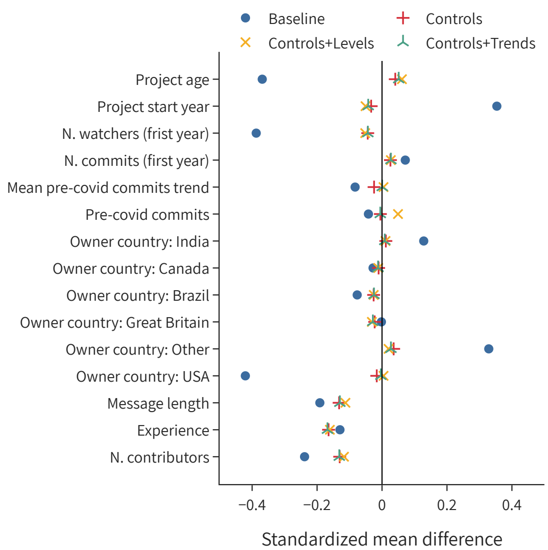
Table B.6 *COVID-19 effects on team performance: matched results*

Method	Dependent: Number of commits					
	Inverse probability weighting			One-to-one matching		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A *Co-located teams*						
Treated	−0.049	−0.018	−0.080	−0.065	0.066	0.073
	(0.114)	(0.113)	(0.112)	(0.122)	(0.135)	(0.127)
Treated × COVID	−0.440**	−0.448**	−0.457**	−0.548**	−0.606***	−0.486**
	(0.197)	(0.198)	(0.197)	(0.229)	(0.220)	(0.248)
Pseudo R2	18,888	18,888	18,888	12,192	12,192	12,192
N	0.125	0.123	0.124	0.138	0.111	0.125
Clusters	787	787	787	508	508	508
Panel B *Mixed teams*						
Treated	0.011	0.017	0.008	−0.014	0.053	0.034
	(0.069)	(0.079)	(0.078)	(0.086)	(0.085)	(0.087)
Treated × COVID	0.143	0.060	0.063	0.265*	0.263*	0.085
	(0.127)	(0.123)	(0.123)	(0.151)	(0.148)	(0.149)
Pseudo R2	0.113	0.117	0.117	0.117	0.116	0.114
N	68,136	68,136	68,136	30,096	30,096	30,096
Clusters	2,839	2,839	2,839	1,254	1,254	1,254
Panel C *Fully distributed teams*						
Treated	0.162***	0.033	0.162***	0.119	0.070	0.057
	(0.058)	(0.054)	(0.060)	(0.081)	(0.075)	(0.079)
Treated × COVID	0.006	0.022	0.012	0.170	0.163	0.113
	(0.122)	(0.121)	(0.122)	(0.132)	(0.131)	(0.130)
Pseudo R2	0.097	0.095	0.097	0.105	0.110	0.088
N	106,968	106,968	106,968	36,576	36,576	36,576
Clusters	4,457	4,457	4,457	1,524	1,524	1,524
Matching variables						
Controls	X	X	X	X	X	X
Pre-treatment outcome levels		X			X	
Pre-treatment outcome trends			X			X

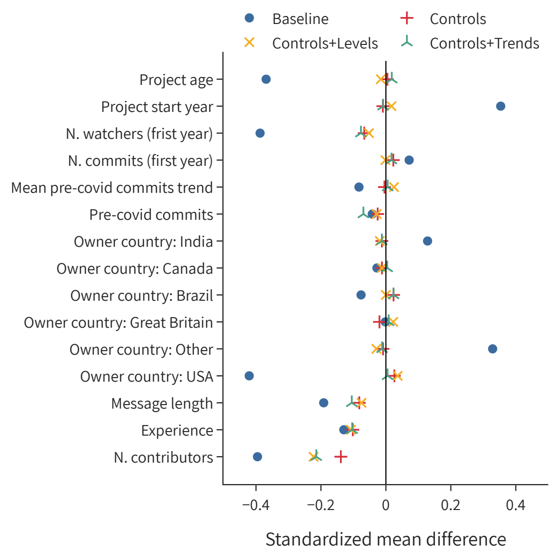
Notes: Columns (1) to (3) estimate weighted regressions using propensity scores and the inverse probability weighting method of Abadie (2005). Columns (4) to (6) estimate regressions on one-to-one matched samples using propensity score matching (Hansen and Klopfer 2006). Pre-treatment outcome levels are measured as quarterly number of commits. Pre-treatment outcome trends are measured as first-differences of quarterly number of commits. Controls include project age (at the beginning of the observational period), number of commits and watchers within a project's first year and country-of-owner fixed effects. Propensity scores are estimated using logistic regression. All regression control for project age (linear & quadratic), month fixed effects and the full set of interacted remaining controls. The sample covers 24 months centered around the hypothetical onset of the coronavirus pandemic. Estimation method: Negative-Binomial ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Figure B.3 *Co-located and distributed teams: balance of covariates*

(a) *Method: IPW*



(b) *Method: one-to-one matching*



Notes: The figures show standardized mean differences of control variables between co-located and distributed teams. The differences are calculated by subtracting the means of distributed teams from those of co-located teams and dividing by the standard deviation of co-located teams. In panel (a) teams are balanced using the inverse probability weighting approach suggested in Abadie (2005) and in panel (b) we apply one-to-one propensity score matching using the optimized approach of Hansen and Klopfer (2006). Propensity scores are estimated using logistic regressions of an indicator for co-location on project age and starting year, number of commits and watchers within a project's first year and project-owner-country fixed effects. Pre-COVID commits is the sum of commits a project has received in the 12 months before March 2020. Mean pre-COVID commits trend measures the project-mean of quarterly first-differences in the number of commits.

Table B.7 *Collaborations by team distribution: matched results*

Method	Dependent: Number of commits						
		Inverse probability weighting			One-to-one Matching		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distributed	0.084	0.037	–0.055	0.045	–0.001	0.015	0.109
	(0.093)	(0.109)	(0.116)	(0.110)	(0.097)	(0.095)	(0.095)
Distributed × COVID	0.372***	0.381***	0.393***	0.369***	0.358***	0.319**	0.354**
	(0.129)	(0.129)	(0.136)	(0.134)	(0.134)	(0.135)	(0.139)
Pseudo R^2	0.061	0.065	0.065	0.065	0.079	0.083	0.083
N	102,696	102,696	102,696	102,696	44,832	44,832	44,832
Clusters	4,279	4,279	4,279	4,279	1,868	1,868	1,868
Matching variables							
Controls		X	X	X	X	X	X
Pre-COVID outcome levels			X			X	
Pre-COVID outcome trends				X			X

Notes: Columns (1) shows a baseline regression without any matching applied. Columns (2) to (4) estimate weighted regressions using propensity scores and the inverse probability weighting method of Abadie (2005). Columns (5) to (7) estimate regressions on one-to-one matched samples using propensity score matching (Hansen and Klopfer 2006). Distributed is a dummy variable that is 1 for mixed and fully distributed teams and 0 for co-located teams. Pre-treatment outcome levels are measured as quarterly number of commits. Pre-treatment outcome trends are measured as first-differences of quarterly number of commits. Controls include project starting year and age (at the beginning of the observational period), number of commits and watchers within a project's first year and country-of-owner fixed effects. Propensity scores are estimated using logistic regression. All regression control for project age (linear & quadratic), month fixed effects and the full set of interacted remaining controls. The sample covers 24 months centered around the onset of the coronavirus pandemic in March 2020. Estimation method: Negative-Binomial ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table B.8 *Collaborations by team distribution: placebo analysis*

	Dependent: Number of commits			
	(1)	(2)	(3)	(4)
D(Locations=2)	–0.054	–0.034	–0.002	–0.001
	(0.113)	(0.119)	(0.119)	(0.107)
D(Locations=3)	0.072	0.126	0.052	0.120
	(0.109)	(0.120)	(0.122)	(0.123)
D(Locations=2) × COVID	0.015	–0.047	–0.131	–0.160
	(0.155)	(0.144)	(0.206)	(0.205)
D(Locations=3) × COVID	0.033	–0.038	–0.021	–0.150
	(0.137)	(0.135)	(0.199)	(0.202)
Placebo COVID year	2019	2019	2018	2018
Controls		X		X
Pseudo R^2	0.008	0.064	0.021	0.086
N	99,312	99,312	83,376	83,376
Clusters	4,138	4,138	3,474	3,474

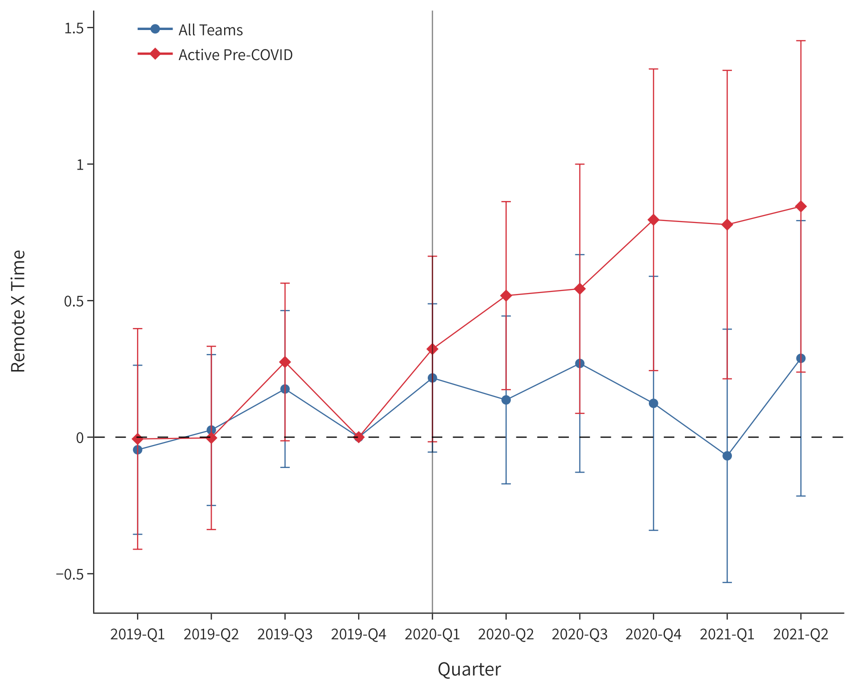
Notes: The sample in columns (1) and (2) includes projects started in 2017 or earlier and the placebo onset of the coronavirus pandemic is set to March 2019. The sample in columns (1) and (2) includes projects started in 2016 or earlier and the placebo onset of the coronavirus pandemic is set to March 2018. Co-located teams are the left-out category. Controls include project age (linear & quadratic) and starting year, number of commits and watchers within a project's first year, country-of-owner fixed effects and month fixed effects. Except for project age and month fixed effects all controls are interacted with the placebo COVID indicator. The sample covers 24 months centered around the placebo onset of the coronavirus pandemic. Estimation method: Negative-Binomial ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table B.9 *Number of contributors by team distribution*

