ESSAYS ON REAL ESTATE DURING CRISES

**Dissertation zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaft**

**eingereicht an der Fakultät für Wirtschaftswissenschaften
der Universität Regensburg**

vorgelegt von: Rupert Klaas Eisfeld

Berichterstatter: Prof. Dr. Tobias Just
Prof. Dr. Kristof Dascher

Tag der Disputation: 16. Juni 2026

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and background

In the aftermath of the global financial crisis of 2007–2008, Germany entered a decade of relative macroeconomic stability, urban prosperity and economic growth (Bundesbank, 2020). However, many of the mechanisms that underpinned growth—expansive monetary policy, increasingly complex global value chains, steady capital inflows into real estate markets, and a persistent belief in Germany's structural resilience—functioned not only as stabilising forces but also as latent risk factors. In hindsight, the long upswing of the 2010s bore its own memento mori: it rested on conditions that were historically exceptional and whose transience was often ignored in public discourse, investor behaviour, and policy design.

The abrupt end of this stability at the beginning of the 2020s exposed the degree to which crises had been conceptually marginalised. The COVID-19 pandemic, emerging in early 2020, exerted a pervasive influence on human and societal life as well as on economic activity throughout Germany in 2020 and 2021. In early 2022, the Russian Federation launched a full-scale invasion of its neighbouring state, Ukraine, precipitating extensive destruction across Ukrainian territory, triggering an unprecedented wave of forced displacement, and generating a global inflationary shock. What is striking about these events is not only their impact but the extent to which they defied dominant expectations in Europe. The pandemic was the most severe public-health crisis in a century, and the invasion produced geopolitical and economic disruptions that many market participants and policy makers had long regarded as inconceivable. For the first time in decades, Western Europe was confronted by events of crisis of a scale that had largely vanished from institutional memory—crises that had ceased to be part of routine modelling, market expectations, or policy preparation.

This neglect is particularly evident in real estate markets. Among both institutional and private investors, real estate is frequently understood as a safe asset—characterised by low volatility, stable income streams, and intrinsic value tied to land and location (Jordá et al. 2019). The prolonged upswing of the 2010s reinforced this perception, attracting record levels of capital and encouraging a form of crisis-blindness in investment strategies and risk modelling. Yet crises are not anomalies; they are recurrent and, as economic history (Minsky, 1986; Carmen & Rogoff, 2009) and pandemic history (Glaeser 2020) repeatedly demonstrate, unavoidable. Risk models that rely on smooth trajectories, trend extrapolation, or assumptions of continuing stability are thus prone to severe misjudgement—especially in an asset class defined by long holding periods, illiquidity, and fixed locations.

The events of the early 2020s therefore offer a double analytical challenge. They reveal structural vulnerabilities that had accumulated unnoticed during a decade of prosperity, and they expose the conceptual limitations of prevailing frameworks for understanding real estate risks, urban dynamics, and systemic shocks. At the same time, they provide an opportunity: the return of crises to the centre of urban and economic life invites a re-examination of the assumptions that have guided investment, planning, and policy making in recent years. This thesis situates itself within this moment of conceptual and empirical rupture. The repercussions of these two crises—the COVID-19 pandemic and the Ukraine shock—on urban systems and the real estate sector constitute the focus of this thesis.

Historically, cities have been more frequently and more severely affected by the impacts of epidemics and pandemics (Hooper, 2020). Population density is the "defining feature" (Carozzi et al., 2022) of cities (Glaeser, 2011). In the case of respiratory virus, it is reasonable to assume that their spread is facilitated when large numbers of people live in close proximity (Tarwater & Martin, 2001; Rocklöv & Sjödin, 2020). In crowded situations —such as those more frequently encountered in urban environments—pathogens transmitted through the air can spread more easily and thus more rapidly (Gerritse, 2020). In general, urban areas are more densely populated, crowded situations involving close physical contact occur more frequently, and the likelihood that viruses are introduced from outside the city is higher than in peripheral regions, as cities also function as major transportation hubs. Thus, since the outset of the pandemic in late 2019 / early 2020 researchers and policy makers across the globe have debated anxiously about the vulnerability of urban populations (McFarlane, 2021).

Therefore, in Chapter 2, we analyse the trajectory of the COVID-19 pandemic in Germany in relation to population density up until one year after the start of the vaccination campaign. Based on county-level data, we first show that density is associated with higher COVID-19 mortality when we control for individual health factors. In this regard, concerns of cities being infection hotspots during pandemics are confirmed. However, this density effect is relatively small compared to health-related factors (air pollution, diabetes, and vaccination rates) and is only about one third of a location effect, measured in terms of longitude and latitude. This location effect captures the impact of regional virus dispersion across space. Even two years after the start of the pandemic, regions in the south of Germany in proximity to Ischgl—a ski resort in the Alps, where après ski parties became superspreader events (Felbermayr et al., 2021)—have to bear a larger cumulative death burden compared to more northern regions. This result confirms that regional dispersion poses a greater challenge for pandemic response regimes in Germany than density is. By tracing the development of influential factors on the COVID-19 mortality, we find that the higher COVID-19 mortality in dense counties as well as southern counties is due to a head start for those counties. Once resilient measures were implemented to protect the population, no density disadvantage of cities can be demonstrated. The mortality risk converges with the personal risk factors. This highlights the importance of early policy respondence onto a pandemic threat and the long-lasting effect of the localization of the first "pandemic seed" in modern epidemiological scenarios.

For real estate investors there are important messages implied with these results: first, conurbations are not posing bigger investment risks than peripheral areas during pandemics, once the necessity of a response is understood. Second, real estate investors thus need to carefully monitor this pandemic management system within a country, as cities are more vulnerable, if responses are too late or too weak. A first reflex among urban planners when confronted with a pandemic threat might be, for understandable reasons, to attempt to disperse the population in order to proactively prevent infections. Our results indicate that such paradigm for urban development is not necessary when effective public crisis management can be expected. Similarly, the scientific community initially discussed that residential location preferences and workplace design had been permanently altered as a consequence of new hygienic priorities (Wolday & Böcker, 2023). In fact, a slight shift in migration patterns from urban areas towards more suburban regions can be observed. However, the reasons now discussed for this development relate exclusively to an increased prevalence of working from home and rising housing costs in metropolitan locations, rather than to infection risk.[3] We conclude from our research that, both in the future and in the event of renewed pandemics, the advantages of urban and, more generally, high-density settlement systems will continue to outweigh their disadvantages.

Given that crises can be expected to occur periodically, and that real estate assets are characterised by long investment horizons, the question arises as to how companies in the real estate sector should manage the risk of major shocks. Ideally, a company's strategy is oriented towards the long-term generation of profits. This long-term orientation on how to receive returns can also be described as the company's business model. The question of whether companies can shield themselves from the consequences of major crises through the design of their business

[3] See, for example, Siedentop (2024) for a review of suburbanisation and its drivers in Germany in the 2020s.

models is central to understanding resilience in real estate capital markets. To address this question, in Chapter 3 we develop a novel framework for distinguishing and quantifying the business models of real estate stock companies in the DACH region (Germany, Austria and Switzerland). The pricing of real estate stocks has attracted considerable attention, particularly during periods of market turbulence. While previous research has explored price fluctuations according to region, investment type, or geographic focus, the influence of underlying business models in the real estate sector on capital market behaviour has not been investigated to date. A deeper understanding of these business models is therefore essential for interpreting stock market reactions during crises and for establishing meaningful benchmarks.

Most research in the real estate stock universe focuses on real estate investment trusts (REITs), whose revenues are primarily derived from rental income generated by owned properties. In contrast, REITs are rare in the DACH region. The majority of real estate stock companies in German-speaking countries are property companies engaging in diverse business activities and combinations thereof. We use real estate stock companies´ revenues along the two dimensions "sector" (asset class) and "position along the value chain" to distinguish between business models and to quantify revenue fragmentation. How the fragmentation of each dimension affects the capital market volatility is relevant for our understanding of how diversification is valued by capital markets. Additionally, different business models and associated stocks might respond differently to varying market phases and external shocks. An analysis of these reactions is therefore a research objective with interesting insights in terms of resiliency during crises such as the COVID-19 pandemic and the war in Ukraine.

Our analysis is based on a sample of 48 real estate stock companies from Germany and Austria from Q1 2016 up until Q4 2022, with a total market capitalization of €89 billion. The revenue data was collected by business report analysis and a survey conducted among the companies in 2022. We use hierarchical clustering after Ward Jr. (1963) to derive compact clusters of real estate stock companies sharing a comparable revenue structure, which can be interpreted as groups of business models. Based on these results we construct a stock market index for each business model. These indices are subjected to a changepoint analysis based on the Pruned Exact Linear Time (PELT) method to identify breaks in the volatility of the return series (Killick et al., 2012). From there, the individual time frames of each index based on the identified breaks are subjected to analysis and specified for the pandemic period. In addition to a better understanding of the complexity and fragmentation of the business models pursued by real estate stock companies and their impact on capital market volatility, this paper provides a basis for a systematic benchmarking approach for investors and regulators based on the identified groups of business models, as well as a better understanding of the resiliency of different business models during different market phases.

However, it is not only companies that are exposed to major shocks. Most frequently, it is individuals who suffer the greatest impact from significant crises. One historical response to better protect against future uncertainties was the establishment of states. Effective institutions enable coordinated management, allowing burdens and responsibilities to be distributed across multiple actors. At the same time, effective institutions protect individuals from excessive exploitation and arbitrariness, which becomes particularly important during periods of major crisis (Acemoglu & Robinson, 2012). This logic can also be applied to real estate markets, which are characterised by a high degree of regulation. Ideally, such regulations facilitate a balance of interests while considering the power relations among the actors involved. Crucially, regulations must be clearly defined and tested before major shocks occur. Only then can they provide sufficient and effective protection during turbulent periods. Conversely, the German experience with rent regulation, particularly regarding the adjustment of *Mietspiegel*, exemplifies how crises expose uncertainties in key regulatory instruments.

A *Mietspiegel* is a register providing an overview of the rents customarily paid for comparable housing within a municipality (§558c of the German Civil Code, BGB), and it serves as the basis for determining the so-called local comparative rent (*ortsübliche Vergleichsmiete*) for a specific dwelling. Only rents that have been newly agreed or changed in the last six years and that are not subject to any special regulations on rent levels may be considered when compiling the *Mietspiegel* (§558(2) BGB). In contrast to simple *Mietspiegel*, qualified *Mietspiegel* are compiled according to recognised scientific principles and therefore possess significantly greater evidentiary value in legal disputes (§558d BGB). The determination of local comparative rents is of fundamental importance for rental practices in Germany, as it also constitutes a key reference point for several regulatory instruments. For instance, under §558(1) BGB, rents in ongoing tenancy agreements may generally be increased only up to the level of the local comparative rent. Moreover, in housing markets classified as “tight” (*angespannte Wohnungsmärkte*) pursuant to §556d BGB, newly concluded rental contracts for buildings constructed before October 2014 (§ 556f BGB) may not exceed the local comparative rent by more than ten percent. According to §§ 22 and 23 of the German Rent Index Ordinance (*Mietspiegelverordnung*, MsV), introduced in 2021, municipalities have two options for adjusting a qualified *Mietspiegel* to market developments after two years: either (1) by applying the change in the Consumer Price Index (CPI) to the *Mietspiegel* to be updated, or (2) by calculating an adjustment factor based on survey data.

In January 2022, shortly before the outbreak of the war in Ukraine, which triggered pronounced inflationary developments in Germany, almost one third of all *Mietspiegel* adjustments relied on sample surveys, while two thirds used the Consumer Price Index (CPI; Sebastian & Range, 2022). High inflation in 2022 and 2023 increased attention to sample-based adjustments, as doubts arose over whether the consumer price index accurately reflects rental developments in each municipality. At the same time, municipalities aim to prevent tenants from facing a double burden caused by simultaneous rent increases and rising living costs. These circumstances raise several unresolved statistical questions regarding sample-based adjustments, for which the MsV offers no guidance.

Although § 23(2) MsV permits a reduced sample size for primary survey data collections, updating via the CPI method is far less costly for municipalities. However, it is unsuitable for adequately reflecting the actual development of local comparative rents. In light of the inflationary shock in 2022 and 2023, during which the CPI increased by more than ten per cent, municipalities increasingly turned to sample-based updates. The methodological approach to such updates had, until recently, remained unclear. Chapter 4 addresses that gap by providing guidance on the advantages and disadvantages of different methodological approaches. We derive, among other things, a formula for determining the reduced sample size required for sample-based updates of qualified *Mietspiegel*. The newly derived formula for planning sample size enables *Mietspiegel* compilers to determine, in a statistically robust and cost-efficient manner, the required number of observations, given predefined expectations regarding the precision of estimates of the local comparative rent. This, in turn, enhances legal certainty for the adjustments made to qualified *Mietspiegel*, as the precision of the sample estimates can be demonstrated on a sound methodological basis.

In summary, this thesis addresses three interconnected dimensions of risk, resilience, and governance relevant to the real estate sector that gained importance during recent years of crises. The first concerns the existential risk arising from crowding in cities and operating real estate under modern pandemic scenarios. We found that this risk can be managed effectively if responses are timely and decisive. However, there remains scope for improvement in international coordination of policy responses. The second aspect relates to the robustness of the stock prices of real estate stock companies when exposed to shocks of various kinds. Our findings demonstrate that stock prices display heterogeneous responses, which may be attributable to differences in business models. This approach could, in future, contribute to the development of more precise investment strategies. The third and final aspect concerns the sample-based adjustment of qualified *Mietspiegel*, which has played an increasingly important role in residential real estate practice during periods of high inflation. We provide guidance for empirical surveys under the simplified requirements of the MsV.

Taken together, these three strands highlight a common theme: the capacity of real estate markets, actors, and institutions to absorb, adapt to, and recover from shocks is critically dependent on structured, data-informed decision-making within clearly defined institutional frameworks. Across scales—from urban systems and companies to individuals and households—the findings underscore that both systemic design and methodological rigor are crucial for safeguarding assets, ensuring stability, and enhancing adaptive resilience in an environment where crises are recurrent rather than exceptional. Ultimately, this thesis contributes to a more integrated understanding of how physical, financial, and regulatory dimensions of the real estate sector interact during crises.

1.2 Research problems and questions

This section introduces the research problems addressed in each paper and compiles the research questions that are examined across the three contributions of this thesis.

Paper 1: Urban Density versus Regional Dispersion: On the Risks in High-Density Conurbations in Germany during COVID-19

It seems reasonable to assume that urban populations are more vulnerable than those in less densely populated regions during the COVID-19 pandemic and consequently suffer a relatively higher death toll. We measure vulnerability in terms of mortality, even if this indicator does not capture vulnerability entirely, since an infection with the coronavirus can be harmful even if it does not lead to death, either due to long-term health complaints after a (mild) infection or because of productivity declines due to increased sickness rates. However, as it is unlikely that all infections are detected during high incidence periods COVID-19-related deaths can be more precisely counted. Additionally, deaths represent the highest toll a population can pay, so we only analyse COVID-19 death numbers. To distinguish between more urban and less urban counties (Stadt- und Landkreise), we employ population size relative to the respective settlement and transport area as an indicator of settlement density—or urban density, respectively. We therefore begin by testing (1) whether there is a positive correlation between settlement density and mortality during the pandemic.

However, settlement density is not the only factor determining the vulnerability of local populations in modern pandemic scenarios. The virus initially entered Germany primarily from Southern Europe. Superspreader events at the ski resort of Ischgl shaped infection dynamics during the early phase of the pandemic (Felbermayr et al., 2021), resulting in higher infection and mortality rates in southern Germany at an earlier stage. The relative geographical position within Germany in relation to infection hotspots may therefore also constitute a factor influencing the vulnerability of local populations. In addition, regional differences in pandemic control measures contributed to varying infection rates, and ultimately, the overall health status of the population plays a decisive role in determining the likelihood that an infection with the coronavirus will result in death. It remains unclear which of these factors exerted the strongest influence on the vulnerability of local populations. We therefore proceed to examine the second research question in a multivariate regression analysis based on standardized data: (2) How is the magnitude of this potential density effect on mortality relative to a broad range of covariates, notably geographic location within Germany?