	Dependent: Number of contributors	
	(1)	(2)
D(Loc.=2)	0.007	0.062
	(0.060)	(0.071)
D(Loc.=3)	0.244***	0.252***
	(0.059)	(0.069)
D(Loc.=2) × COVID	0.301***	0.115
	(0.107)	(0.106)
D(Loc.=3) × COVID	0.409***	0.235**
	(0.109)	(0.106)
Sample	Full	Active during COVID-19
Pseudo R^2	0.034	0.020
N	8,558	3,718

Notes: The dependent variable is the total number of contributors in the 12 months before and the 12 months after the onset of the coronavirus pandemic in March 2020. Column (1) includes all projects without any further restriction. Column (2) only keeps projects which were active during the coronavirus pandemic, i.e. received at least one contribution. Co-located teams are the left-out category. All regression control for project age (at the beginning of the observational period) and starting year, number of commits and watchers within a project's first year and country-of-owner fixed effects. Except for project age all controls are interacted with the COVID indicator. Estimation method: Negative-Binomial ML. Standard errors are clustered at the project level and shown in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Figure B.4 *Remote workers: quarterly effects*



Notes: This figure shows interaction coefficients and 95% confidence intervals of quarterly time dummies and indicators for remote workers. The dependent variable is the number of commits and controls include project age (linear & quadratic) and starting year (interacted with remote). The blue line includes the full sample. The red line limits the sample in the regression to repositories which had commits in at least nine out of the 12 months preceding the pandemic, approximately the 10% most consistently active projects. Estimation method: Negative-Binomial ML.

Appendix C

Appendix to Chapter 3

C.1 Main Appendix

Table C.1 *Summary statistics: full sample*

Variable	Observations (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Management measures					
Style 1 intensity	3,571	0.513	0.288	0.038	0.973
Style 1 intensity (first-difference)	1,762	0.005	0.303	–0.891	0.884
Share of adopted practices (SAP)	3,571	0.470	0.192	0.024	0.902
Share of adopted practices (SAP) (first-difference)	1,762	–0.003	0.127	–0.512	0.488
Industry					
Processing industry	3,595	0.315	0.465	0	1
Metal, electrical, automotive	3,595	0.275	0.447	0	1
Commerce, traffic, communication	3,595	0.161	0.368	0	1
Company related, financial services	3,595	0.153	0.360	0	1
It, communication, other services	3,595	0.096	0.294	0	1
Ownership					
Family, founder	3,551	0.496	0.500	0	1
Management, entrepreneurship	3,551	0.197	0.397	0	1
Financial investor	3,551	0.068	0.252	0	1
Widely held on stock market	3,551	0.067	0.250	0	1
Government or public sector	3,551	0.019	0.136	0	1
Other forms	3,551	0.154	0.361	0	1
Employees					
0-49	3,595	0.022	0.147	0	1
50-99	3,595	0.354	0.478	0	1
100-249	3,595	0.336	0.473	0	1
250-499	3,595	0.168	0.374	0	1
500+	3,595	0.119	0.324	0	1
Competition					
No pressure	3,587	0.035	0.183	0	1
Little pressure	3,587	0.088	0.283	0	1
Medium pressure	3,587	0.372	0.483	0	1
High pressure	3,587	0.506	0.500	0	1
Various					
Multi-plant firm	3,584	0.255	0.436	0	1
Works council	3,590	0.637	0.481	0	1
No collective agreement	3,593	0.397	0.489	0	1
Change of ownership	3,584	0.075	0.264	0	1
Change of management	3,582	0.241	0.428	0	1
Annual result					
Profit	3,595	0.792	0.406	0	1
Loss	3,595	0.081	0.273	0	1
Digital technologies					
Distribution channels	760	0.850	0.357	0	1
Big Data	749	0.198	0.398	0	1
Internet of Things	750	0.187	0.390	0	1

Notes: This table shows summary statistics of management measures and firm characteristics. The statistics are taken from the full sample of all firm-year observations.

Table C.2 *Summary statistics: panel firms*

Variable	Observations (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Management measures					
Style 1 intensity	1,288	0.497	0.294	0.040	0.973
Style 1 intensity (first-difference)	966	0.010	0.314	−0.891	0.884
Share of adopted practices (SAP)	1,288	0.463	0.192	0.049	0.902
Share of adopted practices (SAP) (first-difference)	966	−0.004	0.121	−0.463	0.488
Industry					
Processing industry	1,288	0.318	0.466	0	1
Metal, electrical, automotive	1,288	0.307	0.462	0	1
Commerce, traffic, communication	1,288	0.171	0.376	0	1
Company related, financial services	1,288	0.146	0.353	0	1
It, communication, other services	1,288	0.057	0.233	0	1
Ownership					
Family, founder	1,274	0.521	0.500	0	1
Management, entrepreneurship	1,274	0.192	0.394	0	1
Financial investor	1,274	0.058	0.234	0	1
Widely held on stock market	1,274	0.071	0.258	0	1
Government or public sector	1,274	0.024	0.152	0	1
Other forms	1,274	0.134	0.341	0	1
Employees					
0-49	1,288	0.036	0.186	0	1
50-99	1,288	0.340	0.474	0	1
100-249	1,288	0.329	0.470	0	1
250-499	1,288	0.181	0.385	0	1
500+	1,288	0.114	0.318	0	1
Competition					
No pressure	1,286	0.037	0.190	0	1
Little pressure	1,286	0.093	0.291	0	1
Medium pressure	1,286	0.373	0.484	0	1
High pressure	1,286	0.496	0.500	0	1
Various					
Multi-plant firm	1,286	0.227	0.419	0	1
Works council	1,287	0.670	0.470	0	1
No collective agreement	1,287	0.386	0.487	0	1
Change of ownership	1,286	0.070	0.255	0	1
Change of management	1,285	0.211	0.408	0	1

Notes: This table shows summary statistics of management measures and firm characteristics. The statistics are taken from the panel sample including all firm-year observations of the 322 firms that I observe in every survey wave.

Table C.3 *Summary of individual practices*

Practice	Style 1		Style 2		Share adopted		
	Rank	Loading	Rank	Loading	Full Sample	Style 1	Style 2
Development plans	1	0.040	37	0.000	0.459	0.749	0.114
Employee surveys	2	0.036	35	0.000	0.413	0.640	0.147
Development plans: Implementation	3	0.036	38	0.000	0.408	0.675	0.090
Appraisal interviews	4	0.035	10	0.039	0.703	0.836	0.543
Development plans: Management	5	0.035	41	0.000	0.396	0.660	0.082
Staffing plan	6	0.035	11	0.037	0.688	0.812	0.538
Development plans: Non-management	7	0.034	36	0.000	0.390	0.649	0.081
Appraisal interviews: Management	8	0.034	12	0.036	0.670	0.807	0.505
Target agreements	9	0.034	16	0.033	0.642	0.785	0.471
Internal training	10	0.032	2	0.054	0.791	0.872	0.694
Employee surveys: Communicated to employees	11	0.032	39	0.000	0.359	0.572	0.107
Performance appraisal	12	0.032	13	0.035	0.633	0.763	0.476
Target agreements: Management	13	0.031	19	0.029	0.599	0.751	0.416
Appraisal interviews: Non-management	14	0.031	15	0.033	0.621	0.753	0.462
Performance appraisal: Management	15	0.031	21	0.026	0.556	0.700	0.388
Inefficiency: HR development measures (high)	16	0.031	24	0.020	0.506	0.652	0.331
Analysis of age structure	17	0.031	8	0.040	0.666	0.761	0.547
Promotion of higher educational qualification	18	0.030	14	0.034	0.613	0.709	0.500
On-the-job training	19	0.030	3	0.053	0.763	0.827	0.682
Attending lectures	20	0.030	4	0.047	0.709	0.787	0.615
Performance appraisal: Non-management	21	0.029	17	0.031	0.575	0.698	0.429
Inefficiency: Discussions (high)	22	0.029	1	0.063	0.830	0.847	0.810
Employee surveys: Develop solutions	23	0.028	40	0.000	0.313	0.510	0.083
Conduction of performance appraisal	24	0.025	18	0.029	0.522	0.621	0.404
Target agreements: Non-management	25	0.025	28	0.014	0.405	0.529	0.256
Variable remuneration	26	0.024	9	0.040	0.590	0.667	0.499
Self-directed study (by media)	27	0.023	29	0.013	0.358	0.472	0.224
Increase of women in management set as goal	28	0.019	22	0.024	0.405	0.476	0.318
Recruitment: Social networks	29	0.018	25	0.020	0.366	0.439	0.273
Recruitment: Private agency	30	0.016	27	0.016	0.311	0.381	0.224
Quality/workshop meeting	31	0.016	32	0.008	0.243	0.329	0.140
Inefficiency: Another position (high)	32	0.016	26	0.018	0.326	0.375	0.268
Job rotation	33	0.014	33	0.007	0.213	0.288	0.126
Recruitment agency: Management	34	0.014	31	0.010	0.229	0.296	0.147
Inefficiency: Dismissal (high)	35	0.012	5	0.044	0.488	0.452	0.529
Inefficiency: Dismissal (low)	36	0.012	23	0.020	0.296	0.316	0.272
HR at highest management level	37	0.011	6	0.042	0.460	0.424	0.498
Recruitment agency: Non-management	38	0.009	30	0.010	0.179	0.214	0.135
Inefficiency: Another position (low)	39	0.003	7	0.042	0.363	0.274	0.465
Inefficiency: Discussions (low)	40	0.000	34	0.006	0.050	0.030	0.071
Inefficiency: HR development measures (low)	41	0.000	20	0.029	0.227	0.088	0.385
Advanced training measures					0.921	0.956	0.877
External training					0.874	0.922	0.816
Distribution recommendation for performance appraisal					0.081	0.119	0.037
Distribution recommendation: Non-management					0.071	0.103	0.033
Distribution recommendation: Management					0.067	0.103	0.026

Notes: This table shows a full list of management practices including practice loadings and ranks within each style. By construction the loadings of all practices are strictly positive and sum up to one. The top 10 practices of both styles, ordered from highest to lowest loadings, are shown in bold. It further shows adoption rates of management practices for the full sample as well as Style 1 firms and Style 2 firms. Style 1 firms are firms with Style 1 intensity ≥ 0.5 and Style 2 firms are firms with Style 1 intensity < 0.5. The five practices at the bottom are excluded from the LDA estimation, due to their low TF-IDF scores.