The most effective protection against death from a viral infection is likely an effective vaccination. At the outset of the COVID-19 pandemic, however, no such vaccine was available; it only became gradually accessible from spring 2021 onwards. Individuals who were exposed to the novel coronavirus early therefore faced a higher risk of dying from the infection than those who had the opportunity to be vaccinated. At the same time, it can be assumed that the virus reached urban areas earlier than more remote regions. This raises the question of whether the observed relationships between settlement density, mortality, and other covariates are indeed stable. Finally, focusing on urban areas, we ask: (3) Do these potential correlations between density and mortality vary over time and space? We investigate this research questions within a multivariate regression analysis at different points in time and Generalized Additive Models (Wood, 2017) with a smoothed interaction term of longitude and latitude.

Paper 2: Activity-based Clustering of Real Estate Stock Companies and the Varying Adjustment to Shocks

For a meaningful investigation of differences in stock market reactions to shocks across business activities, we begin by grouping real estate stock companies according to similarities in their revenue structures. Disaggregated company data on revenues by business activity is not readily available for most real estate stock companies in German-speaking countries. Therefore, we hand-collected data from industry experts, business reports, and a survey conducted among real estate companies. The first research question in relation to this is: (1) Can real estate stock companies be meaningfully segmented into clusters based on two pivotal activity dimensions representative of a business model? These dimensions are the specific sector (asset class) a company focuses on and its position along the value chain. Since data on revenues in different real estate segments or sectors are not readily provided, we conduct a survey among real estate companies in the German-speaking region and use standard clustering procedures to find groups with similar revenue structures or, respectively, business models.

Having identified potential clusters, it is important to examine whether these groupings correspond to observable differences in market behaviour of the underlying stocks. Distinct business models are likely to carry differing levels of risk, which may be reflected in the volatility of their stock prices. Analysing these patterns can provide valuable insight into how investors perceive and value diversification within the real estate sector. This leads to the research question: (2) Do the isolated clusters exhibit discernible differences in their stock price volatilities?

Finally, external shocks, such as the COVID-19 pandemic and the recent reversal in interest rates, offer an opportunity to assess the resilience of different business models. Understanding how clusters respond to such shocks can reveal both vulnerabilities and strengths, providing guidance for investment decisions and regulatory strategies in turbulent market conditions. This consideration underpins the next research question: (3) How did these clusters react in terms of volatility and stock price performance during the two most recent shocks—namely, the COVID-19 pandemic and the interest rate reversal?

Paper 3: Anpassung qualifizierter Mietspiegel: methodische Hinweise zur statistischen Absicherung stichprobenbasierter Verfahren

One approach to sample-based adjustments involves using longitudinal data for adjustment surveys rather than comparing two separate cross-sectional surveys. Analysts could obtain or estimate rents for newly moved-in households as of the original survey's reference date, allowing them to track changes over time. This consideration raises the first research question: (1) Should cross-sectional or longitudinal data be used for sample-based *Mietspiegel* adjustments? Additionally, municipalities might re-interview the (gross) sample from the original survey to avoid the costs of drawing a new sample. This strategy motivates the second research question: (2) Can the gross sample from the most recent full survey be reused?

Rental developments are not always uniform, and periods of limited upward or downward movement may occur. Consequently, analysts cannot assume that *Mietspiegel* rents change with every adjustment, which leads to the third research question: (3) Should adjustments be made at all, and which changes in rent levels can be observed actually? If the MsV allows for a reduced sample under §23(2), we must also determine: (4) How should the sample size for the adjustment be defined?

The MsV further permits simplifying assumptions to be made during adjustments (§23(1)). Submarket-specific adjustments may better capture rental dynamics if increases predominantly affect newly constructed housing. This scenario raises the fifth research question: (5) Should adjustments apply uniformly or vary according to submarkets?

Finally, analysts may calculate the local comparative rent using either the table method (§13 MsV) or the regression method (§14 MsV). The regression method relates rents to housing-relevant characteristics, and analysts must ensure that the regression equation remains statistically robust and logically consistent. For adjustments, they can either apply the established regression equation to the reduced sample—recalculating *Mietspiegel* rents while preserving the dependency structure—or construct an index based on the mean values of both samples. This consideration gives rise to the sixth research question: (6) Should adjustments rely on a sample-based index or an updated regression equation?

1.3 Co-authors, submissions and conference presentations

This section provides an overview about co-authors, journal submissions, publication status and conference presentations.

<u>Paper 1: Urban Density versus Regional Dispersion: On the Risks in High-Density Conurbations in Germany during COVID-19</u>

Authors:

- Rupert K. Eisfeld, Tobias Just

Submission Details:

- Journal: Sustainable Cities and Society
- Current Status: published (05/2024)

Conference Presentations:

- -

<u>Paper 2: Activity-based Clustering of Real Estate Stock Companies and the Varying Adjustment to Shocks</u>

Authors:

- Karl-Friedrich Keunecke, Rupert K. Eisfeld and Tobias Just

Submission Details:

- Journal: Journal of Urban Management
- Current Status: under review (04/2025)

Conference Presentations:

- 39th Annual ARES (American Real Estate Society) Meeting, 28th March to 1st April 2023, San Antonio, Texas, United States
- 4th International Conference on Real Estate Development and Management (ICREDM 2025), 3rd to 5th February 2025, Ankara, Türkiye

<u>Paper 3: Anpassung qualifizierter Mietspiegel: methodische Hinweise zur statistischen Absicherung stichprobenbasierter Verfahren</u>

Authors:

- Holger Cischinsky, Max-Christopher Krapp, Steffen Sebastian and Rupert K. Eisfeld

Submission Details:

- Journal: ASta Wirtschafts- und Sozialstatistisches Archiv
- Current Status: under review (10/2025)

Conference Presentations:

- FuB iGES Mietspiegeltagung, 28th April 2025, Berlin, Germany

References

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity and poverty*. London: Profile.

Bundesbank (2020). Monatsbericht Oktober 2020. Jahrgang 72, Nr. 10, Frankfurt am Main.

Carozzi, F., Provenzano, S., & Roth, S. (2022a). Urban density and COVID-19: understanding the US experience. *The Annals of Regional Science*, 1-32.

Felbermayr, G., Julian H. & Chowdhry, S. (2021). Après-ski: The spread of coronavirus from Ischgl through Germany. In: *German Economic Review* 22.4, S. 415–446.

Gerritse, M. (2022). COVID-19 transmission in cities. *European Economic Review*, *150*, 104283.

Glaeser, E. (2011). Triumph of the city: How urban spaces make us human. Pan Macmillan.

Glaeser, E.: Cities and pandemics have a long history. In *City Journal,* Spring 2020, [online] https://www.city-journal.org/cities-and-pandemics-have-long-history [21 January 2026]

Hooper, M.: Pandemics and the future of urban density: Michael Hooper on hygiene, public perception and the "urban penalty." In *Harvard Graduate School of Design Newsletter*, 13 April 2020, [online] https://www.gsd.harvard.edu/2020/04/have-we-embraced-urban-density-to-our-own-peril-michael-hooper-on-hygiene-public-perception-and-the-urban-penalty-in-a-global-pandemic/ [21 January 2026]

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., & Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *The quarterly journal of economics*, *134*(3), 1225-1298.

Killick, Rebecca, Paul Fearnhead & Idris A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. In: *Journal of the American Statistical Association* 107.500, S. 1590–1598.

McFarlane, C. (2021). Repopulating density: COVID-19 and the politics of urban value. In: *Urban Studies*, S. 004209802110148.

Minsky, H. P. (1986). Stabilizing an Unstable Economy. *Hyman P. Minsky Archive*. 144. [online] https://digitalcommons.bard.edu/hm_archive/144 [10 December 2025]

Rocklöv, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine, 27*(3), taaa038.

Sebastian, S., & Range, J. (2022): gif-Mietspiegelreport 2022 – Auswertung der Mietspiegel der zweihundert größten Städte Deutschlands. Wiesbaden: gif-Mietspiegelkommission (Hrsg.).

Siedentop, S. (2024). Die" fünfte Suburbanisierung": Perspektiven suburbaner Raumentwicklung in den 2020er-Jahren. *Raumforschung und Raumordnung/Spatial Research and Planning*, *82*(5), 366-383.

Tarwater, P. M., & Martin, C. F. (2001). Effects of population density on the spread of disease. *Complexity, 6*(6), 29-36.

Ward Jr., Joe H. (1963). Hierarchical grouping to optimize an objective function. In: *Journal of the American statistical association* 58.301, S. 236–244.

Wolday, F., & Böcker, L. (2023). Exploring changes in residential preference during COVID-19: Implications to contemporary urban planning. *Environment and Planning B: Urban Analytics and City Science*, *50*(5), 1280-1297.

Wood, S. N. (2017). *Generalized additive models: an introduction with R.* Second edition. CRC press.

2 Urban Density versus Regional Dispersion: On the Risks in High-Density Conurbations in Germany during COVID-19

Abstract Significant research has been devoted to understanding how environmental factors affect the COVID-19 risks local populations face. Urban density has been identified as a critical factor, posing a particular threat in densely populated countries like Germany, one of Europe's most crowded nations. Furthermore, the high degree of interconnectivity between German regions increases the risk of virus transmission from infection hotspots through regional dispersion. Our research uses linear and non-linear regression models, enhanced with longitude and latitude, to assess the relative impact of population density and regional dispersion of COVID-19 mortality across German counties. Our findings, adjusted for health-related variables, indicate that locational factors have a three times greater impact on mortality than density does. This highlights that pandemic mortality risk is shaped not only by our individual responses but also by the actions of our neighbours. Consequently, our research suggests that, given the increasing global interdependence, regional sprawl cannot be the dominant strategy for planners to avoid the consequences of a pandemic. Instead, the priority should be efficient information management to identify and mitigate interdependence with infection hotspots, which requires enhanced international cooperation.

Keywords: Urban density; COVID-19; mortality; pandemic policy response; Germany

2.1 Introduction

Respiratory viruses such as SARS-CoV-2 spread more readily in densely populated areas compared to less crowded ones (Tarwater & Martin, 2001; Rocklöv & Sjödin, 2020). This has raised concerns about the vulnerability of urban populations since the coronavirus outbreak (McFarlane, 2021). It was feared that COVID-19 would echo historical patterns where cities often became infection epicentres (see Neiderud, 2015; Glaeser, 2020). However, this paper argues that the critical determinant of pandemic impact on an inter-county scale is geographical proximity to infection hotspots, influenced by the timing of non-pharmaceutical interventions (NPIs) and medical measures, especially vaccinations. NPIs like mask-wearing and social distancing have reduced the risks associated with high population densities (Sun et al., 2020; Khavarian-Garmsir et al., 2021). Nonetheless, without NPIs or in cases of poor compliance, urban areas remain at risk of widespread virus transmission (Smith et al., 2021; Zaki et al., 2022).

While density is often viewed as an *"urban penalty"* during pandemics (Hooper, 2020), this assumption is challenged by observations that some densely populated regions reported fewer COVID-19 cases compared to more sparsely inhabited areas (Moosa & Khatatbeh, 2021). This *"density paradox"* is evident in German counties.

By the end of February 2022, Germany recorded 127,539 COVID-19 deaths, equating to 153 per 100,000 inhabitants.[1] The regional variations in these deaths, and their association with settlement density, are presented in Figure 2.1. Specifically, Panel 1 maps the COVID-19 mortality across Germany. Panel 2 explores the relationship between COVID-19 mortality and settlement density (displayed on a logarithmic scale) for 394 out of 400 urban and rural counties. This correlation is statistically significant and negative at the five percent level.[2] However, this correlation can be misleading as density is only one of many factors influencing mortality during the pandemic. For example, despite having comparable settlement densities, Rendsburg-Eckernförde and Saale-Orla-County exhibit dramatically different mortality rates – 31.74 and 621.87 deaths, respectively.[4] It is also important to note the wide variation in settlement density

---

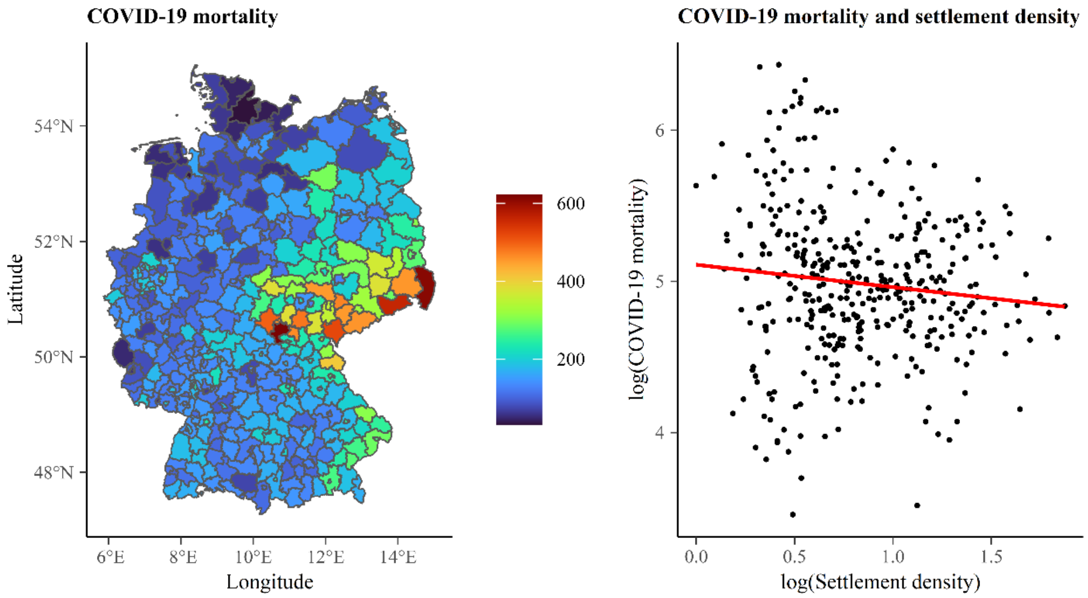
[1] We ended the observation period in February 2022 because our control variables had data limitations. We have found no hint that the geographical distribution of COVID-19 mortalities changed after the end of the observation period.

[2] Using robust standard errors; see Table 2.3, Column (1).

[4] We define the settlement density as a county´s population as of 31 December 2019 per square kilometre of settlement and traffic area. This is a more refined measure of population density since we argue that face-to-face

among German counties, ranging from 612 in Spree-Neisse county to 6439 in Munich, that is, an increase by a factor of ten.

Figure 2.1: COVID-19 mortality in German counties as of February 2022 and the relationship with settlement density.



Notes: COVID-19 mortality is defined as the cumulative number of COVID-19 deaths until the end of February 2022 per 100,000 inhabitants and settlement density as population per km² of settlement and traffic area. The red line in the right panel indicates a univariate regression of the log of COVID-19 mortality on the log of (standardised) settlement density and corresponds to the estimates in Table 2.3, Column (1).

This paper investigates the vulnerability of populations in high-density urban areas to COVID-19 in Germany in the first two years of the pandemic. It aims to discern whether living in densely populated areas increases the risk of mortality compared to other locations within the country. For this research, *mortality* is defined as the rate of deaths within a given timeframe. Three fundamental questions guided our research: First, we test if there is a positive correlation between settlement density and mortality during the pandemic. Second, we quantify the magnitude of this potential density effect on mortality relative to a broad range of covariates, notably geographic location within Germany. Third, we analyse how these correlations vary over time and space, with a focus on high-density urban areas.

---

interactions primarily occur in built-up areas and less in the countryside. Given the administrative structure of German counties, the density differences between urban and rural counties are more pronounced, and it is the differences between urban and rural counties that we are mainly interested in in this paper.

Our analysis reveals a consistent positive correlation between settlement density and COVID-19 mortality across Germany for each month following October 2020. This correlation intensified during the pandemic's first phase when non-pharmaceutical interventions (NPIs) by the federal government, such as *"lockdown-light,"* were either ineffective or implemented too late to curb the virus spread (Schuppert et al., 2021). However, this relationship shifted after Christmas 2020 following the implementation of stricter NPIs, which reduced the heightened threat previously reported in cities. These findings suggest that effective containment measures, coupled with the vaccination campaign, gradually diminished the impact of density on COVID-19 mortality.

In conclusion, our findings affirm that urban areas are inherently more vulnerable to COVID-19 due to the nature of virus transmission. Denser cities are predisposed to a faster spread of infections. In combination with an early arrival of the virus in denser areas, faster virus transmission leads to a head start in infection rates in denser areas which transmitted in higher COVID-19 mortality early in the pandemic. This lead was not caught up by less densely populated areas until effective medical interventions became widespread.

However, this logic of the virus's early arrival also applies to geographic location within Germany, influenced by the spread of the virus from south-east Europe. We examined this geographic effect using regression models enhanced with longitude and latitude, referred to as Geo additive Models. The strategies Germany adopted to contain the virus throughout the pandemic meant that the influence of geographical location ultimately overshadowed the influence of population density on COVID-19 mortality.

To assess the impact of various risk factors on COVID-19 mortality, we used linear and Generalised Additive Models (Wood, 2017) based on standardised covariates to yield effect size estimates. Our analysis shows that in the linear model, the combined influences of longitude and latitude on mortality by February 2022 were three times greater than that of population density. This finding indicates that the proximity of local populations to areas with high infection rates and the regional dispersion of the virus have a more pronounced effect on COVID-19 vulnerability than population density alone.

To our knowledge, this paper is a pioneering investigation into the longer-term evolution of the geographic distribution of COVID-19 mortality across Germany, marking a unique approach by comprehensively accounting for a wide array of health, policy and socio-economic factors. These factors include a regional policy stringency index, the distribution of diabetes prevalence, air pollution and regional vaccination rates. Our findings contribute significantly to the ongoing discussion on urban vulnerability to pandemic-induced mortality. We demonstrate that while cities experience higher conditional COVID-19 mortality compared to less densely populated counties, the relative impact of population density diminishes when considered alongside various arbitrary factors and those influenced by policy.

This is highly relevant for densely populated countries like Germany. Our analysis allows us to assess the impact of distinct pandemic phases and evaluate the effectiveness of various policy responses. Our findings show that, given increasing global interdependence, regional sprawl cannot be the dominant strategy for planners to avoid the consequences of a pandemic. Instead, efficient information management is needed to identify and contain hot spots, emphasizing that controlling the pandemic's spread is more crucial than merely dispersing populations. Achieving this objective necessitates enhanced international cooperation.

The structure of the remainder of this paper is as follows. Section 2 discusses the existing literature, exploring the interplay between population density, COVID-19 outcomes and the trajectory of the pandemic in Germany. Section 3 provides a detailed overview of how the COVID-19 pandemic unfolded in Germany. Section 4 outlines our research methodology, including the control variables used in the analysis and a description of our dataset. Section 5 presents our main findings, following an empirical three-step approach. Section 6 discusses the results as well as policy implications. Finally, Section 7 concludes the paper, addresses its limitations, and suggests directions for future research.

2.2 Related Literature

Throughout the COVID-19 pandemic, population density and urbanity have been focal points in mortality models across various regional analyses. Sigler et al. (2021) estimate density parameters during a six-week period (4 March to 8 April 2020), across 84 countries. Their findings suggest that virus spread followed a descent through the urban hierarchy rather than moving by adjacency, especially in the pandemic's early stages.

Díaz Ramírez et al. (2021) approached this issue from the perspective of excess mortality in an international comparison, considering place-based factors and controlling for spatial autocorrelation. Their analysis revealed that while health-system capacity significantly mitigated excess mortality, population density emerged as a critical risk factor. This is consistent with Moosa and Khatatbeh (2021) who observed a *"density paradox,"* where the expected negative outcomes associated with higher density did not always materialize, echoing the patterns seen in our research (Fig1, Panel 2).

Further, González-Val and Sanz-Gracia (2022) assessed the impact of urban factors, including population density, using data available up to 31 December 2020. They find that the proportion of urban people is more influential in spreading the virus than population density. Additionally, their study highlighted the importance of spatial spillovers and non-linear effects in understanding the spread of COVID-19. Intriguingly, their inclusion of a policy stringency index revealed a positive correlation, suggesting that stringent measures often followed rather than pre-empted spikes in COVID-19 cases.

Jo et al. (2021) provide insight into the initial impact of the pandemic on high-density counties at a sub-national level, proposing that such areas were initially more vulnerable due to either a size effect or a centrality effect. Specifically, they suggest a larger population heightens the likelihood of the virus being brought into the city by residents returning from other regions or through visits by infected individuals. Geng et al. (2021) further expand on this by describing the virus spread as "*space-filling*," emphasizing that the trajectory of the pandemic significantly depends on the stringency of travel restrictions and the implementation of NPIs.

Numerous papers have identified a positive relationship between population density and COVID-19 mortality (Wong & Lin, 2020; Arbel et al., 2022; Diao et al., 2021; Badhra et al. 2021; Dahal et al., 2021; Pilkington et al., 2021; Feinhandler et al., 2020; Smith et al., 2021). Despite these findings, there remains a notable gap in the literature: a comparative analysis assessing the magnitude of population density's impact relative to other factors. This omission leaves a critical question unanswered: In the broader context of COVID-19 mortality, how significant is the influence of population density compared to other determinants? Addressing this question is crucial for understanding the practical implications of population density on pandemic outcomes.

Wheaton and Kinsella Thompson (2020) and Carozzi et al. (2020; 2022a) focused their research on US metropolitan statistical areas (MSA). They found the effects associated with high population density began to wane shortly after the initial outbreak of COVID-19. In parallel, Carozzi et al. (2022b) observed similar patterns within UK districts, indicating that the negative impact of density on pandemic outcomes may diminish over time across different settings.

Carozzi et al. (2022a) explored different mechanisms, notably social connectedness, to understand the nuanced impact of population density on COVID-19 outcomes. Their findings highlight social connectivity as a significant factor that mediates the effects of urban density on the pandemic's spread. Using instrumental variables techniques to infer causal interference, they concluded that there is no direct causal relationship between urban density and COVID-19 mortality. This suggests that while urban density is closely associated with regional factors that may increase vulnerability, it does not directly escalate the risk of mortality from COVID-19.

Hamidi et al. (2020) studied the impact of activity density[5] on COVID-19 mortality using data from 913 metropolitan counties in the United States up to mid-May 2020. Using structural equation modelling, their study differentiated between the direct effect of density on mortality and the indirect effect mediated through the infection rate. Surprisingly, they found that densely populated counties, despite higher infection rates, exhibited a lower mortality risk. They attributed this negative effect to better healthcare systems and more effective management of NPIs in densely populated areas. These findings are consistent with Díaz Ramírez et al. (2021) and Moosa and Khatatbeh (2021). However, Hamidi et al. (2020) only compare metropolitan areas, not rural regions, so it is unclear if their estimates can be generalised to a complete country.

Neiderud (2015) emphasises that cities, as hubs of mobility, social interactions and trade, demand a nuanced analysis that accounts for the interplay between density, urban size, and connectivity in studying the vulnerability of local populations to the COVID-19 pandemic. Notably, connectivity often correlates positively with city size, which could confound analyses that do not account for this relationship. Our study builds on this premise by examining population size and testing for excessive collinearity to ensure interconnected variables do not skew our findings.

Density is ultimately only an indirect indicator of infection risk in a region. However, personal interactions within city hubs, such as shopping areas and households, are more important than density. This perspective is supported by micro-level analyses within urban conurbations, including studies on shopping districts (Zhou et al., 2022) and household transmission dynamics (Kashem et al. 2021; Hamidi & Hamidi, 2021). Additionally, research for example by Menéndez & García (2020) draws attention to the heightened vulnerability of socially deprived districts. Those results collectively suggest that while high-density areas may face significant risks at a pandemic's outset, particularly in the absence of effective non-pharmaceutical interventions

[5] Defined as the sum of population and employment density.

(NPIs), the overall risk to cities remains small and manageable (Khavarian-Garmsir et al., 2021; Liu et al. 2021a).

However, the geographic location within an urban area, particularly in relation to the initial outbreak site, may significantly influence COVID-19 risk. For example, several studies document a clear correlation between COVID-19 infection and deaths and the spatial distance to the Huanan Seafood Market, where the virus is believed to have originated (Worobey et al., 2022). Such spatial dispersion of the virus was particularly pronounced during the early phase of the pandemic, prior to the implementation of effective NPIs, as highlighted by studies like Peng et al. (2020), You et al. (2020), and Liu et al. (2021b). This principle of spatial risk factors may extend beyond urban settings to encompass entire nations. For example, in Germany, the virus's initial entry from the south exemplifies how national-level geographic factors can mirror the dynamics observed within individual cities.

Felbermayr et al. (2021) traced a superspreader event at the Ischgl ski resort in late February that, as a *"pandemic seed,"* significantly affected the geographical distribution of COVID-19 cases and deaths. Our research examines whether this influence diminishes over time. Ehlert (2021) used county-level data in Germany up to 15 June 2020 for a spatial analysis of socio-economic determinants of COVID-19. The study found a significant correlation between population density and case numbers, though not deaths. Krenz and Strulik (2021) introduce the *"benefit of remoteness"* concept in mitigating regional COVID-19 distribution, indicating that early pandemic mobility reductions could have markedly decreased infection rates. Finally, Prinz and Richter (2022) investigate the link between long-term exposure to air pollution and COVID-19 infection and mortality up to February 2021. They found a strong correlation between fine particulate matter and COVID-19 infections and, to a lesser extent, COVID-19 mortality. Their analysis also considered population density and location effects, such as proximity to hot spots on the Czech Republic border, which appear to increase COVID-19 mortality.

Our research builds on and expands these findings by extending the analysis period, incorporating the effects of the vaccination campaign, and accounting for health-related risk factors such as diabetes prevalence and the non-linear effects of population density. Our study builds on Felbermayr (2021), Krenz und Strulik (2021) and Prinz and Richter (2022), which underscore the significance of proximity to hotspots in determining a local population's vulnerability to COVID-19 suggesting that the location relative to hotspots may play a more significant role in a local population's vulnerability to COVID-19 than the *"urban penalty."*

Our work diverges by closely examining how the influence of these factors evolves. Understanding this evolution is crucial for designing and developing pandemic response policies. Furthermore, Teller (2021) published a meta-analysis of 15 studies on the relationship between urban density and COVID-19 outcomes. Teller highlights this relationship's dynamic nature, showing that the outcomes are not constant but change over time.

Alidadi and Sharifi's (2023) comprehensive literature review delved into the relationship between urban density and COVID-19 outcomes, analysing 134 papers written up to the end of 2021. Their review revealed that 58 % of these studies found a positive relationship between density and COVID-19 outcomes, such as infections and mortality rates. Despite this majority, the overall mixed results studies indicate that the relationship between density and COVID-19 mortality is not definite. The authors caution against oversimplifying the complex role of cities during a pandemic, emphasising the need to consider the specific contexts of each study as the relationship between density and COVID-19 decreased over time. An interesting observation from their review is that while cities might get hit first, this does not necessarily mean they get hit harder.

Given these insights, Alidadi and Sharifi's work serves as a foundation for our subsequent research examining the impact of urban density on COVID-19 outcomes over an extended period, with a specific focus on Germany. Such research should account for critical risk factors to understand better the nuances of how population density influences local populations' susceptibility to the pandemic. This approach is crucial for developing a more accurate and comprehensive understanding of the dynamics between urban density and pandemic outcomes.

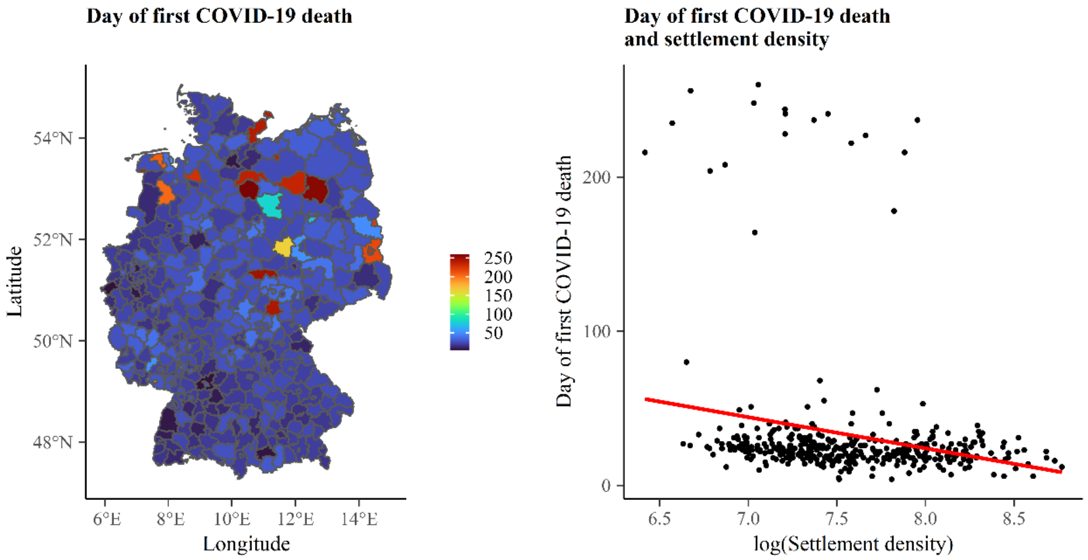
In summary, the influence of urban density on COVID-19 outcomes, as highlighted by Alidadi and Sharifi (2023), is markedly context-dependent, varying with the pandemic's phase, settlement patterns, the effectiveness of non-pharmaceutical interventions (NPIs), and overall population health. This underscores the need for more nuanced studies on the prolonged effects of density throughout the pandemic, based on the most influential control variables.

Our research aims to address this gap by focusing on previously unexplored dimensions. First, we investigate the longer-term effects of the COVID-19 pandemic regarding (urban) density and controlling for health factors, medical interventions, policy responses and geographical location within Germany. Second, we endeavour to compare the influence of density effects against other influential factors, especially to the effect of the relative location within the country. Third, we aim to discern the relative burdens experienced during different periods of the pandemic. Such a comparative analysis is essential for improving risk assessment and the strategic design of targeted NPIs.

2.2 The COVID-19 pandemic in Germany

As understanding of the virus grew, behavioural adaptations occurred both autonomously and due to external mandates, and crucial medical measures, notably vaccinations, were developed. Consequently, the vulnerability of cities fluctuated over time, influenced significantly by the effectiveness of policy responses (Teller, 2021).

Figure 2.2: Day of first confirmed COVID-19 death after 1 March 2020 and relationship with settlement density.



Notes: The day of first confirmed COVID-19 death after 1 March 2020 and settlement density are defined as population per km² of settlement and traffic area. The red line in the right panel indicates a univariate regression of the day of the first confirmed COVID-19 death after 1 March 2020 on the log of (standardised) settlement density.