Table C.4 *Overview of management practices*

Practice	Question text
Development plans	Are there any development plans for employees in your establishment/office?
Employee surveys	Does your establishment/office regularly conduct employee surveys?
Development plans: Implementation	Do you systematically review the implementation of the development plans?
Appraisal interviews	Do you conduct structured appraisal interviews in your establishment/office at least once a year?
Development plans: Management	For whom are development plans available? (management staff)
Staffing plan	Does your establishment/office have a staffing plan?
Development plans: Non-management	For whom are development plans available? (employees without management responsibility)
Appraisal interviews: Management	With whom do you conduct the structured appraisal interviews? (management staff)
Target agreements	Does your establishment have target agreements?
Internal training (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Internal training courses, seminars or workshops
Employee surveys: Communicated to employees	Are the results of the survey communicated to all employees?
Performance appraisal	Is a review of the performance of the employees carried out by the respective supervisor in your establishment/office at least once a year?
Target agreements: Management	For whom are the target agreements available? (management staff)
Appraisal interviews: Non-management	With whom do you conduct the structured appraisal interviews? (employees without management responsibility)
Performance appraisal: Management	For whom are the annual performance appraisals issued? (management staff)
Inefficiency: HR development measures (high)	How do you and your management staff deal with employees, whose performance is not satisfactory? HR development measures are purposefully offered to correct performance problems.
Analysis of age structure	Do you systematically analyze the age structure of employees in your establishment/office?
Promotion of higher educational qualification	Have you actively promoted employees' qualification activities leading to a higher educational qualification, e.g. by releasing from work or partially bearing costs? This includes e.g. further training to master craftsmen, technician, postgraduate program, MBA, doctorate.
On-the-job training (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Further training on the job (instruction, familiarization training)
Attending lectures (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Participation in lectures, symposia, fairs, etc.

Notes: This table lists survey questions to the related management practices. A detailed data report of the last LPP survey wave (2018) can be found in Mackeben et al. (2020b). (IAB BP) indicates that these practices are taken from the IAB Establishment Panel (Bellmann et al. 2019). The remaining practices are all taken from the LPP (Mackeben et al. 2020a).
This table is continued on the next page.

Table C.4 *Overview of management practices (cont'd)*

Practice	Question text
Performance appraisal: Non-management	For whom are the annual performance appraisals issued? (employees without management responsibility)
Inefficiency: Discussions (high)	How do you and your management staff deal with employees, whose performance is not satisfactory? The management staff openly discusses the problems with the employee in question.
Employee surveys: Develop solutions	Is there a systematic process to develop solutions for flaws, which were identified in the employee surveys?
Conduction of performance appraisal	Is the performance appraisal generally conducted by just one superior or collectively by a group of superiors (evaluation round), meaning not only by one superior?
Target agreements: Non-management	For whom are the target agreements available? (employees without management responsibility)
Variable remuneration	Does your establishment/office have a salary system with variable proportions?
Self-directed study (by media)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Self-directed study (e.g. by means of computer-aided self-learning programs or reference books)
Increase of women in management set as goal	Do you pursue the goal to increase the proportion of women in management positions?
Recruitment: Social networks	Have you directly addressed applicants employed by another company via social networks such as Xing, LinkedIn etc. in the past two years?
Recruitment: Private agency	Have you recruited applicants in the past two years, who were employed by another company, with the help of a private recruitment agency or HR consulting?
Quality/workshop meeting (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Quality circles, workshop circles, learning workshop, continuous improvement teams
Inefficiency: Another position (high)	How do you and your management staff deal with employees, whose performance is not satisfactory? We try to find another position in the establishment/office if there are permanent performance problems.
Job rotation (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? Job rotation

Notes: This table lists survey questions to the related management practices. A detailed data report of the last LPP survey wave (2018) can be found in Mackeben et al. (2020b). (IAB BP) indicates that these practices are taken from the IAB Establishment Panel (Bellmann et al. 2019). The remaining practices are all taken from the LPP (Mackeben et al. 2020a).
This table is continued on the next page.

Table C.4 *Overview of management practices (cont'd)*

Practice	Question text
Recruitment agency: Management	Have you recruited applicants in the past two years, who were employed by another company, with the help of a private recruitment agency or HR consulting? (management staff)
Inefficiency: Dismissal (high)	How do you and your management staff deal with employees, whose performance is not satisfactory? Employees who permanently show poor working performance will be dismissed or urged to leave the establishment/office.
Inefficiency: Dismissal (low)	How do you and your management staff deal with employees, whose performance is not satisfactory? Employees who permanently show poor working performance will be dismissed or urged to leave the establishment/office.
HR at highest management level	On which level is the management of the human resources (HR) located in your company? On the first management level, that means executive board or management?
Recruitment agency: Non-management	Have you recruited applicants in the past two years, who were employed by another company, with the help of a private recruitment agency or HR consulting? (employees without management responsibility)
Inefficiency: Another position (low)	How do you and your management staff deal with employees, whose performance is not satisfactory? We try to find another position in the establishment/office if there are permanent performance problems.
Inefficiency: Discussions (low)	How do you and your management staff deal with employees, whose performance is not satisfactory? The management staff openly discusses the problems with the employee in question.
Inefficiency: HR development measures (low)	How do you and your management staff deal with employees, whose performance is not satisfactory? HR development measures are purposefully offered to correct performance problems.
Advanced training measures (IAB BP)	Did your establishment/office support training courses in the first half of this year?
External training (IAB BP)	For which of the following internal or external training courses did your establishment release staff and cover the expenses in full or in part? External training courses, seminars or workshops
Distribution recommendation for performance appraisal	Do you have recommendations regarding distribution of performance appraisal? Recommendations regarding performance appraisal include information on what percentage of employees should, for instance, receive the best performance appraisal, the second-best performance appraisal etc.
Distribution recommendation: Non-management	Do you have recommendations regarding distribution of performance appraisal? Recommendations regarding performance appraisal include information on what percentage of employees should, for instance, receive the best performance appraisal, the second-best performance appraisal etc. (employees without management responsibility)
Distribution recommendation: Management	Do you have recommendations regarding distribution of performance appraisal? Recommendations regarding performance appraisal include information on what percentage of employees should, for instance, receive the best performance appraisal, the second-best performance appraisal etc. (management staff)

Notes: This table lists survey questions to the related management practices. A detailed data report of the last LPP survey wave (2018) can be found in Mackeben et al. (2020b). (IAB BP) indicates that these practices are taken from the IAB Establishment Panel (Bellmann et al. 2019). The remaining practices are all taken from the LPP (Mackeben et al. 2020a).

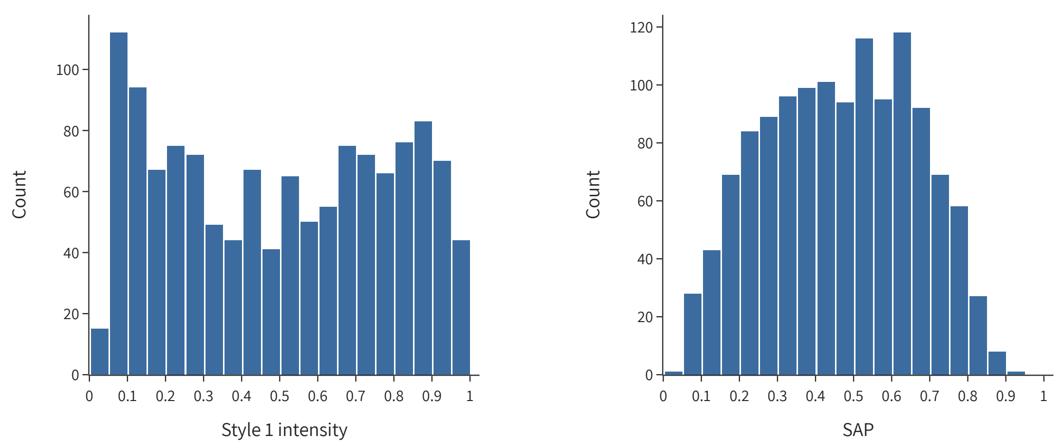
Table C.5 *Management styles and SAP: correlations*

	Style 1	Style 1 (alt.)	SAP
Style 1	1.000		
Style 1 (alt.)	0.709	1.000	
SAP	0.636	0.663	1.000

Notes: This table shows correlations of management measures. Style 1 indicates the intensity of the structured management style, resulting from LDA. Style 1 (alt.) is an alternative measure without TF-IDF elimination. SAP is simply the share of adopted practices.