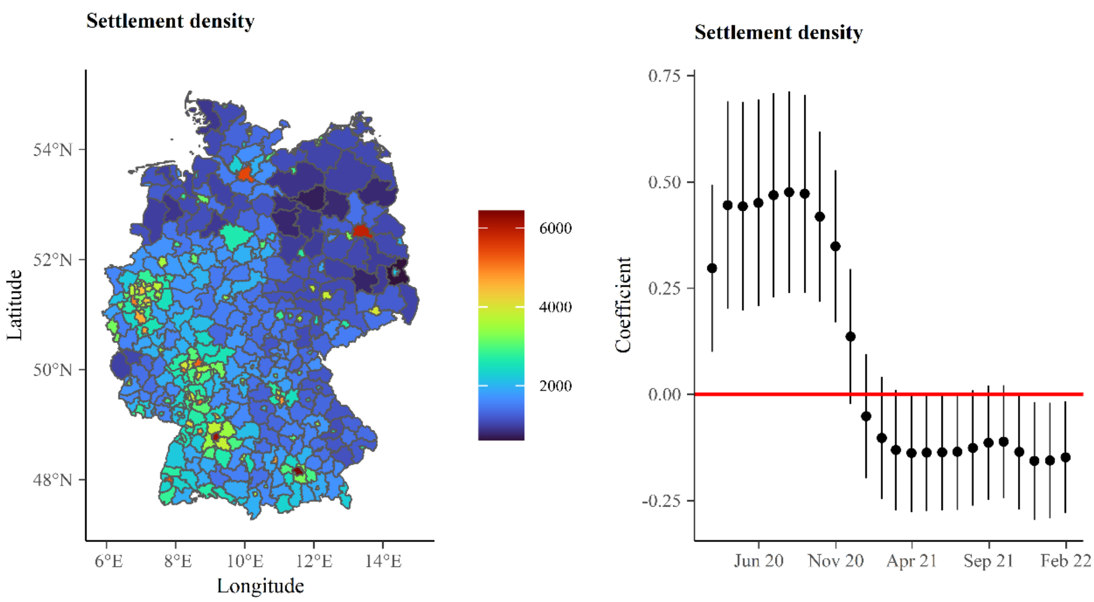
For Germany, there was a strong correlation between the number of days to the first reported COVID-19-related death after 1 March 2020 and settlement density (see Figure 2.2). Analysis reveals that counties with higher population densities reported their initial COVID-19 fatalities sooner than those with lower densities. Out of 400 counties, only 18 (4.5 %) reported no COVID-19 deaths during the pandemic's first wave,[6] including Schwerin, which ranks as the

[6] The first wave is approximately 150 days after 1 March 2020, the start of our dataset.

only one of these 18 counties in the top quartile for settlement density. Notably, all 18 counties are in the northern half of Germany.

Figure 2.3 depicts the trajectory of COVID-19 mortality in relation to settlement density in Germany, spanning the 24 months between March 2020 and February 2022. Our analysis identifies two distinct phases in this trajectory. Initially, until January 2021, a pattern emerged where counties with higher population densities experienced a relatively higher death toll. This pattern shifted in the subsequent phase, with higher mortality observed in counties with lower population densities. This shift is illustrated in Panel 2 of Figure 2.2, which depicts the development of the regression coefficient and its robust 95 %-confidence interval for a univariate regression of cumulative COVID-19 mortality on settlement density over the observation period.[7]

Figure 2.3: Settlement density and univariate regression coefficients of log of settlement density (standardised) on the log of COVID-19 mortality over time.



Notes: Regional distribution of settlement density on 31 December 2019 (left) and univariate regression coefficients of log of COVID-19 mortality (cumulated) on log of (standardised) settlement density (right).

[7] Both variables are on a log scale; the last estimate is the same as in Figure 2.1, Panel 2 and Table 2.3, Column (1).

2.3 Material and methods

We define the overall vulnerability of a county's population with respect to mortality as the cumulated number of confirmed COVID-19 deaths up to specific month T, normalised per 100,000 persons. Therefore, we define the COVID-19 mortality M in county i for a month T after February 2020 as the cumulative number of COVID-19 deaths R up to month T, per 100,000 persons N_i at the end of 2019:

$$M_i^T = \sum_1^T \frac{R_{i;T} * 100{,}000}{N_{i;2019}} \tag{2.1}$$

This indicator has limitations in fully capturing the extent of vulnerability. Recognising that a coronavirus infection can have consequences beyond death is essential. These consequences include long-term health issues even after a mild infection or economic repercussions due to increased sickness rates affecting productivity. Despite these constraints, mortality was chosen as the primary focus for our modelling for two key reasons: First, deaths are the most extreme outcome of a pandemic and second, mortality data tend to be more reliably recorded than infection rates. Furthermore, the Federal Statistical Office in 2022 reported that COVID-19 death numbers correspond closely to excess mortality observed during the infection spikes from 2020 to 2022. This correlation supports the use of mortality as a primary indicator of pandemic impact in our study.

To analyse the dynamic nature of the pandemic's impact over time, we employed cross-sectional regressions for each month from March 2020 to February 2022. This approach allows us to study the evolution of the coefficients, with each regression model assessing the cumulative impact of the pandemic up to the month in question.[8]

2.3.1 Other control variables

In assessing the impact of settlement density on COVID-19 mortality, we acknowledge the correlation that exists between population density metrics and various cross-regional variables. To address potential collinearity issues, particularly those affecting the estimation of standard errors for density coefficients, we strategically exclude non-health indicators that are highly correlated with our density measure.[9] However, this approach has certain limitations. Specifically, by focusing on broader measures of population density and excluding highly

[8] Note that we add a constant of 1 to the dependent variable for the regressions because of the logarithm.

[9] See the correlation matrix in Appendix A1.

correlated variables, we omit a detailed examination of crowdedness in more specific contexts, such as within households or public transportation systems, which have been highlighted in research like that of Hamidi & Hamidi (2021) and Harris (2021).[10] Consequently, our analysis may not fully capture the nuances of how different forms of crowdedness influence COVID-19 vulnerability.

In addition to the limitations concerning the specificity of crowdedness, our analysis does not directly incorporate the factor of connectivity, which, according to Jo et al. (2021), significantly influences the spread of SARS-CoV-2. It is important to clarify that in our study, the use of settlement density primarily serves as a spatial classification tool rather than establishing a causal relationship between density and COVID-19 vulnerability. Nevertheless, our approach enables a comparative analysis of the spatial associations between COVID-19 mortality and various spatial variables, notably longitude and latitude. These variables are relevant as they capture the dynamics of virus spread through regional dispersion among adjacent counties, as depicted in our Geoadditive Model. This comparative approach between (urban) densities and regional dispersion is crucial in understanding the conditional geographical distribution of COVID-19 mortality.

Our study incorporates several critical control variables to assess the conditional geographic distribution of COVID-19 mortality across Germany. We include each county's size, represented through the population count. Additionally, we factor in the stringency of NPIs deployed in each region, quantified through a regional policy stringency index. This index ranges from 0 to 100, with higher values implying stricter measures. To account for the time lag in the effectiveness and implementation of these policies, we calculate this measure as the average stringency level of the previous pandemic months, lagged by one month.

The capacity and quality of the public health infrastructure are pivotal in combating a pandemic or public health emergency (Kapitsinis, 2021). Our analysis includes two specific measures to capture the capabilities and vulnerabilities within the local healthcare system. First, we use the number of hospital beds per 1000 persons to indicate the local public health system's capacity to respond to the pandemic. Second, considering the heightened vulnerability of nursing home residents – where many people living in close quarters– we included the density of nursing home

[10] Those indicators are, among others, crowdedness indicators like the number of rooms per capita, overcrowding in households, household size, share of multifamily housing units and average living space, or connectivity indicators like commuting rates or accessibility measures.

residency in our models. This metric aims to capture the potential for outbreaks and the resulting high casualty rates in these settings, as Schweickert et al. (2021) identified.

To assess the individual health factors associated with COVID-19 mortality, our model includes several individual health variables known to elevate the likelihood of severe outcomes following an infection with SARS-CoV-2. This includes the proportion of the population aged 75 and older in each county (Sasson, 2021) and prevalence of type 2 diabetes within local populations, given the significant evidence linking this condition to heightened risk of severe COVID-19 (see Abdi et al., 2020; Kathe & Wani, 2021; Kinafar et al., 2022).

Furthermore, studies have linked long-term exposure to high air pollution levels to increased risk of contracting COVID-19 (Pozzer et al., 2020; Yu et al., 2021; Prinz & Richter, 2022; Díaz Ramírez et al., 2022). We include the long-term air pollution level, measured as the average particulate matter concentration (PM 2.5) between 2015 and 2020.

The most effective protection against viral infection is an effective vaccination. Research, including studies by Watson et al. (2022) and Shabalina and Matveeva (2023), has consistently demonstrated that vaccines against SARS-CoV-2 substantially reduce the mortality risk associated with the virus. In Germany, public access to COVID-19 vaccines commenced at the end of 2020 and by late summer 2021, vaccination was available for the entire population. A high vaccination rate in a community is pivotal in reducing the local population's vulnerability to COVID-19. Acknowledging this, we control for regional vaccination rates at the place of residence over time.

2.3.2 Data and descriptive statistics

Our study's data is primarily from the Corona Data platform provided by infas360. This platform aggregates COVID-related data from the Robert Koch Institute alongside socioeconomic data from the Federal Statistical Office of Germany and includes a monthly regional policy stringency index (Infas360, 2022). The scope of our data on COVID-19-confirmed deaths extends from 01 March 2020 to 28 February 2022, a period longer than most other studies conducted in Germany and incorporates one year of the vaccination campaign. This comprehensive dataset enables us to analyse the pandemic's substantial phases and assess the long-term effects of population density and geographical location on COVID-19 mortality across the country. To address potential reporting biases, particularly those stemming from variations in data recording during weekends and holidays, we aggregate the data monthly. Calculations per capita are based on population figures as of 31 December 2019.

In addition to the primary data source from the Corona Data platform, we calculate the longitude and latitude for each county based on a map of administrative borders. The prevalence of Type 2 Diabetes Mellitus among the adult population as of 2019 is sourced from the AOK Research Institute (2019). Long-term exposure to air pollution is measured using the average particulate matter 2.5 concentration level (PM 2.5) from 2015 to 2020. This data was provided by the German Environment Agency (2022) and averaged for each county (Table 2.1).

To further enhance our analysis, we constructed monthly vaccination rates for each county derived from data provided by infas360. These rates reflect the percentage of the adult population that has been fully vaccinated, based on two infas360 surveys conducted in February/March 2021 and January 2022. Infas360 used small-area-data methods to regionalize these survey results.[11] Given the temporal gap between the surveys, we apply a linear interpolation method to estimate vaccination rates for the intervening months. This enables us to model the gradual increase in vaccination coverage over time, offering a close approximation of the vaccination campaign's rollout trend throughout 2021.[12]

---

[11] To our knowledge, this is the only reliable source of county-level vaccination rate estimates in Germany, referring to the place of residence, not the place of vaccination, available at more than one point in time. More information can be found in infas360 (2021).

[12] Note that the vaccination campaign started in late December 2020 but was quantitatively limited.

Table 2.1: Variable definitions and data sources.

Variable	Definition	Time	Sources
COVID-19 deaths	Number of COVID-19 deaths per county	March 2020 – February 2022	Robert Koch Institute, Federal Statistical Office of Germany
Settlement density	Population per km² of settlement and traffic area	2019	infas360 GmbH, Regional database Germany, Federal Statistical Office of Germany
Population	County population	2019	Regional database Germany, Federal Statistical Office of Germany
Longitude	Longitude of a county's centroid	2020	Own calculations based on administrative borders map provided by the Federal Agency for Cartography and Geodesy
Latitude	Latitude of a county's centroid	2020	Own calculations based on administrative borders map provided by the Federal Agency for Cartography and Geodesy
Policy stringency index	Index (0-100) for the severity of NPIs per month.	March 2020 – January 2022	infas360 GmbH
Hospital beds	Hospital beds per 1,000 persons	2019	Regional database Germany, Federal Statistical Office of Germany
Nursing care	Persons in stationary nursing care per 10,000 inhabitants	2019	Federal Institute for Research on Building, Urban Affairs and Spatial Development, long-term care statistics of the Federal Government and the states, own calculations
Population aged 75+	Share of persons aged 75+	2019	Regional database Germany, Federal Statistical Office of Germany, own calculations
Diabetes prevalence	Share of persons suffering diabetes mellitus type-2	2019	AOK Research Institute (2019)
PM 2.5 concentration	Average particle concentration of 2.5 microns between 2015 and 2020	Average 2015-2020	German Environment Agency (2022), own calculations
Vaccination rate	Share of fully vaccinated persons at the place of residence	February/ March 2021 & January 2022	infas360 GmbH

We adjust the dataset for missing data points, leading to a final set of 394 out of 400 counties, spanning the 24-month study period. The descriptive statistics for the complete observation period are displayed in Table 2.2; county maps for all variables can be found in Appendix A2.

Table 2.2: Descriptive statistics.

Variable	Mean	St. Dev.	Median	Min	Max
COVID-19 mortality*	166.1	92.5	144.1	31.7	621.9
Settlement density	2,168.2	1,057.5	1,878	612	6,439
Longitude	9.9	2.0	9.8	6.2	14.8
Latitude	50.6	1.8	50.6	47.6	54.8
Population	209,514.1	246,750.5	156,885.5	36,789	3,669,491
Policy stringency index**	40.7	2.3	40.9	37.3	46.8
Hospital beds	6.4	3.9	5.6	0.2	29.2
Nursing care	102.1	26.7	99.0	47.8	212.1
Population aged 75+	11.8	1.7	11.6	7.8	17.9
Diabetes prevalence	8.9	2.0	8.5	4.8	15.4
PM 2.5 concentration	9.6	1.1	9.5	6.8	13.4
Vaccination rate***	0.9	0.1	0.9	0.7	1.0
Notes:	*Based on 394 counties in Germany. *Cumulated, February 2022. **Average value March 2020 - January 2022. *** January 2022.*				

2.3.3 Model

In the final regression equation, we add a constant of one to the COVID-19 mortality (M^T) to facilitate the use of a logarithmic scale. This adjustment allows us to interpret the regression parameter as elasticities for variables in logarithmic form and as semi-elasticities for proportional variables. For our estimations, we standardised the explanatory data, enabling a direct comparison of effect size by examining the estimated coefficients.[11]

Our methodological approach leads us to the following regression equation, applied for each month T from March 2020 to February 2022. In sum, we estimate 24 multivariate models – one for each month since the start of the pandemic up to month T after February 2020:

[11] Log-terms are shifted after standardisation according to $x - min(x) + 1$.

$$ln(M^T + 1) \ = \ \beta_0^T + X\beta_X^T \ + Z^{T-1}\beta_Z^T + \ \varepsilon^T \tag{2.2}$$

In the regression equation, β_0^T represents the intercept for each specific month T. The term $X\beta_X^T$ encompasses all time-invariant variables including settlement density and longitude and latitude terms and their corresponding month-specific regression parameters. In addition, β_Z^T denotes a set of corresponding month-specific regression parameters for each time-varying variable, which we collect in Z^{T-1}, such as the policy stringency index and the vaccination rate. Lastly, ε^T denotes the normally distributed errors in the model for each month T.

In the next step of our analysis, we relax the assumption of linear relationships by considering potential non-linear relationships between the logarithm of COVID-19 mortality and all explanatory variables. To estimate these non-linear relationships, we use the framework of Generalised Additive Models (GAM). Each regression parameter can be substituted in this approach with a smoothing term $s(\cdot)$ with appropriate characteristics (Wood, 2017). We use Thin-Plate-Regression-Splines for univariate smooth estimations (Wood, 2003).

To compare with the linear model presented earlier, our strategy involves initially estimating model (2.2) with all parameters smoothed. This helps identify which factors have a non-linear relationship with COVID-19 mortality. Following this, we estimate a final model that includes linear and non-linear components. We use the effective-degrees-of-freedom (edf) to determine which smoothing term can be reduced to a linear factor; any factor with an edf greater than one is considered in the non-linear part of the model. In our smoothing process, we avoid shrinkage and retain all covariates in the final model.

A critical advantage of the Generalised Additive Model is its ability to model unobserved geographic factors as a smoothed surface across longitude and latitude factors. Our approach employs a Spline-on-the-Sphere for estimating a longitude-latitude smooth, a method well-suited to respecting the Earth's curvature (Wood, 2017). This method allows us to 'search' for unobserved geographic factors and its origins in the distribution of COVID-19 mortality.

Specifically, this method can be instrumental in examining whether proximity to certain locations, such as the Ischgl ski resort – a noted early hotspot of COVID-19 transmission – continues to play a significant role in the geographic distribution of COVID-19 mortality even after two years of the pandemic. Should the distance from Ischgl (or similar points of interest) remain a critical geographic factor, our spatial smooths would reveal higher values in proximity to Ischgl and diminish with distance.

To conduct this analysis, we map the estimated spatial smooths across Germany at different points in time. Note that we do not log-transform the longitude and latitude to map the estimate of the spatial smooth term directly from the results. Thus, we estimate the following model:

$$ln(M^T + 1) \ = \beta_0^T + s(X^*) \ + s(Z^{T-1}) + \ s(x_{longitude}, x_{latitude}) + \ \varepsilon^T \qquad (2.3)$$

In this model, the smoothing terms are indicated by $s(\cdot)$, and X^* includes the same time-invariant variables as X in Formula (2.2), except for longitude and latitude. Time-varying factors are grouped under Z^{T-1}. The term $s(x_{longitude}, x_{latitude})$ represents a smoothed interaction term of the longitude and latitude, which reflects the (unobserved) conditional geography of COVID-19 mortality. By doing so, we contribute valuable insights that can inform future policy decisions and public health strategies, especially in high-density urban areas where the interplay between density, policy, health and geography is complex. For the model's estimation, we use restricted maximum likelihood estimation, which is more robust against overfitting than other estimation approaches (Wood, 2011).

2.4 Results

2.4.1 Linear regression results for complete observation period

Table 2.3 represents the linear model estimates for the logarithm of COVID-19 mortality as of February 2022. Column (1) details the results from a univariate regression like the one depicted in Figure 2.1. Similarly, column (2) displays the results for a model with only the logarithm of longitude and latitude. In column (3), we add variables to the first model that are not directly associated with increased individual mortality risk from COVID-19. Conversely, column (4) includes individual risk factors. Column (5) combines both categories of variables – those associated with and without direct individual risk – but excludes the logarithms of longitude and latitude. Finally, column (6) represents the complete model, encompassing all covariates.

Given the correlation between settlement density and many control variables, we conducted a series of collinearity tests. The results are reassuring: none of the VIF values for the settlement density variable nor the mean VIF across all explanatory variables reported in Table 2.3 indicate elevated collinearity.

The coefficient for settlement density is negative and significantly different from zero in the univariate model in column (1) and the model excluding individual risk factors. However, this changes to a significantly positive coefficient when individual health factors are controlled for, as shown in column (3). This shift suggests that when socioeconomic and location factors are not considered, the impact of settlement density on COVID-19 mortality is underestimated.

As seen in column (4), the omission of latitude and longitude in the model leads to an overestimation of the settlement density coefficient. This is because population density is generally higher in southern Germany than in the north. When longitude and latitude factors are included, as in column (5), the adjusted R^2 increases by more than ten percentage points.

In column (6), both location factors significantly differ from zero at any conventional confidence level. Using standardised data, we can directly compare the effect size by examining the coefficient estimates: One standard deviation closer to the south and east has a combined effect three times greater than an increase of one standard deviation in settlement density.

In the comprehensive model, the effects of population size and hospital beds are not statistically different from zero and negligible in size. This leads us to conclude that there is no direct (population) size effect on increasing COVID-19 mortality, nor is there a direct effect on the quality of the local healthcare system. This finding is plausible for a country like Germany, where regional differences in healthcare are minor, and no region experienced a significant breakdown of the local healthcare system during the COVID-19 pandemic.

The policy stringency index's slightly negative yet significant coefficient indicates the general efficiency of the German response to the pandemic: counties with higher average policy stringency exhibit lower COVID-19 mortality, *ceteris paribus*, compared to counties with more laissez-faire approaches. It is important to note that this observation does not necessarily reflect the absolute efficiency of the policy measures implemented, as we do not have data on the counterfactual scenario.[12]

Among individual risk factors in relation to COVID-19 mortality, long-term exposure to air pollution, measured as average PM 2.5 concentration, emerges as having a comparable impact on to that of settlement density. Furthermore, while a higher proportion of the population aged 75+ is a considered risk factor, its effect appears less significant compared to the influence of nursing care presence and diabetes prevalence within the population. Those two variables are linked to age as most nursing care targets the elderly and the probability of suffering diabetes

[12] For example, see Kosfeld et al. (2021) for an assessment of different policy measures in the first pandemic wave using a difference-in-difference approach.

mellitus type II increases with age. From this, two conclusions emerge. First, care facilities for the elderly emerge as particularly vulnerable environments, contributing significantly to the pandemic's death toll. Secondly, the analysis suggests that age, in isolation, may not be as critical a risk factor as previously thought; instead, it is the presence of underlying medical conditions, such as type II diabetes, that considerably heightens the risk.

Additionally, the study found a negative correlation between vaccination rates and COVID-19 mortality, suggesting that counties with lower vaccination rates experience higher mortality.

Table 2.3: Regression results for end of February 2022.

	Dependent variable:					
	log (COVID-19 mortality)					
	(1)	(2)	(3)	(4)	(5)	(6)
Log (Settlement density)	-0.148^{**}		-0.197^{**}	0.292^{***}	0.389^{***}	0.241^{***}
	(0.068)		(0.084)	(0.060)	(0.069)	(0.072)
Log (Longitude)		0.539^{***}				0.229^{***}
		(0.060)				(0.056)
Log (Latitude)		-0.214^{***}				-0.558^{***}
		(0.060)				(0.069)
Log (Population)			-0.077		-0.208^{***}	-0.052
			(0.081)		(0.066)	(0.064)
Policy stringency index			0.104^{***}		0.019	-0.060^{***}
			(0.023)		(0.020)	(0.021)
Hospital beds			0.073^{***}		-0.041^{*}	-0.027
			(0.025)		(0.022)	(0.020)
Nursing care				0.034	0.040	0.062^{***}
				(0.026)	(0.027)	(0.024)
Population aged 75+				-0.057^{*}	-0.053	-0.023
				(0.033)	(0.033)	(0.028)
Diabetes prevalence				0.265^{***}	0.255^{***}	0.298^{***}
				(0.035)	(0.035)	(0.033)
Log (PM 2.5 concentration)				-0.136^{**}	-0.061	0.186^{**}
				(0.066)	(0.072)	(0.076)
Vaccination rate				-0.196^{***}	-0.196^{***}	-0.104^{***}
				(0.025)	(0.026)	(0.025)
VIF settlement density	-	-	1.602	1.481	2.301	2.902
Mean VIF	-	1.005	1.371	2.059	2.045	2.253
R^2	0.013	0.211	0.087	0.475	0.491	0.609
Adjusted R^2	0.010	0.207	0.078	0.467	0.479	0.598
Akaike Inf. Crit.	564.839	478.647	539.937	325.908	320.051	220.025
Notes:	*Constant not reported. All explanatory variables are standardised, and log-terms are shifted by x - min(x) + 1. Robust standard errors: *p < 0.05; **p < 0.01; ***p < 0.001.*					

2.4.2 Development of linear regression parameters over time

Implementing the methodology detailed in Section 4.3, we conducted 24 cross-sectional regression models corresponding to the full model showcased in column (6) of Table 2.3. These models were executed for every month from March 2020 to February 2022. Figs 4 and 5 display the results for each regression coefficient and the corresponding robust 95 %-confidence intervals.

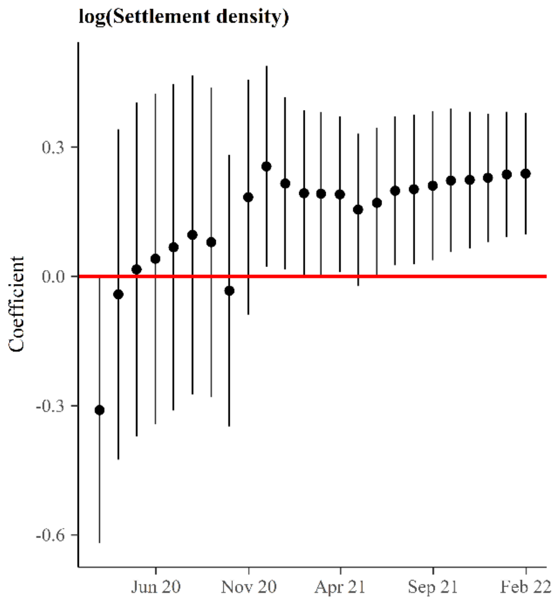
Our study identifies two distinct phases in the trajectory of the COVID-19 pandemic in Germany, each characterised by different spatial dynamics, social behaviours, and policy interventions. The first period, lasting from the pandemic's onset until the end of 2020, was marked by a lack of effective medical interventions and a reliance on broad, stringent containment strategies. The second period, which commenced at the beginning of 2021, saw the initiation of the vaccination campaign and containment measures more focused on vulnerable locations and situations.

The start of the vaccination campaign towards the end of 2020 coincided with a strict lockdown beginning December 16. This lockdown curbed the interregional spread of the virus, effectively mitigating the impact of the first and second waves. The success of this stringent measure contrasts starkly with the previous *"lockdown-light"* employed during the fall and early winter of 2020. The lighter restrictions were less effective and, as Schuppert et al. (2021) detailed, allowed the virus to spread throughout the country and within counties. This is reflected in the increasing coefficient of the policy stringency index during the autumn of 2020.

The observed increase in coefficient estimates for longitude and latitude corresponds with the geographic distribution of initial COVID-19 deaths, notably in northern and eastern Germany. Similarly, the significant coefficients for PM 2.5 concentration and diabetes prevalence shed light on the pandemic's differential impact across the country. Both factors were more prevalent in the northeast than southern regions and explain the sharp drop in the settlement density coefficient in the univariate models (see Figure 2.3).

However, southern, and high-density counties had a lead in COVID-19 cases that did not completely level off by the end of the observation period. This indicates that a relative "surplus" of deaths accumulated in more densely populated counties and southern and eastern counties during November and December 2020. This surplus persisted until the end of our study period. Cities and counties in southeast Germany suffered a relatively higher conditional mortality risk.

Figure 2.4: Coefficient estimates for the (log of) settlement density in the full model over time.

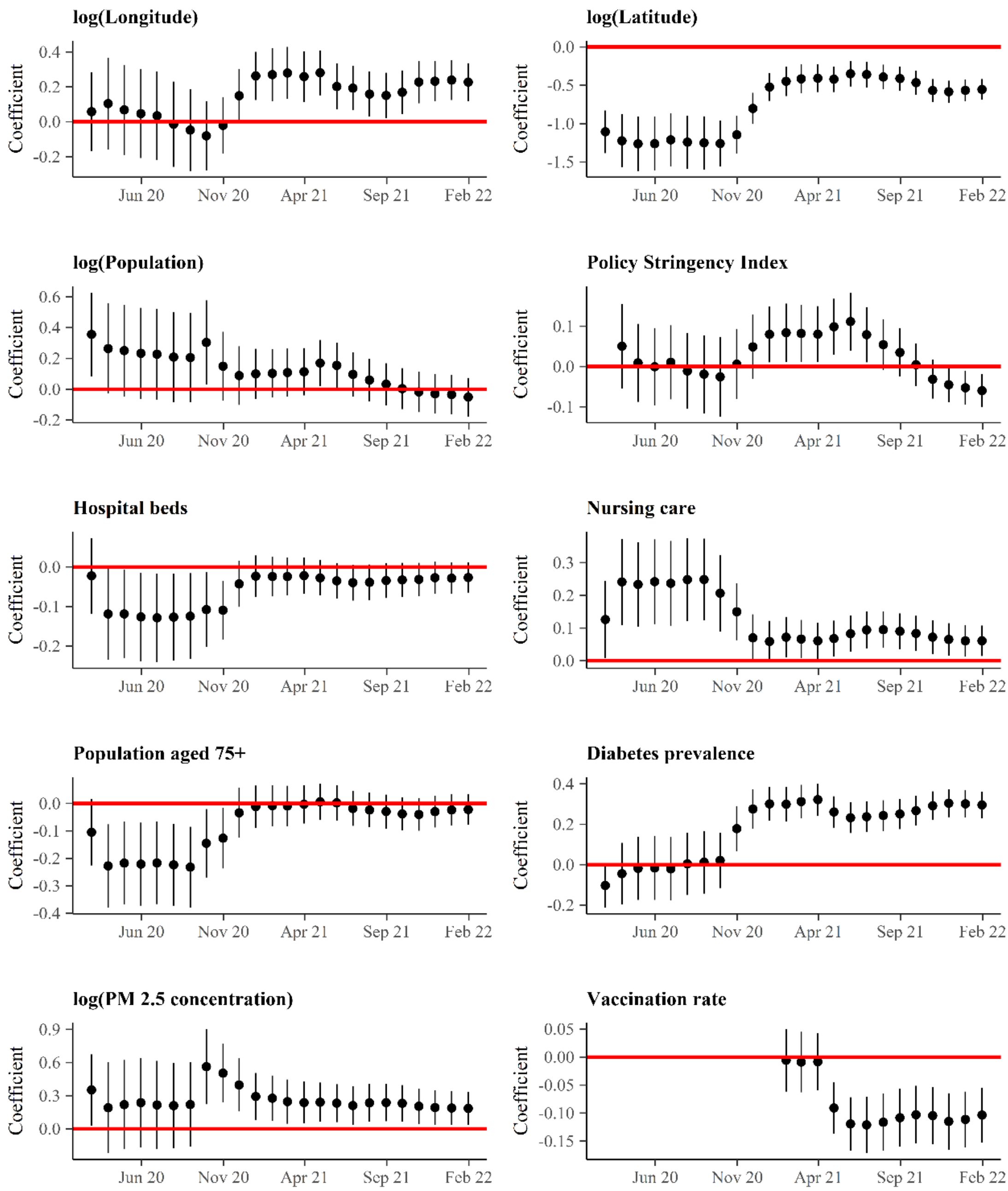


Notes: Coefficient estimates for the log of (standardised) settlement density in the full model from Table 2.3, Column (6) over time.

In the initial phase of the pandemic, the virus caught vulnerable individuals unprepared, including those with medical disorders and/or those in care facilities. These groups contributed strongly to the mortality dynamics during this period. A population size effect was notable only in the first two months. While the coefficient on density increased during this early phase, the coefficient on population decreased. This indicates that densely populated areas experienced an early arrival of the virus, but density, or more specifically crowdedness, fostered virus transmission, depending on population behaviour (Gerritse, 2022).

In the latter period, COVID-19 deaths became more a matter of personal risk assessment. The vaccination campaign, initiated in late December 2020, was initially limited by vaccine availability, but coincided with the decay of the second wave. Vaccinations began in the most affected counties, potentially explaining the three insignificant coefficients of vaccination rates in spring 2021. Subsequently, we observed a significantly negative impact of vaccination rates on COVID-19 mortality, indicating the effectiveness of the vaccination campaign in reducing deaths.

Figure 2.5: Other coefficient estimates in the full model over time.



Notes: Other coefficient estimates in the full model from Table 2.3, Column (6) over time.

2.4.3 Results of generalised additive models with spatial smooth interaction term

Using Generalised Additive Models means we could move beyond the restrictive assumption of linearity in our analysis. Initially, we constructed a model covering the entire observation period until February 2022. In it, we incorporated smooth terms for all exogenous variables, including a spatial smooth interaction term, to explore these complex relationships (detailed results are provided in Appendix A3).

The effective degrees of freedom (edf) indicate the complexity of the relationship between each variable and COVID-19 mortality. For several variables, such as the logarithm of population size, the proportion of the population aged 75+, and the logarithm of long-term air pollution levels, the edf values were at or very close to one, indicating a linear relationship with COVID-19 mortality. This suggests that these relationships can be adequately modelled as linear, leading to their inclusion in the linear component of our refined model. However, for settlement density, with an edf close to but not precisely one, we allowed for a non-linear relationship with COVID-19 mortality in the optimal model.