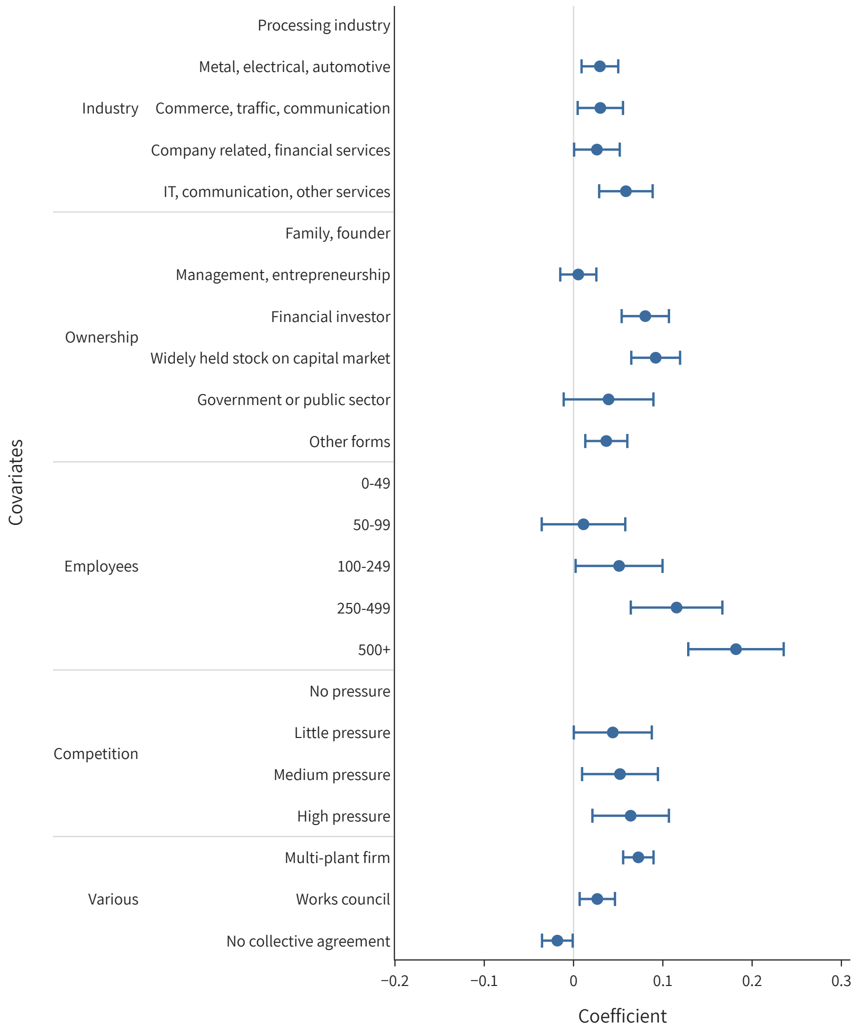
Figure C.1 *Distributions of management measures – panel firms*

(a) *Style 1 intensity* **(b)** *Share of adopted practices (SAP)*



Notes: This figure shows histograms of both management measures. Panel (a) shows the distribution of Style 1 intensity and panel (b) the distribution of the SAP. The statistics are taken from the panel sample including all firm-year observations of the 322 firms that I observe in every survey wave.

Figure C.2 *Share of adopted practices (SAP) correlates*



Notes: This figure shows coefficients and 95% confidence intervals from an OLS regression of the SAP on firm characteristics. All regressors are either dummies or categorical variables. Reference categories are those without coefficient indicators. The regression is estimated on a pooled sample including all firm-year observations. The number of observations is $N = 3,508$ and standard errors are clustered at the firm-level.

Table C.6 *Management levels: Style 1 intensity – full sample*

	Dependent: Style 1 intensity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry (Ref: Processing industry)								
Metal, electrical, automotive	0.046**							0.019
	(0.017)							(0.015)
Commerce, traffic, communication	0.030							0.031
	(0.020)							(0.018)
Company, financial services	0.022							0.021
	(0.020)							(0.019)
IT, communication, other services	0.043*							0.043*
	(0.021)							(0.021)
Principal owner (Ref: Family, founder)								
Management, entrepreneurship		0.053***						0.032*
		(0.015)						(0.015)
Financial investor		0.121***						0.073***
		(0.022)						(0.021)
Listed on stock market		0.217***						0.103***
		(0.020)						(0.020)
Government or public sector		0.108**						0.044
		(0.042)						(0.040)
Other forms		0.095***						0.040*
		(0.017)						(0.017)
Competition (Ref: No pressure)								
Little pressure			0.011					0.014
			(0.032)					(0.030)
Medium pressure			0.051					0.037
			(0.030)					(0.028)
High pressure			0.056					0.031
			(0.030)					(0.028)
Firm size (Ref: Employees: 0-49)								
Employees: 50-99				−0.028				−0.026
				(0.040)				(0.039)
Employees: 100-249				0.049				0.026
				(0.040)				(0.040)
Employees: 250-499				0.127**				0.083*
				(0.042)				(0.041)
Employees: 500+				0.224***				0.161***
				(0.042)				(0.042)
Dummy indicators								
Multiplant firm					0.132***			0.076***
					(0.012)			(0.013)
Works council						0.138***		0.055***
						(0.012)		(0.014)
No Collective agreement							−0.097***	−0.019
							(0.012)	(0.013)
Intercept	0.488***	0.462***	0.465***	0.458***	0.478***	0.425***	0.551***	0.353***
	(0.012)	(0.009)	(0.029)	(0.039)	(0.007)	(0.010)	(0.008)	(0.049)
Adj. R^2	0.003	0.044	0.002	0.081	0.039	0.053	0.027	0.129
Observations	3,571	3,530	3,563	3,571	3,561	3,567	3,569	3,508
Cluster	1,773	1,761	1,771	1,773	1,769	1,772	1,773	1,754

Notes: The dependent variable is the intensity of Style 1. The regressions are based on the full sample of firm-year observations. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.7 *Management levels: Style 1 intensity – panel firms*

	Dependent: Style 1 intensity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry (Ref: Processing industry)								
Metal, electrical, automotive	0.049							0.034
	(0.032)							(0.029)
Commerce, traffic, communication	0.036							0.045
	(0.038)							(0.033)
Company, financial services	0.060							0.057
	(0.040)							(0.035)
IT, communication, other services	0.010							0.022
	(0.049)							(0.047)
Principal owner (Ref: Family, founder)								
Management, entrepreneurship		0.052						0.034
		(0.028)						(0.027)
Financial investor		0.070						0.039
		(0.045)						(0.043)
Listed on stock market		0.237***						0.115**
		(0.035)						(0.036)
Government or public sector		0.167*						0.120
		(0.070)						(0.065)
Other forms		0.125***						0.062
		(0.036)						(0.035)
Competition (Ref: No pressure)								
Little pressure			−0.015					−0.012
			(0.051)					(0.047)
Medium pressure			0.045					0.025
			(0.050)					(0.046)
High pressure			0.052					0.019
			(0.051)					(0.047)
Firm size (Ref: Employees: 0-49)								
Employees: 50-99				−0.036				−0.031
				(0.055)				(0.053)
Employees: 100-249				0.045				0.018
				(0.058)				(0.057)
Employees: 250-499				0.099				0.050
				(0.061)				(0.059)
Employees: 500+				0.202**				0.135*
				(0.061)				(0.063)
Dummy indicators								
Multiplant firm					0.167***			0.103***
					(0.023)			(0.025)
Works council						0.147***		0.067*
						(0.024)		(0.027)
No Collective agreement							−0.103***	−0.017
							(0.023)	(0.025)
Intercept	0.467***	0.445***	0.456***	0.454***	0.459***	0.399***	0.537***	0.337***
	(0.023)	(0.016)	(0.049)	(0.054)	(0.013)	(0.019)	(0.016)	(0.072)
Adj. R^2	0.003	0.053	0.003	0.062	0.056	0.055	0.028	0.133
Observations	1,288	1,274	1,286	1,288	1,286	1,287	1,287	1,268
Cluster	322	322	322	322	322	322	322	322

Notes: The dependent variable is the intensity of Style 1. The regressions are based on the panel sample firm-year observations of firms which I observe in every survey wave. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.8 *Management levels: Share of adopted practices (SAP) – full sample*

	Dependent: SAP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry (Ref: Processing industry)								
Metal, electrical, automotive	0.052***							0.030**
	(0.013)							(0.010)
Commerce, traffic, communication	0.029							0.030*
	(0.015)							(0.013)
Company, financial services	0.029							0.026*
	(0.015)							(0.013)
IT, communication, other services	0.051**							0.059***
	(0.016)							(0.015)
Principal owner (Ref: Family, founder)								
Management, entrepreneurship		0.021						0.005
		(0.011)						(0.010)
Financial investor		0.122***						0.080***
		(0.014)						(0.014)
Listed on stock market		0.192***						0.092***
		(0.014)						(0.014)
Government or public sector		0.081**						0.039
		(0.031)						(0.026)
Other forms		0.083***						0.037**
		(0.012)						(0.012)
Competition (Ref: No pressure)								
Little pressure			0.039					0.044*
			(0.024)					(0.022)
Medium pressure			0.061*					0.052*
			(0.024)					(0.022)
High pressure			0.083***					0.064**
			(0.024)					(0.022)
Firm size (Ref: Employees: 0-49)								
Employees: 50-99				0.010				0.011
				(0.026)				(0.024)
Employees: 100-249				0.071**				0.051*
				(0.027)				(0.025)
Employees: 250-499				0.150***				0.115***
				(0.028)				(0.026)
Employees: 500+				0.235***				0.182***
				(0.028)				(0.027)
Dummy indicators								
Multiplant firm					0.119***			0.073***
					(0.009)			(0.009)
Works council						0.106***		0.027**
						(0.009)		(0.010)
No Collective agreement							−0.084***	−0.018*
							(0.009)	(0.009)
Intercept	0.442***	0.430***	0.403***	0.390***	0.440***	0.403***	0.504***	0.279***
	(0.009)	(0.007)	(0.023)	(0.026)	(0.005)	(0.007)	(0.006)	(0.034)
Adj. R^2	0.011	0.085	0.009	0.152	0.073	0.070	0.046	0.245
Observations	3,571	3,530	3,563	3,571	3,561	3,567	3,569	3,508
Cluster	1,773	1,761	1,771	1,773	1,769	1,772	1,773	1,754

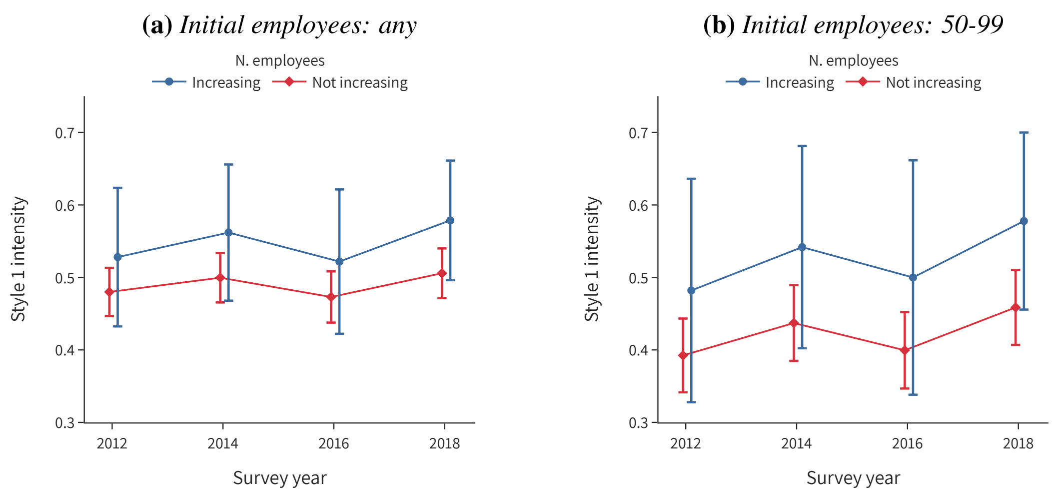
Notes: The dependent variable is the SAP. The regressions are based on the full sample of firm-year observations. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.9 *Management levels: Share of adopted practices (SAP) – panel firms*

	Dependent: SAP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Industry (Ref: Processing industry)								
Metal, electrical, automotive	0.050*							0.037
	(0.025)							(0.020)
Commerce, traffic, communication	0.024							0.029
	(0.029)							(0.025)
Company, financial services	0.059							0.052*
	(0.032)							(0.025)
IT, communication, other services	0.027							0.038
	(0.037)							(0.034)
Principal owner (Ref: Family, founder)								
Management, entrepreneurship		0.028						0.015
		(0.022)						(0.020)
Financial investor		0.107***						0.082***
		(0.024)						(0.023)
Listed on stock market		0.214***						0.099***
		(0.026)						(0.025)
Government or public sector		0.099						0.065
		(0.055)						(0.045)
Other forms		0.100***						0.047
		(0.026)						(0.025)
Competition (Ref: No pressure)								
Little pressure			0.043					0.042
			(0.041)					(0.031)
Medium pressure			0.068					0.048
			(0.043)					(0.032)
High pressure			0.084					0.051
			(0.044)					(0.033)
Firm size (Ref: Employees: 0-49)								
Employees: 50-99				0.006				0.006
				(0.039)				(0.037)
Employees: 100-249				0.065				0.042
				(0.039)				(0.038)
Employees: 250-499				0.147***				0.107**
				(0.041)				(0.041)
Employees: 500+				0.234***				0.184***
				(0.043)				(0.044)
Dummy indicators								
Multiplant firm					0.155***			0.106***
					(0.018)			(0.017)
Works council						0.107***		0.024
						(0.019)		(0.020)
No Collective agreement							−0.092***	−0.019
							(0.018)	(0.017)
Intercept	0.433***	0.420***	0.392***	0.386***	0.427***	0.391***	0.498***	0.287***
	(0.018)	(0.012)	(0.042)	(0.037)	(0.010)	(0.015)	(0.012)	(0.053)
Adj. R^2	0.011	0.101	0.007	0.153	0.114	0.068	0.054	0.282
Observations	1,288	1,274	1,286	1,288	1,286	1,287	1,287	1,268
Cluster	322	322	322	322	322	322	322	322

Notes: The dependent variable is the SAP. The regressions are based on the panel sample firm-year observations of firms which I observe in every survey wave. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Figure C.3 *Change of management by employee growth*



Notes: This figure shows Style 1 intensities across survey-years split by firms that move to a higher workforce category and firms that don't. The markers are slightly shifted to enhance the readability of the figure. Panel (a) shows mean values and 95% confidence intervals of Style 1 intensity for all initial workforce sizes. Panel (b) shows mean values and 95% confidence intervals of Style 1 intensity only for firms with initially 50-99 employees.

Table C.10 *Management dynamics: Style 1 – change of ownership or managers*

	Dependent: Total change of Style 1 intensity[a]				
	(1)	(2)	(3)	(4)	(5)[a]
Change of ownership	–0.004		0.004	–0.002	–0.016
	(0.038)		(0.044)	(0.050)	(0.028)
Change of management		–0.017	–0.019	–0.015	0.016
		(0.035)	(0.036)	(0.041)	(0.026)
Intercept	0.030	0.038	0.038	0.176	0.164*
	(0.020)	(0.025)	(0.025)	(0.097)	(0.070)
Controls				**X**	**X**
Adj. R^2	-0.003	-0.002	-0.005	0.014	-0.016
Observations	322	322	322	318	318

Notes: The dependent variable is the total change of Style 1 intensity from 2012 to 2018. Change of ownership indicates that the ownership structure of a firm has changed at least once between 2012 and 2018. Change of management indicates that the management staff of a firm has changed at least once between 2012 and 2018. Robust standard errors are reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.
[a] The dependent variable in column (5) is total change of Style 1 intensity in absolute terms. Thus, the coefficients in column (5) indicate changes of Style 1 intensity in any direction.

Table C.11 *Management dynamics: SAP – change of ownership or managers*

	Dependent: Total change of SAP[a]				
	(1)	(2)	(3)	(4)	(5)[a]
Change of ownership	–0.033		–0.045*	–0.046*	0.008
	(0.020)		(0.022)	(0.023)	(0.017)
Change of management		0.014	0.027	0.021	–0.016
		(0.017)	(0.019)	(0.022)	(0.015)
Intercept	–0.003	–0.018	–0.015	0.094	0.134***
	(0.010)	(0.013)	(0.013)	(0.050)	(0.036)
Controls				**X**	**X**
Adj. R^2	0.005	-0.001	0.009	0.014	0.002
Observations	322	322	322	318	318

Notes: This table summarizes regressions results of management measures on indicators for changes of principal owners or managers. Robust standard errors are reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

[a] The dependent variable in column (5) is the absolute total change of the SAP.

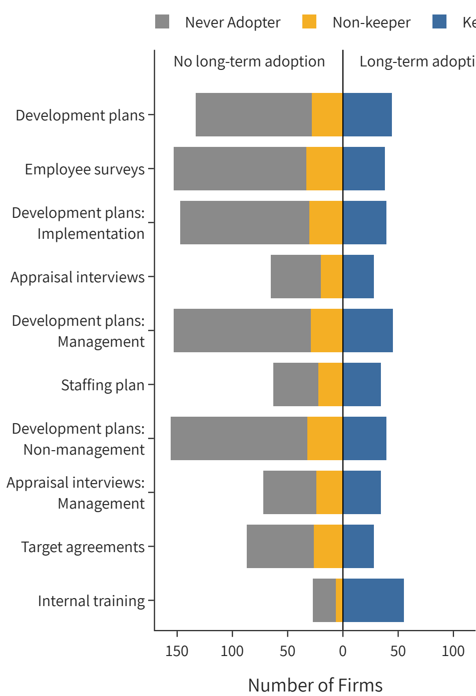
Table C.12 *Management dynamics*

	Dependent: Change Style 1 intensity			Dependent: Change SAP		
	FD (1)	FD (2)	Total (3)	FD (4)	FD (5)	Total (6)
Industry (Ref: Processing industry)						
Metal, electrical, automotive	0.017	0.000	0.009	0.005	0.001	–0.001
	(0.014)	(0.020)	(0.045)	(0.006)	(0.007)	(0.022)
Commerce, traffic, communication	0.007	0.001	–0.004	0.001	0.012	0.029
	(0.016)	(0.017)	(0.053)	(0.007)	(0.009)	(0.027)
Company, financial services	–0.005	–0.004	–0.081	–0.005	0.003	–0.007
	(0.019)	(0.019)	(0.057)	(0.008)	(0.009)	(0.026)
IT, communication, other services	–0.023	–0.041	–0.211**	–0.013	–0.024	–0.084*
	(0.023)	(0.027)	(0.079)	(0.012)	(0.013)	(0.039)
Principal owner (Ref: Family, founder)						
Management, entrepreneurship	0.012	0.031	0.075	0.006	0.007	0.029
	(0.017)	(0.024)	(0.049)	(0.007)	(0.067)	(0.023)
Financial investor	0.013	–0.001	0.068	0.014	0.000	0.034
	(0.028)	(0.042)	(0.088)	(0.012)	(0.021)	(0.042)
Listed on stock market	0.030	0.024	0.136	0.019	0.012	–0.005
	(0.024)	(0.031)	(0.078)	(0.011)	(0.014)	(0.037)
Government or public sector	0.019	0.055	0.155*	0.021*	0.014	0.046
	(0.032)	(0.032)	(0.077)	(0.011)	(0.012)	(0.028)
Other forms	–0.003	–0.015	0.113*	0.010	–0.004	0.041
	(0.019)	(0.024)	(0.054)	(0.009)	(0.011)	(0.029)
Competition (Ref: No pressure)						
Little pressure	–0.059	–0.039	–0.167	–0.010	0.007	–0.109*
	(0.039)	(0.049)	(0.098)	(0.018)	(0.021)	(0.054)
Medium pressure	–0.031	0.006	–0.163	–0.018	–0.009	–0.123**
	(0.033)	(0.045)	(0.092)	(0.016)	(0.018)	(0.044)
High pressure	–0.039	–0.001	–0.135	–0.025	–0.018	–0.146***
	(0.033)	(0.039)	(0.091)	(0.016)	(0.018)	(0.044)
Firm size (Ref: Employees: 0-49)[a]						
Employees: 50-99	–0.029	–0.051		0.028	0.022	
	(0.033)	(0.037)		(0.017)	(0.020)	
Employees: 100-249	–0.024	–0.062	–0.086	0.029	0.020	0.001
	(0.032)	(0.036)	(0.044)	(0.016)	(0.020)	(0.024)
Employees: 250-499	–0.030	–0.069	–0.012	0.023	0.016	0.001
	(0.034)	(0.039)	(0.059)	(0.017)	(0.020)	(0.029)
Employees: 500+	–0.033	–0.046	–0.094	0.027	0.031	0.022
	(0.036)	(0.042)	(0.072)	(0.017)	(0.021)	(0.031)
Dummy indicators						
Multiplant firm	0.001	–0.001	–0.033	–0.002	–0.004	0.013
	(0.014)	(0.021)	(0.045)	(0.007)	(0.009)	(0.023)
Works council	0.012	0.027	0.015	–0.006	0.005	–0.006
	(0.015)	(0.017)	(0.044)	(0.007)	(0.008)	(0.021)
No Collective agreement	0.003	0.016	0.048	–0.007	–0.007	0.020
	(0.015)	(0.017)	(0.040)	(0.012)	(0.008)	(0.021)
Change of owner	–0.029	–0.006	–0.002	0.007	0.018	–0.046*
	(0.026)	(0.038)	(0.050)	(0.012)	(0.016)	(0.023)
Change of management	0.004	–0.004	–0.015	0.000	–0.009	0.021
	(0.018)	(0.023)	(0.041)	(0.007)	(0.009)	(0.022)
Intercept	0.024	0.029	0.159	0.000	–0.005	0.069
	(0.051)	(0.065)	(0.101)	(0.024)	(0.033)	(0.052)
Panel firms		X	X		X	X
Adj. R^2	-0.007	-0.014	0.014	-0.004	-0.009	0.014
Observations	1,730	949	318	1,730	949	318
Cluster	922	320		922	320	