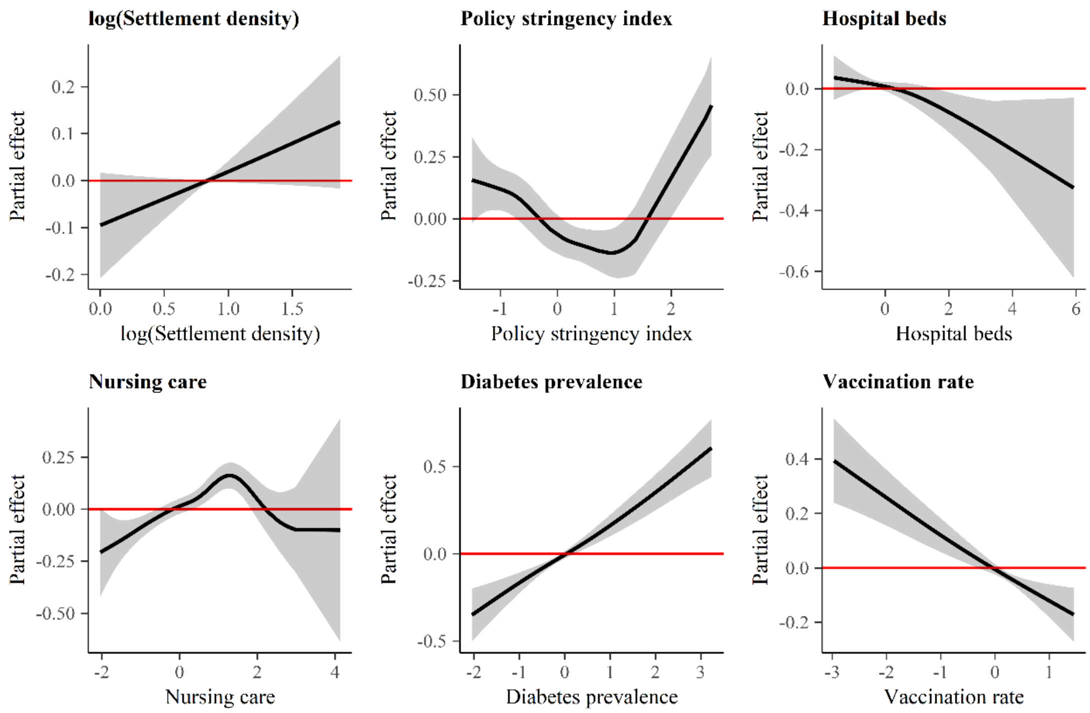
We then estimated an optimal model that integrates linear and non-linear elements. The findings are detailed in Table 2.4. A significant portion of the geographic variation in COVID-19 mortality is attributed to the geographic factor, as opposed to specific control factors. Despite this, the non-linear model explains a higher proportion of the variance in COVID-19 mortality, as indicated by an adjusted R^2 of 85 %, a substantial increase from the 60 % observed in the full linear model. While the coefficient for diabetes prevalence remains significant, the other linear terms are no longer statistically significant and exhibit reduced effect sizes compared to those in the linear model.

Table 2.4: Regression results for the optimal Generalised Additive Model.

	Dependent variable:			
	COVID-19 mortality			
Parametric coefficients:				
	Estimate	Std. Error	t	p-Value
Log (Population)	-0.09699	0.05700	-1.702	0.0898
Population aged 75+	0.04254	0.02252	1.889	0.0598
Log (PM 2.5 concentration)	0.19571	0.10417	1.879	0.0612
Approximate significance of smooth terms:				
	edf	Ref. df	F	p-Value
s(log(Settlement density))	1.055	1.101	2.865	0.0960
s(Longitude, latitude)	48.059	149.000	2.858	0.0000***
s(Policy stringency index)	3.658	4.331	6.755	0.0000***
s(Hospital beds)	1.807	2.290	3.045	0.0389*
s(Nursing care)	5.029	6.141	5.284	0.0000***
Diabetes prevalence	1.875	2.329	21.343	0.0000***
s(Vaccination rate)	1.293	1.487	18.839	0.0000***
Adjusted R^2	0.847			
Notes:	*Link-function: log. Optimal model with all explanatory variables standardised and log-terms are shifted by x - min(x) + 1. Constant not reported. Approximate significance of smooth terms: *p < 0.05; **p < 0.01; ***p < 0.001.*			

In the optimal model's non-linear section, the smooth terms analysis reveals that all except for the smooth associated with settlement density are significantly different from zero. It is important to note that a smooth term is meant to be significantly different from zero when its confidence interval does not include the zero line across the entire distribution of observations. For settlement density, this specific criterion is not met, as shown in Figure 2.6.

Figure 2.6: Non-linear part of the optimal model from Table 2.4.



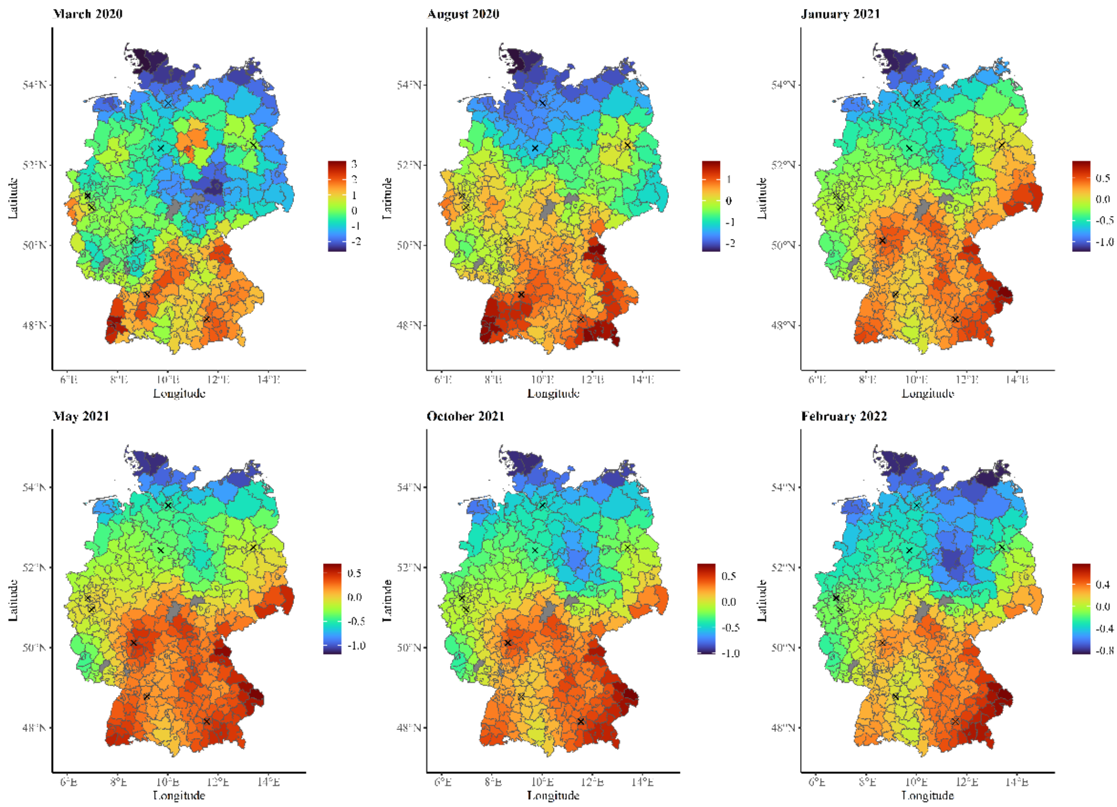
Finally, we examine the results for the spatial smooth interaction term from the optimal model discussed in the previous section, analysing it at different points in time. This term is pivotal as it captures the conditional geographic distribution of COVID-19 mortality across Germany, offering insights into how the impact of the pandemic has varied across different regions at various stages. The model's structure mirrors the linear model, integrating both time-varying and static variables, as outlined in Section 5.2. Figure 2.7 presents a series of maps, each representing the spatial smooth interaction term at six different months between March 2020 and February 2022.

The analysis of the spatial smooth interaction term within the optimal model reveals a pronounced north-south gradient in the conditional geography of COVID-19 mortality throughout Germany, a feature evident across all examined periods. However, this gradient is more pronounced in the pandemic's first phase, when the virus spread from the south of Europe. In the earlier months, particularly in March 2020, and to a lesser extent up to August 2020, the impact of the Ischgl hotspot is distinctly visible, registering the highest values of the spatial smooth interaction term.

As the pandemic progressed to the end of its first phase in January 2021, the influence of Ischgl and other intra-German hotspots significantly diminished. This gave way to a more marked eastwest gradient in COVID-19 mortality, with high values along the Austrian and Czech borders. This pattern remains consistent until the end of the observation period in February 2022.

Given that the response to the pandemic was primarily implemented at the national level and considering that location within Germany is the most influential factor for COVID-19 mortality, the observed border effect underscores an essential lesson: the need for enhanced European integration in future pandemic response strategies. This highlights the necessity for Germany and its neighbouring countries to create a cohesive pandemic strategy in a region characterized by open borders.

Figure 2.7: Spatial smooth interaction term based on the optimal model from Table 2.4 at six different time points.



Notes: Black crosses indicate the centroids of Germany's eight most populated counties. From north to south, they are Hamburg, Berlin, Hannover (region), Düsseldorf, Cologne, Frankfurt, Stuttgart and Munich.

However, as the south of Germany is more densely populated than the north, failing to control for location in terms of longitude and latitude or through comparable methods would lead to an overestimation of the effect of density on local population vulnerability. The findings from the Generalised Additive Models show that cities in Germany, broadly speaking, did not suffer from a higher conditional COVID-19 mortality risk. This insight, coupled with the significant influence of the spatial interaction term, further suggests that the risk of COVID-19 mortality is more strongly tied to vicinity and local interactions, i.e. regional dispersion, rather than an inherent *"urban penalty."* This observation is particularly relevant when considering the non-pharmaceutical interventions put in place, underscoring the importance of local context in understanding pandemic patterns.

2.5 Discussion and policy implications

High-density urban areas are susceptible to the spread of infectious diseases with similar transmission routes to coronaviruses. However, these same transmission characteristics lead to a sharply decreasing relative risk over time in such areas compared to rural regions. Throughout the pandemic, population density was less significant in determining mortality risk. Our study shows that regional proximity to hot spots is more important than urban density and that these hot spots can exhibit unpredictable patterns.

Most studies exploring the relationship between urban density and COVID-19 mortality have overlooked the crucial role of location and failed to compare these factors' impact adequately. Our findings suggest that, within Germany, the spread of respiratory viruses is more significantly influenced by regional dispersion rather than by population density. This contrasts with the spread dynamics observed in countries like Oman, where lower population density and a network of metropolitan interconnections present different transmission dynamics (Mansour et al., 2021).

Our findings reinforce lessons from past pandemics: High-density urban areas are not inherently disadvantaged in pandemics with transmission patterns similar to coronaviruses. The perceived disadvantages of high density are relatively minor and can be effectively addressed through appropriate policies. Consequently, the concern that people might relocate from urban areas in for safety is unfounded, and such a move could even be counterproductive. Building sustainable cities resilient to future pandemics hinges on enhancing informational connectivity, not population dispersion. Physical dispersion can offer temporary relief but is most effective when accompanied by a broader, comprehensive strategy.

Our research underscores the importance of precise and targeted policy interventions in managing pandemics, yielding five critical policy implications: Firstly, effective pandemic management hinges on rapidly recognising and comprehensively isolating hot spots of infections. The timely collection and analysis of infection data is paramount to identifying these areas quickly and accurately. Second, gaining insights into how hot spots connect to cities is crucial. Policies should be informed by data on the interconnectedness of these areas, focusing on regional proximity and transport links, to understand and address the spread. Third, extensive contact restrictions must be implemented quickly in regions with a high degree of connection to identified hot spots. This entails understanding high-interaction locations within cities and prioritising their isolation. This requires inner-city connectivity data to anticipate possible spread patterns. Fourth, dispersing populations may offer temporary protection but the effectiveness of other measures is crucial as the virus will proliferate through regional dispersion. Finally, effective pandemic management cannot stop at territorial borders. This applies equally to national and regional borders. International coordination is more crucial than inner-city restrictions for addressing the proximity to and influence of hot spots.

2.6 Conclusion

In this paper, we explored the influence of urban density on COVID-19 vulnerability across German regions and yielded significant insights that contribute to the ongoing discussion on pandemic management. There were three questions fundamental to our study.

First, we examined how urban density related to the spread of the pandemic. We discovered that urban density significantly influenced the spread of COVID-19, with high-density areas experiencing earlier and more rapid transmission. This density effect tapered off over time. As the pandemic progressed, the virus also spread into less populated regions. Our work shows that without effective NPIs, cities experienced a higher proportion of early infections. This excess did not fully diminish when effective medical interventions became available later. Finally, cities faced an elevated mortality risk, particularly when policy responses to contain the virus were delayed or ineffective in the aftermath of superspreader events.

Second, our findings show that while population density does impact on COVID-19 spread and mortality, its influence is modest compared to health factors and geographic location. Specifically, the linear model conducted two years after the pandemic onset indicates that the effects of population density account for only a third of the impact associated with location factors, such as longitude and latitude. The findings suggest that in a country like Germany, which is densely populated and highly interconnected, the risk of COVID-19 mortality is more determined by proximity to regions of high infections rather than population density.

As previously mentioned, the third question highlights the significant time- and space-dependent effects on the spread of COVID-19. This reinforces the results for the second research question, as the NPIs implemented throughout Germany during the observation period were, partly, enacted inefficiently and with delays. This lack of immediate response allowed the virus to spread within cities and across regions.

While our study provides insightful observations regarding the impact of population density on COVID-19 mortality, it is crucial to acknowledge certain limitations that must be considered when interpreting our results. First, our study uses the density variable primarily as a spatial classification scheme to distinguish between cities and rural areas and, therefore, cannot be interpreted directly in terms of causal relationships. Second, we could not separate the effects of crowdedness and population density due to high correlations between these variables. Similarly, data limitations prevented a clear separation of the impacts of interregional connectivity, with adjustments made only for population size, longitude and latitude. Third, we did not consider how the pathogenicity of the virus changed over time due to the emergence of new variants. Additionally, variations in the average time from infection to death and how these shifts affected the intertemporal dynamics under investigation were outside the scope of the analysis.

Future research could explore questions not addressed in this study. Specifically, it could investigate whether our findings hold for the spread of other diseases and examine how different pandemic response regimes interact spatially. This would enrich our understanding of disease spread and effective pandemic management in diverse policy and cross-border contexts.

Acknowledgements

The authors would like to thank AOK Research Institute, infas360 GmbH and the German Environment Agency for providing data. The authors have no conflicts of interest.

References

Abdi, A., Jalilian, M., Sarbarzeh, P. A., & Vlaisavljevic, Z. (2020). Diabetes and COVID-19: A systematic review on the current evidences. *Diabetes research and clinical practice, 166*, 108347.

Alidadi, M., & Sharifi, A. (2023). Beyond the blame game: Unraveling the complex relationship between density and COVID-19 through a systematic literature review. *Cities*, *141*, 104519.

AOK Research Institute (2019). *Gesundheitsatlas Deutschland – Diabetes mellitus Typ 2: Verbreitung in der Bevölkerung Deutschlands und seinen Regionen. Ursachen, Folgen und Präventionsmöglichkeiten.* [online] https://gesundheitsatlas-deutschland.de/ [27 November 2023]

Arbel, Y., Fialkoff, C., Kerner, A., & Kerner, M. (2022). Do population density, socio-economic ranking and Gini Index of cities influence infection rates from coronavirus? Israel as a case study. *The Annals of Regional Science, 68*(1)*,* 181-206.