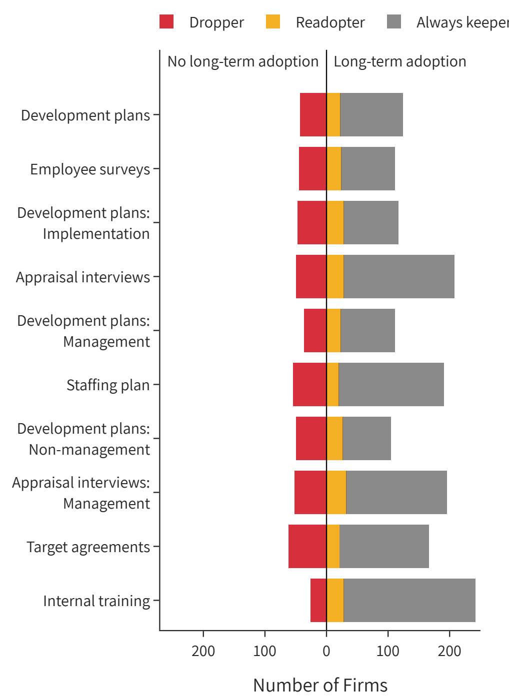
Notes: The dependent variables in columns (1)–(2) and (4)–(5) are the first-difference of Style 1 intensity and SAP, respectively. The dependent variables in columns (3) and (6) are the total difference of Style 1 intensity and SAP. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.
[a]The reference category "Employees: 0-49" was only introduced in the second wave. In the regressions of columns (3) and (6) firm characteristics are evaluated at the first wave and therefore in these regressions the reference category is "Employees: 50-99".

Figure C.4 *Practice adopters and droppers*

(a) *Structured practices – Adopters*



(b) *Structured practices – Droppers*



Notes: This figure shows patterns of adopting and dropping the ten most characteristic practices of management Style 1. Panel (a) considers all firms which had a practice adopted at the beginning of the sample and shows how many of those adopted and permanently kept (blue), adopted but dropped again (yellow) or never adopted (gray) this practice. Panel (b) considers all firms which had a practice not adopted at the beginning of the sample and shows how many of those firms dropped (red), dropped but readopted (yellow) or always kept (gray) this practice.

C.2 Additional management correlates

This section introduces additional correlates with management styles. The aspects analyzed here are neither firm nor environment characteristics, but could be important complements or outcomes of structured management.[1] Although, some of these results are covered in the accompanying paper (Englmaier et al. 2023), they strongly support my discussion in Section 3.5 of this paper. To ensure Chapter 3 can be read independently, I briefly describe the methodology and results of these additional analyses.

I estimate conditional correlations of additional firm-level variables and management measures using the following specification:

$$y_{it} = \alpha + \beta * \theta_{it} + \boldsymbol{X}_{it}\boldsymbol{\delta} + \eta_t + \varepsilon_{it}, \tag{C.1}$$

where y_{it} is the variable of interest for firm i in year t, θ_{it} is one of the management measures, $\boldsymbol{X}_{it}$ are firm-level controls[2] and η_t denote year-fixed effects. The regressions are estimated using the pooled sample of all firm-year observations and dependent variables as well as management measures are z-score standardized.

First, I consider firm success. Balance sheet data or productivity measures are not available in my setting, but the employer survey asks whether the annual result of the past year was positive, neutral or negative. I define two indicators: *Profit* and *Loss*, which are one if the annual result was positive or negative, respectively, and zero otherwise. The results are summarized in Table C.15 and show that structured management is associated with a higher likelihood for positive and lower likelihood for negative annual results. These correlations are consistent with previous (causal) findings (Bloom et al. 2012a) that structured management leads to higher productivity. Second, I analyze the correlation of structured management and

[1] I do not claim to identify causal channels from structured management to the respective variables and in some parts explicitly suggest effects in both directions.

[2] Controls include indicators of firm characteristics: Industry, size, region, multiplant firm, ownership, collective agreement, works council, competition and changes of ownership and management.

the use of digital tools and data, which as discussed in Section 3.5 could be complementing each other. Table C.16 documents consistent positive correlations of both management measures with the usage of three digital tools: Digital distributions channels, big data and the Internet of Things.[3]

I further describe correlations with variables from the employee survey, which surveys multiple employees of the firms covered in my previous analysis. It thus complements the firm-level data with linked employee-level information on, among other topics, corporate culture, job satisfaction and commitment, work-life balance as well as health, personal characteristics and socio-demographic variables. A more detailed description of the survey can be found in Kampkötter et al. (2016). Table C.13 and Table C.14 show summary statistics of the available information. Here, I focus on two topics, employee satisfaction and corporate culture, and estimate the following specifications:

$$y_{jit} = \alpha + \beta * \theta_{it} + \boldsymbol{X}_{it}\boldsymbol{\gamma} + \boldsymbol{Z}_{jt}\boldsymbol{\delta} + \eta_t + \varepsilon_{it}. \tag{C.2}$$

y_{jit} denotes the variable of interest for employee j of firm i at time t, θ_{it} is one of the management measures, $\boldsymbol{X}_{it}$ and $\boldsymbol{Z}_{jt}$ are firm-level[4] and employee-level[5] controls and η_t denote year-fixed effects. The regressions are estimated using the pooled sample of slightly less than 15,000 employee-firm-year observations and dependent variables as well as management measures are z-score standardized. First, Table C.17 focuses on indicators of employee satisfaction. Structured management is associated with lower levels of turnover intention and

[3] All three variables are dummies indicating whether the respective technology is used.

[4] Firm-level controls include indicators for industry, size, region, mutliplant firm, ownership, collective agreement.

[5] Employee-level controls include indicators for management position, functional area, employment situation, full-time employment, education, training qualification, net income, year of birth, gender, household size and relationship status.

higher levels of job as well as income satisfaction.[6] These correlations provide suggestive evidence that employees prefer structured management styles. Second, I analyze correlations of management styles and corporate culture, more specifically how employees assess qualities of their supervisors. Table C.18 summarizes the results. I document strong positive correlations between Style 1 intensity (SAP) and supervisors being perceived as fair and understanding, confident in their employees and offering good guidance.[7] As discussed in Section 3.5, I expect synergy effects between corporate culture and management practices, which would make having the right corporate culture an important requirement for the success of structured management styles.

[6]Job and income satisfaction are measured on a scale from 0 ("totally unhappy") to 10 ("totally happy"). Turnover intention measures how often employees think about changing their job and ranges from 1 ("daily") to 5 ("never"). In the regressions all three variables are z-score standardized.

[7]All indicators are measured on an agreement scale from 1 ("does not apply at all") to 5 ("fully applies").

Table C.13 *Summary statistics: employee survey – controls*

Variable	Observations (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Demographics					
Year of birth	19,469	1,967.649	10.386	1,942	1,998
Female	19,469	0.274	0.446	0	1
Household size	19,441	2.774	1.221	1	14
In relationship	19,437	1.156	0.363	1	2
Education					
No qualification	19,424	0.005	0.069	0	1
Lower secondary school	19,424	0.218	0.413	0	1
Intermediate secondary school	19,424	0.417	0.493	0	1
Vocational diploma	19,424	0.113	0.317	0	1
A-level	19,424	0.239	0.427	0	1
Other	19,424	0.008	0.087	0	1
Training qualification					
Apprenticeship	19,454	0.456	0.498	0	1
Vocational training	19,454	0.092	0.289	0	1
College of advanced vocational studies	19,454	0.206	0.404	0	1
University of applied science	19,454	0.099	0.299	0	1
University degree	19,454	0.113	0.317	0	1
Other	19,454	0.005	0.069	0	1
None	19,454	0.021	0.143	0	1
Bachelor	19,454	0.008	0.091	0	1
Employment situation					
Worker	19,464	0.370	0.483	0	1
Employee	19,464	0.630	0.483	0	1
Full-time employment	19,448	0.873	0.333	0	1
Part-time employment	19,448	0.127	0.333	0	1
Management position	19,446	0.292	0.455	0	1
Functional area					
Production	12,982	0.410	0.492	0	1
Sales, marketing	12,982	0.113	0.317	0	1
Administration	12,982	0.167	0.373	0	1
Services	12,982	0.310	0.462	0	1

Notes: This table shows summary statistics of employee characteristics. The statistics are from the full sample including all firm-employee-year observations.

Table C.14 *Summary statistics: employee survey – outcomes*

Variable	Observations (1)	Mean (2)	St. Dev. (3)	Min (4)	Max (5)
Satisfaction					
Job satisfaction	19,457	7.458	1.775	0	10
Income satisfaction	19,453	6.872	2.096	0	10
Turnover intention	19,441	1.594	0.920	1	5
Perceived job security	19,444	2.571	0.615	1	3
Commitment					
Stay rest of my life at firm	19,417	4.087	1.142	1	5
Emotionally attached to firm	19,390	3.786	1.214	1	5
Consider problems at work my own	19,435	2.856	1.301	1	5
Personal meaning	19,439	3.762	1.175	1	5
Part of the company family	19,370	3.782	1.204	1	5
Feel a sense of belonging to firm	19,405	3.880	1.181	1	5
Fairness					
Income	19,424	3.540	1.149	1	5
Decision procedures	19,268	3.406	0.995	1	5
Supervisor	19,408	3.924	0.952	1	5
Work-life balance					
Time pressure at work	19,451	3.579	1.214	1	5
Work interferes with private life	19,445	2.211	1.165	1	5
Work interferes with private responsibilities	19,450	2.270	1.201	1	5
Work strain interferes with private life	19,451	2.425	1.205	1	5
Put off doing things at work	19,445	1.621	0.833	1	5
Things at work don't get done	19,440	1.460	0.724	1	5
Private life interferes with work	19,453	1.600	0.927	1	5
Corporate culture					
Create meaning through work	7,756	3.790	1.065	1	5
Supervisors show understanding	19,429	3.722	0.982	1	5
Supervisors offer good guidance	19,409	3.529	1.034	1	5
Supervisors show confidence	19,418	3.773	1.010	1	5
Good understanding of corporate culture	19,366	3.793	1.002	1	5
Long-term plans are clear	19,383	3.563	1.182	1	5

Notes: This table shows summary statistics of employee-level outcomes. The statistics are from the full sample including all firm-employee-year observations.

Table C.15 *Management effects: annual result*

	Dependent variable			
	Profit		Loss	
	(1)	(2)	(3)	(4)
Style 1 intensity	0.022**		–0.016**	
	(0.008)		(0.005)	
SAP		0.054***		–0.029***
		(0.009)		(0.006)
Adj. R^2	0.044	0.055	0.036	0.043
Observations	3,453	3,492	3,453	3,492

Notes: All specifications include employer controls and year-fixed effects. Style 1 intensity and SAP are both z-score standardized. The dependent variable in columns (1) and (2) is and indicator if a firm's annual results was positive, rather than neutral or negative. The dependent variable in columns (3) and (4) is and indicator if a firm's annual results was negative. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.16 *Management effects: digital technologies*

	Dependent variable					
	Distribution channels		Big Data		Internet of Things	
	(1)	(2)	(3)	(4)	(5)	(6)
Style 1 intensity	0.029*		0.054***		0.054***	
	(0.015)		(0.015)		(0.015)	
SAP		0.087***		0.095***		0.065***
		(0.015)		(0.016)		(0.015)
Adj. R^2	0.039	0.082	0.074	0.104	0.076	0.083
Observations	730	730	719	719	720	720

Notes: All specifications include employer controls and year-fixed effects. Style 1 intensity and SAP are both z-score standardized. In columns (1) and (2) the dependent variable indicates usage of digital distribution channels. In columns (3) and (4) the dependent variable indicates usage of Big Data. In columns (5) and (7) the dependent variable indicates usage of the Internet of Things. All dependent variables are self-reported. Standard errors are clustered at the firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.17 *Management effects: employee satisfaction*

	Dependent variable					
	Turnover intention		Job satisfaction		Income satisfaction	
	(1)	(2)	(3)	(4)	(5)	(6)
Style 1 intensity	−0.040*		0.056***		0.056***	
	(0.016)		(0.014)		(0.016)	
SAP		−0.080***		0.070***		0.118***
		(0.016)		(0.015)		(0.018)
Adj. R^2	0.098	0.101	0.037	0.037	0.134	0.141
Observations	11,472	11,472	11,475	11,475	11,475	11,475

Notes: All specifications include employer controls, employee controls and year-fixed effects. Style 1 intensity, SAP and dependent variables are all z-score standardized. Standard errors are clustered at the employee-firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%.

Table C.18 *Management synergies: corporate culture*

	Dependent variable: Supervisor qualities							
	Fairness		Understanding		Guidance		Confidence	
	(1)	(2)	(3)	(4)	(5)	(6)	(5)	(6)
Style 1 intensity	0.062***		0.066***		0.061***		0.044**	
	(0.013)		(0.013)		(0.015)		(0.014)	
SAP		0.080***		0.091***		0.088***		0.079***
		(0.014)		(0.014)		(0.015)		(0.015)
Adj. R^2	0.019	0.020	0.027	0.029	0.039	0.041	0.027	0.029
Observations	11,451	11,451	11,463	11,463	11,457	11,457	11,461	11,461

Notes: All specifications include employer controls, employee controls and year-fixed effects. Style 1 intensity, SAP and dependent variables are all z-score standardized. Standard errors are clustered at the employee-firm level and reported in parentheses. ***, **, * indicate significance at 1%, 5% and 10%. The original statements are as follows and original agreement scales range from one (fully agree) to five (fully disagree): *Fairness*: My direct supervisor treats me fairly in all aspects of work. *Understanding*: Supervisors show understanding for employees. *Guidance*: Supervisors offer good guidance to employees. *Confidence*: Supervisors show confidence in employees.

Bibliography

Abadie, Alberto (Jan. 2005). “Semiparametric Difference-in-Differences Estimators”. *The Review of Economic Studies* 72(1), pp. 1–19.

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Mathias Dolls, and Pablo Zarate (2022). *Working from home around the world*. Working Paper 30446. National Bureau of Economic Research.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu (June 2017). “Reference-Dependent Preferences: Evidence from Marathon Runners”. *Management Science* 63(6), pp. 1657–1672.

Allen, Marcus T. and William H. Dare (Dec. 2004). “The Effects of Charm Listing Prices on House Transaction Prices”. *Real Estate Economics* 32(4), pp. 695–713.

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

— (Dec. 2014). *Mastering ’Metrics: The Path from Cause to Effect*. Princeton University Press.

Athey, Susan (Jan. 2018). “The Impact of Machine Learning on Economics”. *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, pp. 507–547.

Bahar, Dany, Prithwiraj Choudhury, Do Yoon Kim, and Wesley W Koo (2022). “Innovation on Wings: Nonstop Flights and Firm Innovation in the Global Context”. *Forthcoming in Management Science*.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun (Apr. 2020). “CEO Behavior and Firm Performance”. *Journal of Political Economy* 128(4), pp. 1325–1369.

Barrero, Jose Maria, Nicholas Bloom, and Steven J Davis (2021). *Why working from home will stick*. Working Paper 28731. National Bureau of Economic Research.

Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga (2016). "A practitioners' guide to gravity models of international migration". *The World Economy* 39(4), pp. 496–512.

Belenzon, Sharon and Mark Schankerman (Oct. 2008). *Motivation and Sorting in Open Source Software Innovation*. CEPR Discussion Papers 7012. C.E.P.R. Discussion Papers.

Bellmann, Lutz, Peter Ellguth, Philipp Grunau, Christian Hohendanner, Susanne Kohaut, Ute Leber, Iris Möller, Barbara Schwengler, Jens Stegmaier, Matthias Umkehrer, Viktoria Nußbeck, Sabine Oberg, Theresa Koch, and Hannes Walz (2019). *IAB-Betriebspanel (IAB BP) - Version 9319 v1*. de.

Blader, Steven, Claudine Gartenberg, and Andrea Prat (July 2019). "The Contingent Effect of Management Practices". *The Review of Economic Studies* 87(2), pp. 721–749.

Blei, David M and John D Lafferty (2009). "Topic models". *Text mining*. Chapman and Hall/CRC, pp. 101–124.

Blei, David M, Andrew Y. Ng, and Michael I. Jordan (Mar. 2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3, pp. 993–1022.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen (May 2019). "What Drives Differences in Management Practices?" *American Economic Review* 109(5), pp. 1648–1683.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (Nov. 2012a). "Does Management Matter? Evidence from India". *The Quarterly Journal of Economics* 128(1), pp. 1–51.

Bloom, Nicholas, Ruobing Han, and James Liang (2022). *How hybrid working from home works out*. Working Paper 30292. National Bureau of Economic Research.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenen (July 2014). "The new empirical economics of management". *Journal of the European Economic Association* 12(4), pp. 835–876.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts (Apr. 2020). "Do Management Interventions Last? Evidence from India". *American Economic Journal: Applied Economics* 12(2), pp. 198–219.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen (Jan. 2015). "The Impact of Competition on Management Quality: Evidence from Public Hospitals". *The Review of Economic Studies* 82(2), pp. 457–489.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen (Feb. 2012b). "Americans Do IT Better: US Multinationals and the Productivity Miracle". *American Economic Review* 102(1), pp. 167–201.

— (Nov. 2012c). "The Organization of Firms Across Countries". *The Quarterly Journal of Economics* 127(4), pp. 1663–1705.

— (June 2016). *Management as a Technology?* Working Paper 22327. National Bureau of Economic Research.

Bloom, Nicholas and John Van Reenen (2007). "Measuring and Explaining Management Practices Across Firms and Countries". *The Quarterly Journal of Economics* 122(4), pp. 1351–1408.

— (2011). "Chapter 19 - Human Resource Management and Productivity". Ed. by David Card and Orley Ashenfelter. Vol. 4. Handbook of Labor Economics. Elsevier, pp. 1697–1767.

Böckerman, Petri and Pekka Ilmakunnas (2012). "The Job Satisfaction-Productivity Nexus: A Study Using Matched Survey and Register Data". *Industrial and Labor Relations Review* 65(2), pp. 244–262.

Bojinov, Iavor, Prithwiraj Choudhury, and Jacqueline N Lane (2021). "Virtual watercoolers: A field experiment on virtual synchronous interactions and performance of organizational newcomers". *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (21-125).

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (2018). "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico". *Journal of Political Economy* 126(2), pp. 635–687.

Busse, Meghan R., Nicola Lacetera, Devin G. Pope, Jorge Silva-Risso, and Justin R. Sydnor (May 2013). "Estimating the Effect of Salience in Wholesale and Retail Car Markets". *American Economic Review* 103(3), pp. 575–79.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021). "Difference-in-Differences with multiple time periods". *Journal of Econometrics* 225(2). Themed Issue: Treatment Effect 1, pp. 200–230.

Card, David (Jan. 1990). "The Impact of the Mariel Boatlift on the Miami Labor Market". *Industrial and Labor Relations Review* 43(2), p. 245.

— (1992). "Do Minimum Wages Reduce Employment? A Case Study of California, 1987-89". *Industrial and Labor Relations Review* 46(1), pp. 38–54.

Castro, Silvia, Florian Englmaier, and Maria Guadalupe (June 2022). *Fostering Psychological Safety in Teams: Evidence from an RCT*. Working Paper.

Chava, Sudheer and Vincent W. Yao (2017). "Cognitive Reference Points, the Left-Digit Effect, and Clustering in Housing Markets". *SSRN Electronic Journal*.