Bhadra, A., Mukherjee, A., & Sarkar, K. (2021). Impact of population density on COVID-19 infected and mortality rate in India. *Modeling earth systems and environment, 7*(1)*,* 623-629.

Carozzi, F., Provenzano, S., & Roth, S. (2020). *Urban density and COVID-19.* IZA Discussion Papers, No. 13440, Institute of Labor Economics (IZA). Bonn.

Carozzi, F., Provenzano, S., & Roth, S. (2022a). Urban density and COVID-19: understanding the US experience. *The Annals of Regional Science*, 1-32.

Carozzi, F., Provenzano, S., & Roth, S. (2022b). Density and the Spread of COVID-19 in Cities: Lessons from the United States and the United Kingdom. In *European Cities After COVID-19: Strategies for Resilient Cities and Real Estate* (pp. 107-115). Cham: Springer International Publishing.

Dahal, S., Luo, R., Swahn, M. H., & Chowell, G. (2021). Geospatial Variability in Excess Death Rates during the COVID-19 Pandemic in Mexico: Examining Socio Demographic, Climate and Population Health Characteristics. *International Journal of Infectious Diseases, 113,* 347-354.

Diao, Y., Kodera, S., Anzai, D., Gomez-Tames, J., Rashed, E. A., & Hirata, A. (2021). Influence of population density, temperature, and absolute humidity on spread and decay durations of COVID-19: A comparative study of scenarios in China, England, Germany, and Japan. *One Health, 12,* 100203.

Díaz Ramírez, M., Veneri, P., & Lembcke, A. C. (2022). Where did it hit harder? Understanding the geography of excess mortality during the COVID-19 pandemic. *Journal of Regional Science*, *62*(3), 889-908.

Ehlert, A. (2021). The socio-economic determinants of COVID-19: A spatial analysis of German county level data. *Socio-Economic Planning Sciences, 78,* 101083.

Federal Statistical Office: Sonderauswertung zu Sterbefallzahlen der Jahre 2020 bis 2023, in *Federal Statistical Office*, 21 November 2023 [online] https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Sterbefaelle-Lebenserwartung/sterbefallzahlen.html?nn=209016#Vorjahre%20Methodik [27 November 2023]

Feinhandler, I., Cilento, B., Beauvais, B., Harrop, J., & Fulton, L. (2020). Predictors of death rate during the COVID-19 pandemic. *Healthcare*, *8*(3), 339.

Felbermayr, G., Hinz, J., & Chowdhry, S. (2021). Après-ski: The spread of coronavirus from Ischgl through Germany. *German Economic Review, 22*(4)*,* 415-446.

Geng, X.; Gerges, F.; Katul, G. G.; Bou-Zeid, E.; Nassif, H. & Boufadel, M. C. (2021). Population agglomeration is a harbinger of the spatial complexity of COVID-19 *Chemical Engineering Journal, Elsevier BV, 420*, 127702

German Environment Agency: Air pollution levels in Germany, in *German Environment Agency*, 2021, [online] https://gis.uba.de/maps/resources/apps/lu_schadstoffbelastung/index.html?lang=de [27 November 2023]

Gerritse, M. (2022). COVID-19 transmission in cities. *European Economic Review*, *150*, 104283.

Glaeser, E.: Cities and pandemics have a long history. In *City Journal,* Spring 2020, [online] https://www.city-journal.org/cities-and-pandemics-have-long-history [21 January 2026]

González-Val, R., & Sanz-Gracia, F. (2022). Urbanization and COVID-19 incidence: A cross-country investigation. *Papers in Regional Science, 101*(2), 399-415.

Hooper, M.: Pandemics and the future of urban density: Michael Hooper on hygiene, public perception and the "urban penalty." In *Harvard Graduate School of Design Newsletter*, 13 April 2020, [online] https://www.gsd.harvard.edu/2020/04/have-we-embraced-urban-density-to-our-own-peril-michael-hooper-on-hygiene-public-perception-and-the-urban-penalty-in-a-global-pandemic/ [21 January 2026]

Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *Journal of the American Planning Association, 86(4),*, 495-509.

Hamidi, S., & Hamidi, I. (2021). Subway ridership, crowding, or population density: determinants of COVID-19 infection rates in New York City. *American journal of preventive medicine, 60*(5), 614-620.

Harris, J. E. (2021). Critical role of the subways in the initial spread of SARS-CoV-2 in New York City. *Frontiers in public health*, *9*, 754767.

Infas360(2022): Corona Data platform. In: *Healthcare Data Platform*, February 2021, [online] https://www.healthcare-datenplattform.de/ [27 November 2023]

Infas360 (2021): Impfquote und Impfbereitschaft – Ergebnisse nach Bundesländern im Juli 2021 auf Basis einer Infas-Befragung. *Themenreport Corona-Datenplattform Nr. 3.,* Ausgabe August 2021.

Jo, Y., Hong, A., & Sung, H. (2021). Density or connectivity: what are the main causes of the spatial proliferation of COVID-19 in Korea?. *International Journal of Environmental Research and Public Health, 18*(10)*,* 5084.

Kathe, N. J., & Wani, R. J. (2021). Determinants of COVID-19 case fatality rate in the United States: spatial analysis over one year of the pandemic. *Journal of Health Economics and Outcomes Research, 8*(1), 51.

Kapitsinis, N. (2021). The underlying factors of excess mortality in 2020: a cross-country analysis of pre-pandemic healthcare conditions and strategies to cope with COVID-19. *BMC Health Services Research, 21*(1)*,* 1-19.

Khavarian-Garmsir, A. R., Sharifi, A., & Moradpour, N. (2021). Are high-density districts more vulnerable to the COVID-19 pandemic? *Sustainable Cities and Society, 70*, 102911.

Kianfar, N., Mesgari, M. S., Mollalo, A., & Kaveh, M. (2022). Spatio-temporal modeling of COVID-19 prevalence and mortality using artificial neural network algorithms. *Spatial and Spatio-temporal Epidemiology, 40,* 100471.

Kosfeld, R., Mitze, T., Rode, J., & Wälde, K. (2021). The COVID-19 containment effects of public health measures: A spatial difference-in-differences approach. *Journal of Regional Science, 61*(4)*,* 799-825.

Krenz, A., & Strulik, H. (2021). The benefits of remoteness–digital mobility data, regional road infrastructure, and COVID-19 infections. *German Economic Review, 22(3),* 257-287.

Liu, C., Liu, Z., & Guan, C. (2021a). The impacts of the built environment on the incidence rate of COVID-19: A case study of King County, Washington. *Sustainable cities and society*, *74*, 103144.

Liu, W., Wang, D., Hua, S., Xie, C., Wang, B., Qiu, W., Xu, T., Ye, Z., Yang, M., Xiao, Y., Feng, X., Shi, T., Li, M., & Chen, W. (2021b). Spatiotemporal analysis of COVID-19 outbreaks in Wuhan, China. *Scientific reports*, *11*(1), 13648.

Mansour, S., Al Kindi, A., Al-Said, A., Al-Said, A., & Atkinson, P. (2021). Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). *Sustainable cities and society*, *65*, 102627.

Menéndez, E. P., & Higueras García, E. (2020). Urban sustainability versus the impact of Covid-19: A Madrid case study. *disP-The Planning Review*, *56*(4), 64-81.

Moosa, I. A., & Khatatbeh, I. N. (2021). The density paradox: Are densely-populated regions more vulnerable to COVID-19?. *The International Journal of Health Planning and Management, 36*(5), 1575-1588.

Neiderud, C. J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infection Ecology & Epidemiology, 5*(1), 27060.

McFarlane, C. (2021). Repopulating density: COVID-19 and the politics of urban value. *Urban Studies,* 00420980211014810.

Peng, Z., Wang, R., Liu, L., & Wu, H. (2020). Exploring urban spatial features of COVID-19 transmission in Wuhan based on social media data. *ISPRS International Journal of Geo-Information*, *9*(6), 402.

Pilkington, H., Feuillet, T., Rican, S., de Bouillé, J. G., Bouchaud, O., Cailhol, J., Bihan, H., Lombrail, P., & Julia, C. (2021). Spatial determinants of excess all-cause mortality during the first wave of the COVID-19 epidemic in France. *BMC public health, 21*(1), 1-10.

Pozzer, A., Dominici, F., Haines, A., Witt, C., Münzel, T., & Lelieveld, J. (2020). Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research*, *116*(14), 2247-2253.

Prinz, A. L., & Richter, D. J. (2022). Long-term exposure to fine particulate matter air pollution: An ecological study of its effect on COVID-19 cases and fatality in Germany. *Environmental Research*, *204*, 111948.

Rocklöv, J., & Sjödin, H. (2020). High population densities catalyse the spread of COVID-19. *Journal of Travel Medicine, 27*(3), taaa038.

Sasson, I. (2021). Age and COVID-19 mortality. *Demographic Research, 44,* 379-396.

Schuppert, A., Polotzek, K., Schmitt, J., Busse, R., Karschau, J., & Karagiannidis, C. (2021). Different spreading dynamics throughout Germany during the second wave of the COVID-19 pandemic: a time series study based on national surveillance data. *The Lancet Regional Health-Europe, 6,* 100151.

Schweickert, B., Klingeberg, A., Haller, S., Richter, D., Schmidt, N., Abu Sin, M. & Eckmanns, T. (2021). COVID-19-Ausbrüche in deutschen Alten- und Pflegeheimen. *Epidemiolgisches Bulletin*, *18,* 3-29.

Sigler, T., Mahmuda, S., Kimpton, A., Loginova, J., Wohland, P., Charles-Edwards, E., & Corcoran, J. (2021). The socio-spatial determinants of COVID-19 diffusion: the impact of globalisation, settlement characteristics and population. *Globalization and Health, 17*(1), 1-14.

Shabalina, S. A., & Matveeva, O. V. (2023). Comparison of Vaccination and Booster Rates and Their Impact on Excess Mortality During the COVID-19 Pandemic in European Countries. *Frontiers in Immunology, 14,* 1151311.

Smith, T. P., Flaxman, S., Gallinat, A. S., Kinosian, S. P., Stemkovski, M., Unwin, H. J. T., ... & Pearse, W. D. (2021). Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. *Proceedings of the National Academy of Sciences, 118*(25), e2019284118.

Sun, Z., Zhang, H., Yang, Y., Wan, H., & Wang, Y. (2020). Impacts of geographic factors and population density on the COVID-19 spreading under the lockdown policies of China. *Science of The Total Environment, 746,* 141347.

Tarwater, P. M., & Martin, C. F. (2001). Effects of population density on the spread of disease. *Complexity, 6*(6), 29-36.

Teller, J. (2021). Urban density and COVID-19: towards an adaptive approach. *Buildings and Cities, 2*(1), 150-165.

Watson, O. J., Barnsley, G., Toor, J., Hogan, A. B., Winskill, P., & Ghani, A. C. (2022). Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. *The Lancet Infectious Diseases, 22*(9), 1293-1302.

Wheaton, W. C., & Kinsella Thompson, A. (2020). Doubts about density: COVID-19 across cities and towns. Available at SSRN 3586081.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*(1), 95-114.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3-36.

Wood, S. N. (2017). *Generalized additive models: an introduction with R.* Second edition. CRC press.

Wong, D. W., & Li, Y. (2020). Spreading of COVID-19: Density matters. *Plos one, 15*(12), e0242398.

Worobey, M., Levy, J. I., Malpica Serrano, L., Crits-Christoph, A., Pekar, J. E., Goldstein, S. A., Rasmussen, A. L., Kraemer, M. G., Newman, C., Koopmans, M. P. G., Suchard, M. A., Wertheim, J. O., Lemey, P., Robertson, D. L. , Garry, R. F., Holmes, E. C. , Rambaut, A., & Andersen, K. G. (2022). The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science*, *377*(6609), 951-959.

You, H., Wu, X., & Guo, X. (2020). Distribution of COVID-19 morbidity rate in association with social and economic factors in Wuhan, China: Implications for urban development. *International journal of environmental research and public health*, *17*(10), 3417.

Yu, X., Wong, M. S., Kwan, M. P., Nichol, J. E., Zhu, R., Heo, J., ... & Kan, Z. (2021). COVID-19 infection and mortality: Association with PM2. 5 concentration and population density—An exploratory study. *ISPRS International Journal of Geo-Information, 10*(3), 123.

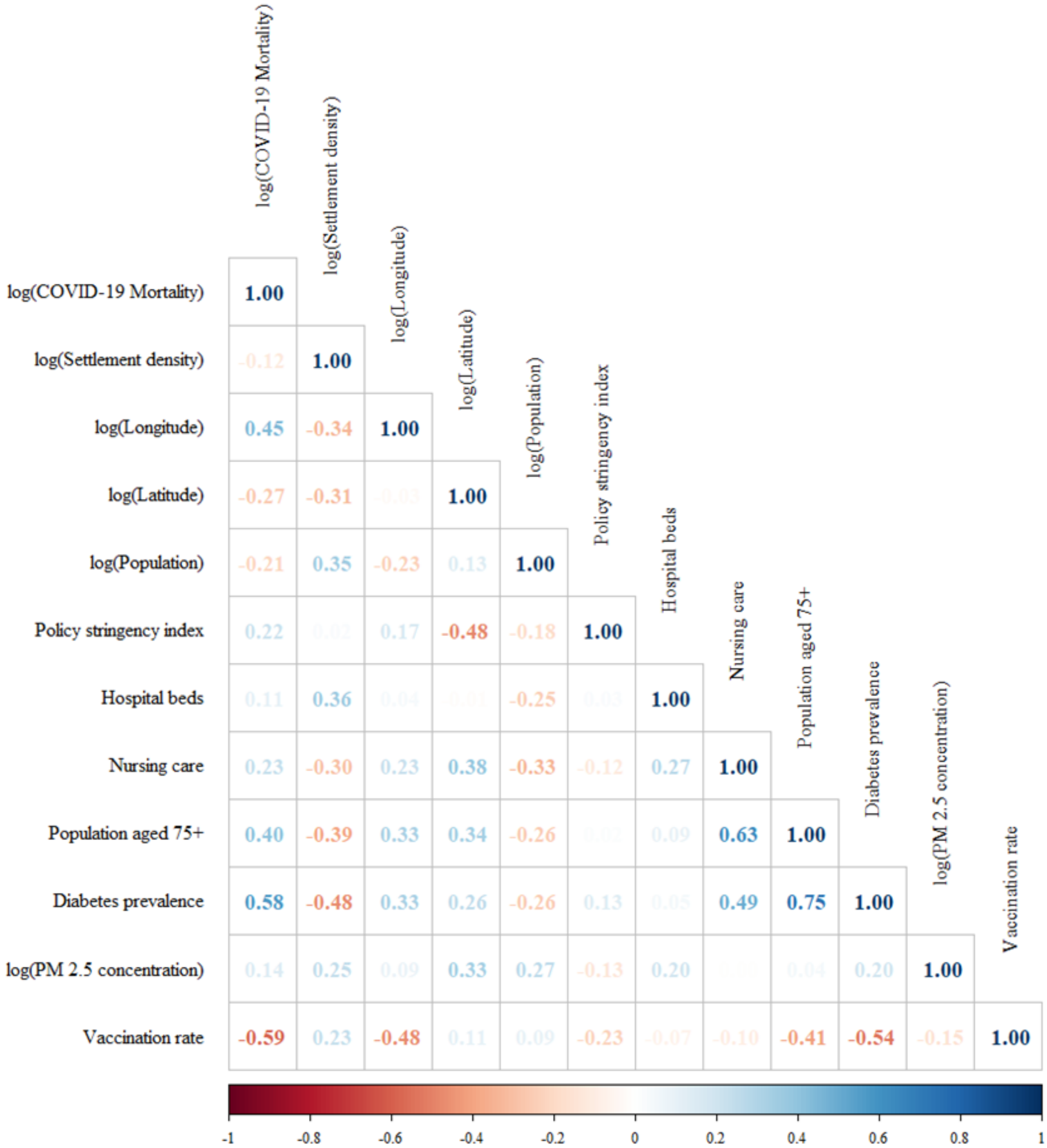
Zaki, B. L., Nicoli, F., Wayenberg, E., & Verschuere, B. (2022). In trust we trust: The impact of trust in government on excess mortality during the COVID-19 pandemic. *Public Policy and Administration,* 09520767211058003.