Chetty, Raj, Adam Looney, and Kory Kroft (Aug. 2009). "Salience and Taxation: Theory and Evidence". *American Economic Review* 99(4), pp. 1145–1177.

Daw, Jamie R. and Laura A. Hatfield (2018). "Matching and Regression to the Mean in Difference-in-Differences Analysis". *Health Services Research* 53(6), pp. 4138–4156.

DeFilippis, Evan, Stephen Michael Impink, Madison Singell, Jeffrey T Polzer, and Raffaella Sadun (2020). *Collaborating during coronavirus: The impact of COVID-19 on the nature of work*. Working Paper 27612. National Bureau of Economic Research.

DellaVigna, Stefano (May 2009). "Psychology and Economics: Evidence from the Field". *Journal of Economic Literature* 47(2), pp. 315–372.

Emanuel, Natalia, Emma Harrington, and Amanda Pallais (2022). "The Power of Proximity: Office Interactions Affect Online Feedback and Quits, Especially for Women and Young Workers". Working Paper.

Englmaier, Florian, E. Galdon-Sanchez Jose, Ricard Gil, Michael Kaiser, and Helene Strandt (2022). "Management practices and firm performance during the Great Recession - Evidence from Spanish survey data". Working Paper.

Englmaier, Florian, Michael Hofmann, and Stefanie Wolter (2023). "Win-Win: Management styles, firm performances and employee satisfaction". Mimeo.

Englmaier, Florian, Arno Schmöller, and Till Stowasser (2018a). "Does Bidding for Complex Goods Reflect All Relevant Information? Field Evidence from Online Gaming". Working paper.

— (2018b). "Price Discontinuities in an Online Market for Used Cars". *Management Science* 64(6), pp. 2754–2766.

Gelman, Andrew and Guido Imbens (May 2018). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs". *Journal of Business & Economic Statistics* 37(3), pp. 447–456.

Gibbons, Robert and John Roberts (Jan. 1, 2013). *Handbook of Organizational Economics*. Princeton Univers. Press.

Gibbs, Michael, Friederike Mengel, and Christoph Siemroth (2021). "Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals". *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2021-56).

Gilbert, Thomas, Shimon Kogan, Lars Lochstoer, and Ataman Ozyildirim (2012). "Investor Inattention and the Market Impact of Summary Statistics". *Management Science* 58(2), pp. 336–350.

Gousios, Georgios (2013). "The GHTorrent dataset and tool suite". *Proceedings of the 10th Working Conference on Mining Software Repositories*. MSR '13. San Francisco, CA, USA: IEEE Press, pp. 233–236.

Halkos, George and Dimitrios Bousinakis (Jan. 2010). "The effect of stress and satisfaction on productivity". *International Journal of Productivity and Performance Management* 59(5), pp. 415–431.

Hansen, Ben B. and Stephanie Olsen Klopfer (2006). "Optimal Full Matching and Related Designs via Network Flows". *Journal of Computational and Graphical Statistics* 15(3), pp. 609–627.

Head, Keith and Thierry Mayer (2014). "Gravity equations: Workhorse, toolkit, and cookbook". *Handbook of International Economics*. Vol. 4. Elsevier, pp. 131–195.

Hergueux, Jérôme and Nicolas Jacquemet (June 2015). “Social preferences in the online laboratory: a randomized experiment”. *Experimental Economics* 18(2), pp. 251–283.

Holland, Paul W. (Dec. 1986). “Statistics and Causal Inference”. *Journal of the American Statistical Association* 81(396), pp. 945–960.

Hu, Albert G. Z. and Adam B. Jaffe (June 2003). “Patent citations and international knowledge flow: the cases of Korea and Taiwan”. *International Journal of Industrial Organization* 21(6), pp. 849–880.

Ichniowski, Casey and Kathryn Shaw (Mar. 2003). “Beyond Incentive Pay: Insiders’ Estimates of the Value of Complementary Human Resource Management Practices”. *Journal of Economic Perspectives* 17(1), pp. 155–180.

Imbens, Guido W. and Joshua D. Angrist (1994). “Identification and Estimation of Local Average Treatment Effects”. *Econometrica* 62(2), pp. 467–475.

Imbens, Guido W. and Donald B. Rubin (1997). “Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance”. *The Annals of Statistics* 25(1), pp. 305–327.

Jaffe, Adam B and Manuel Trajtenberg (1999). “International knowledge flows: evidence from patent citations”. *Economics of Innovation and New Technology* 8(1-2), pp. 105–136.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (Aug. 1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations”. *The Quarterly Journal of Economics* 108(3), pp. 577–98.

Kampkötter, Patrick, Jens Mohrenweiser, Dirk Sliwka, Susanne Steffes, and Stefanie Wolter (Aug. 2016). “Measuring the use of human resources practices and employee attitudes: The Linked Personnel Panel”. *Evidence-based HRM: a Global Forum for Empirical Scholarship* 4, pp. 94–115.

Koch, Tobias, Dr. Oliver Ehrentraut, Marion Neumann, and Ante Pivac (2017). *Wohnraumbedarf in Deutschland und den regionalen Wohnungsmärkten - Endbericht*. German. Prognos AG.

Lacetera, Nicola, Devin G Pope, and Justin R Sydnor (Aug. 2012). “Heuristic Thinking and Limited Attention in the Car Market”. *American Economic Review* 102(5), pp. 2206–2236.

LaLonde, Robert J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". *The American Economic Review* 76(4), pp. 604–620.

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

Lerner, Josh and Jean Tirole (May 2001). "The open source movement: Key research questions". *European Economic Review* 45(4-6), pp. 819–826.

— (Spring 2005). "The Economics of Technology Sharing: Open Source and Beyond". *Journal of Economic Perspectives* 19(2), pp. 99–120.

Lu, Xuan, Wei Ai, Yixin Wang, and Qiaozhu Mei (2023). "Team Resilience under Shock: An Empirical Analysis of GitHub Repositories during Early COVID-19 Pandemic". *arXiv preprint arXiv:2301.12326*.

Luca, Michael (2011). "Reviews, Reputation, and Revenue: The Case of Yelp.Com". *Harvard Business School Working Paper, No. 12-016*.

Lynn, Michael, Sean Masaki Flynn, and Chelsea Helion (June 2013). "Do consumers prefer round prices? Evidence from pay-what-you-want decisions and self-pumped gasoline purchases". *Journal of Economic Psychology* 36, pp. 96–102.

Mackeben, Jan, Kevin Ruf, Sefanie Wolter, Philipp Grunau, and Tobias Haepp (2020a). *Linked Personnel Panel (LPP) - Version 1819 v1*. en. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB).

Mackeben, Jan, Kevin Ruf, Stefanie Wolter, Philipp Grunau, and Tobias Haepp (2020b). "LPP - Linked Personnel Panel 1819: Quality of work and economic success. Longitudinal study in German establishments (data documentation on the fourth wave). (FDZ-Datenreport 11/2020 (en))".

Mayda, Anna Maria (2010). "International migration: A panel data analysis of the determinants of bilateral flows". *Journal of Population Economics* 23(4), pp. 1249–1274.

McDermott, Grant R and Benjamin Hansen (2021). *Labor Reallocation and Remote Work During COVID-19: Real-time Evidence from GitHub*. Working Paper 29598. National Bureau of Economic Research.

McDonald, Nora and Sean Goggins (2013). "Performance and Participation in Open Source Software on GitHub". *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. Paris, France: ACM, pp. 139–144.

Meng, Charlotte C. (2019). "Every Digit Counts: Heuristic Thinking and Biased Reference Points". Working paper.

Migali, Silvia, Fabrizio Natale, Guido Tintori, Sona Kalantaryan, Sara Grubanov-Boskovic, Marco Scipioni, Fabio Farinosi, Cristina Cattaneo, Barbara Benandi, Marco Follador, et al. (2018). *International Migration Drivers*.

Palmer, Karen and Margaret Walls (May 2015). "Limited Attention and the Residential Energy Efficiency Gap". *American Economic Review* 105(5), pp. 192–95.

Pope, Devin G. (Dec. 2009). "Reacting to rankings: Evidence from "America's Best Hospitals"". *Journal of Health Economics* 28(6), pp. 1154–1165.

Pope, Devin G., Jaren C. Pope, and Justin R. Sydnor (Sept. 2015). "Focal points and bargaining in housing markets". *Games and Economic Behavior* 93, pp. 89–107.

Ramos, Raul (2017). "Modelling migration". *The Econometrics of Multi-Dimensional Panels*. Springer, pp. 377–395.

Repetto, Luca and Alex Solís (Nov. 2019). "The Price of Inattention: Evidence from the Swedish Housing Market". *Journal of the European Economic Association*.

Rubin, Donald B. (Oct. 1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66(5), pp. 688–701.

Simon, Herbert A. (Feb. 1955). "A Behavioral Model of Rational Choice". *The Quarterly Journal of Economics* 69(1), pp. 99–118.

Sirmans, G. Stacy, David A. Macpherson, and Emily N. Zietz (2005). "The Composition of Hedonic Pricing Models". *Journal of Real Estate Literature* 13(1), pp. 3–43.

Sonnemans, Joep (Nov. 2006). "Price clustering and natural resistance points in the Dutch stock market: A natural experiment". *European Economic Review* 50(8), pp. 1937–1950.

Sopranzetti, Ben J and Vinay Datar (Oct. 2002). "Price clustering in foreign exchange spot markets". *Journal of Financial Markets* 5(4), pp. 411–417.

Splawa-Neyman, Jerzy, D. M. Dabrowska, and T. P. Speed (1923/1990). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9". *Statistical Science* 5(4), pp. 465–472.

Thompson, Peter and Melanie Fox-Kean (Mar. 2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment". *American Economic Review* 95(1), pp. 450–460.

Tinbergen, Jan (1962). "Shaping the world economy; suggestions for an international economic policy".

Yang, Longqi, David Holtz, Sonia Jaffe, Siddharth Suri, Shilpi Sinha, Jeffrey Weston, Connor Joyce, Neha Shah, Kevin Sherman, Brent Hecht, et al. (2022). "The effects of remote work on collaboration among information workers". *Nature human behaviour* 6(1), pp. 43–54.