Zhou, L., Li, S., Li, C., Shen, G., Yang, H., Zhu, P., Han, H., & Li, B. (2022). Spatial congruency or mismatch? Analyzing the COVID-19 potential infection risk and urban density as businesses reopen. *Cities*, *123*, 103615.

A. Appendix

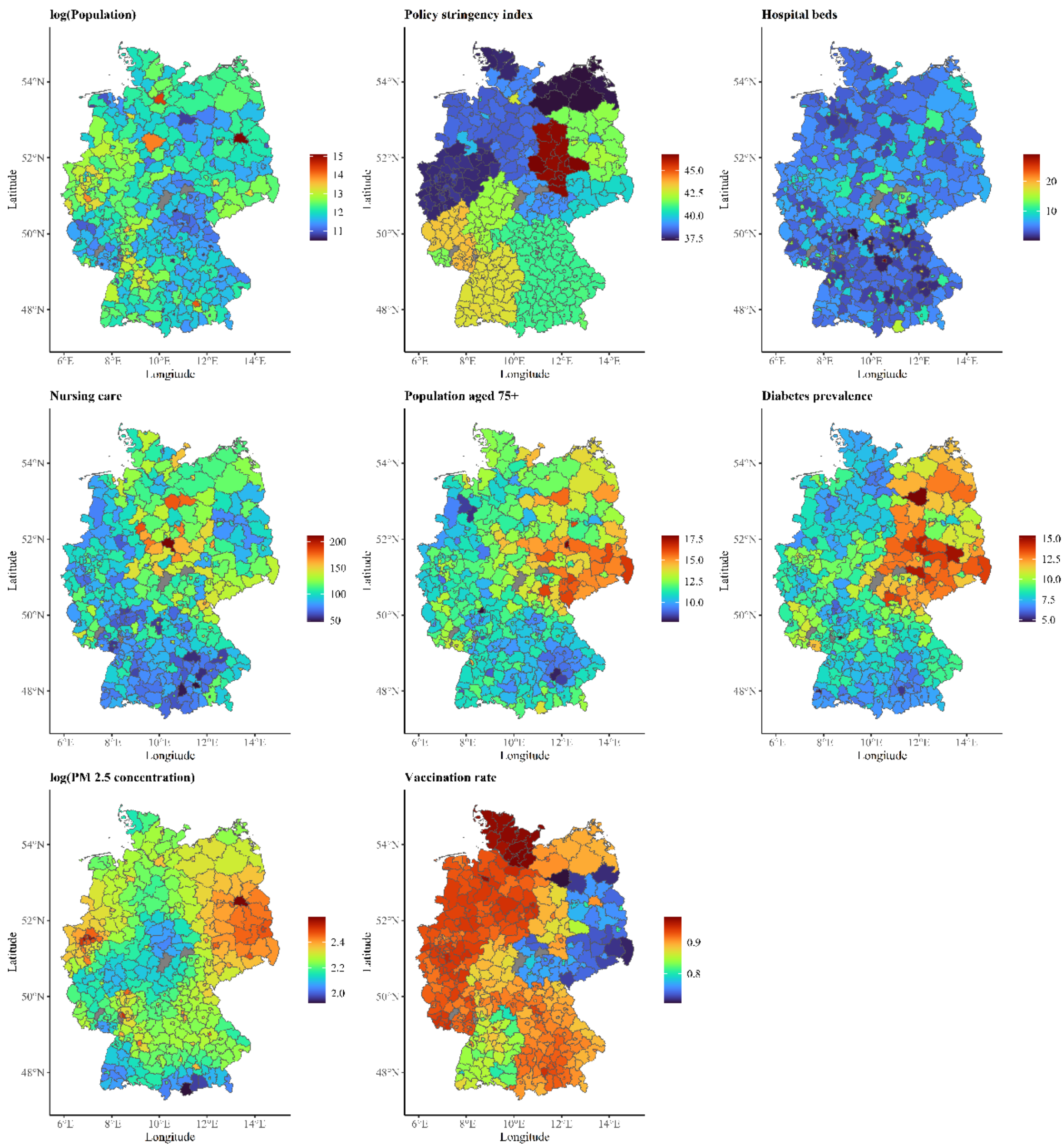
A1 Correlation matrix of regression variables

Figure 2.8: Correlation matrix of regression variables as of end of February 2022.



A2 Additional maps of regression variables

Figure 2.9: Additional maps of regression variables as of end of February 2022



A3 Regression results for the Generalised Additive Model with all variables smoothed

Table 2.5: Regression results for the Generalised Additive Model with all variables smoothed.

	Dependent variable:			
	COVID-19 mortality			
Approximate significance of smooth terms:				
	edf	Ref.df	t	p-Value
s(log(Settlement density))	1.054	1.100	2.865	0.0960
s(log(Population))	1.001	1.001	2.892	0.0899
s(Longitude, Latitude)	48.058	149.000	2.858	0.0000***
s(log(Policy stringency index))	3.658	4.331	6.755	0.0000***
s(Hospital beds)	1.807	2.290	3.045	0.0388*
s(Nursing care)	5.029	6.141	5.284	0.0000***
s(Population aged 75+)	1.002	1.004	3.556	0.0602
s(Diabetes prevalence)	1.876	2.329	21.341	0.0000***
s(log(PM 2.5)concentration)	1.001	1.002	3.522	0.0613
s(Vaccination rate)	1.292	1.485	18.853	0.0000***
Adjusted R^2	0.846			
Notes:	*Link-function: log. Model based on Formula (2.3) with all explanatory variables standardised and log-terms are shifted by x - min(x) + 1. Constant not reported. Approximate significance of smooth terms: *p < 0.05; **p < 0.01; ***p < 0.001.*			

3 Activity-based Clustering of Real Estate Stock Companies and the Varying Adjustment to Shocks

Abstract Real estate stock companies' price fluctuations have been analysed extensively, e.g. by regional focus and investment type. However, the activities on which the real estate stock corporations focus have not yet been thoroughly analysed. The intensity to which these companies participate in various real estate segments of the value chain and within different sectors may lead to differing capital market performances. This is because individual activities may be more susceptible to shocks or valued differently by capital markets. We cluster real estate stock companies in the German-speaking region (DACH-region) according to their revenue structures by two dimensions: business segments (along the value chain) and sectors (real estate asset classes). Using a novel dataset our analysis identifies four statistically robust clusters among real estate stock companies. These four clusters can be viewed as different business models. We then compute stock indices for these four clusters and analyse how these indices perform during crises. We employ Granger Causality Tests, changepoint analysis and three volatility measures to investigate these reactions. Our results underscore that these clusters react differently to the selected shocks. These insights serve as a basis for systematic benchmarking and enrich the understanding of stock market performances and resilience of different real estate business models across market phases.

Keywords: real estate stock companies; business models; stock market performance; benchmark

3.1 Introduction

Real estate stocks can contribute to mixed-asset portfolios with higher returns at comparable risk levels (Lee & Stevenson, 2005). This becomes particularly clear when different company or sector characteristics are taken into account when compiling the portfolio of property stock corporations in order to utilise the different risk/return profiles for optimal risk diversification. To measure the risk diversification attributes of real estate stock companies, they are typically grouped based on investment focus. For instance, Eichholtz et al. (1995) apply the properties in the respective portfolio, grouped by region, sectors (like residential or office properties) or investment type (low-risk or risk-oriented).

Such classifications are particularly useful for equity real estate investment trusts (REITs), as they largely share a common business model centred around the leasing of buildings. Yet, a decisive element might be overlooked when extending the definition to a broader universe of real estate stock companies. According to the FTSE EPRA NAREIT EMEA Developed market 44% of market capitalisation of the European real state stock market is constituted by non-REITS. The underrepresentation of REITs in European markets stems from various factors, including historical regulatory restrictions and investor preferences (Falkenbach et al., 2013). Structurally real estate stock companies operate similarly to REITs but given that REITs are more strictly regulated, real estate stock companies can have more differentiated business models. In fact, real estate stock companies have enjoyed greater flexibility in their business structures and investment strategies, enabling them to derive income from a wider pool of business activities. Specifically, a significant portion of the revenues might arise from other activities within the real estate value chain, including project development, construction or consulting. We argue that categorising real estate stock companies based solely on the above criteria – region, sector and investment type – is insufficient given their more complex risk-return-profiles. For example, companies centred on transactions could be more vulnerable during periods of surging interest rates than companies which focus on buy-and-hold strategies or services. These differences can be expected to be mirrored in the respective stock prices of companies (Fama, 1970).

Amit & Zott (2001) posited that a business model encapsulates a cohesive framework for exploring and exploiting new venture opportunities. Central to this concept are a value proposition and a revenue model, manifesting in specific activities across the value chain, encompassing the rental, construction or investment markets. Excluding real estate stock companies from a comprehensive analysis of the relevant investment universe would severely underestimate the overall capital flows directed towards real estate market-related companies. This is of critical significance for practitioners and especially investors.

In this paper, we address three research questions:

(1) Can real estate stock companies be meaningfully segmented into clusters based on two pivotal activity dimensions representative of a business model? These dimensions are the specific sector (asset class) a company focuses on and its position along the value chain.
(2) Do the isolated clusters exhibit discernible differences in their stock price volatilities?
(3) How did these clusters react regarding volatility and stock price performance during the recent two shocks, namely the COVID-19 pandemic and the interest rate reversal?

To answer these questions, we used the constituents of the German real estate stock market index, DIMAX, and the Austrian real estate stock market index, IATX, as of 22nd of November 2022, to define the German and Austrian real estate equity universe, respectively. We focus on the German-speaking markets, as in these countries the share of non-REITs in the overall real estate stock universe is particularly large. First, we collected categorial revenue data from the relevant companies for the time period spanning from 01st January 2016 until 31st December 2022, in order to cover similar time period before and during the two different crises of interest, the COVID-19 pandemic and the interest rate shock that followed. We cleaned the data set and ended with data for 48 companies´ revenue segmented by sector (such as residential, office, retail, hotel, logistics and special use) and by activities along the value chain (development, construction, transaction, rental, services or other). We then converted this data to metrical percentage shares of revenue in each segment and clustered the companies using Wards-algorithm. Our analysis revealed four distinct clusters: (1) Residential portfolio holders, (2) mixed portfolio holders, (3) investor/asset manager, and (4) developers. For these clusters, we calculated cluster-specific weighted performance indices. We then analysed the lead-lag structure for these clusters by using Granger-causality tests and changepoint analyses (PELT analyses). Lastly, we compute three different volatility/performance metrics and juxtaposed them across phases endogenously derived by the changepoint analyses.

To the best of our knowledge, this study is the first to cluster German real estate stock companies according to their positioning in the value chain. Additionally, it is the first study to apply PELT and Granger causality analyses in examining the lead-lag dynamics of German real estate stock companies during two most recent shocks: the COVID-19 pandemic and interest rate reversal.

This paper unfolds as follows: Chapter 2 provides a concise overview of the literature on REITs and real estate stock companies, focusing on the diversification potential of REITs, complemented by an introduction to business model literature, which forms the framework for our activity-based approach. Chapter 3 presents the formal cluster analysis, leading to the identification of four clusters. Chapter 4 unveils the performance and volatility development for the indices of these clusters, including both the Granger and the PELT analyses. Chapter 5 offers concluding remarks.

3.2 Related literature

Investing in real estate, whether directly or indirectly, can diversify risk due to distinctive risk-return characteristics (Webb et al., 1988; Lee, 2010). Similarly, managing a portfolio of diverse properties offers risk diversification benefits (Eichholtz et al., 1995). Researchers typically gauge this diversification potential by examining the stock market performance of real estate stock companies, considering factors such as geographical reach and sectoral focus, company type and organisational structure.

In the international real estate universe, real estate investment trusts (REITs) – a tax subset of real estate stock companies – have been extensively studied. Hartzell et al. (2012) observed that geographically diversifying REITs tend to have lower valuations than those with a tighter geographical focus. Studies by Hoesli & Oikarinen (2012) and Yunus et al. (2012) show that, over extended periods, listed real estate closely mirrors direct real estate performances. The tight long-run relationship between REITs and direct real estate is influenced by the underlying real estate property sectors (Chiang, 2010; Pagliari et al., 2005; Riddiough et al., 2005).

Since different real estate sectors follow distinct cycles, portfolios can derive diversification benefits. Research by Kallberg et al. (2002), Hoesli & Kustrim (2013) and Liow et al. (2018) shows how listed real estate responds to external shocks and uncertainty. Notably, these different responses can be traced back to differences across property sectors (see Hoesli & Malle, 2022).

And since investors often construct their optimal portfolios, REITs and real estate stock companies in general tend to pursue focused strategies on specific property types. While the emphasis traditionally has been on more liquid and tradable assets like offices, shopping centres and warehouses, Moss et al. (2015) highlight a growing interest in hotels, apartments and niche sectors, such as healthcare facilities and data centres. Each property sector serves a distinct industry or activity (Yavas & Yildirim, 2011), which can impact stock market performance during particular market phases.

We contend that real estate stock companies with varied activities beyond mere regional or sectoral investing focuses will manifest distinct risk-return traits potentially beneficial for portfolio allocation. An activity-centric assessment of real estate stock companies might capture their specific business model nuances more aptly than only distinguishing between regional or sectoral investment emphasis alone. This is because real estate stock companies can engage in a broader range of activities than REITs, which largely follow similar business models. Although a generally accepted definition of business models does not exist (Peric et al., 2017), common elements emerge in definitions like those of Osterwalder et al. (2005), Ovans (2015) or Drucker (1994). Wirtz et al. (2016) break this down to ‘a simplified and aggregated representation of the relevant activities of a company’. Birkner & Just (2023) summarise that a business model can be described by the value proposition of a company, her business structure, the underlying revenue model and entrepreneurial spirit. The regional and sectoral investment foci and the position on the value chain are parts of the value proposition and the revenue model, which are ultimately reflected in the revenue structure of a company, i.e. how revenue is distributed across the various sectors respectively activities along the value chain.[13]

Building on this literature, we define business models of real estate stock companies as long-term strategic activities aimed at creating value within a process.

The activities of real estate companies can differ significantly: for example, a residential portfolio company with a high proportion of rental activity differs from a company that mainly manages real estate transactions or from a company that primarily acts as a commercial real estate developer. All three may share the same regional focus and may operate in the same asset class (residential) and yet reveal different risk-return-profiles and may be subject to different cycles depend to different degrees from capital market developments. Thus, specific price movements of real estate companies with different bundles of activities are plausible. We assume, that the valuation of a real estate company differs with regards to the underlying activities. Moreover, diverse business activities might lead to different stock market valuations because of the distinct temporal structures of the underlying cash flows. Few studies examine the impact of real estate business models on company performance or volatility. Liu et al. (2018)

[13] The term “revenue structure” is used as a proxy for a breakdown of the total revenue of a company into individual segments.

pioneered a business model framework centered on the components influencing a firm's value creation capabilities. By examining the Chinese real estate construction industry along seven parameters, they distinguished five standard business models for construction companies. More recently, Geltner et al. (2022) found that companies can create value by focusing on specific segments of the real estate value chain, embracing inherent risks. These activities span various real estate sectors like residential, office, and logistics. Our study builds on this literature, broadening the analysis to encompass all real estate stock companies.

Our research connects with earlier literature on real estate capital markets. Moss & Farelly (2015) and Lee (2010) suggest that REIT performance varies across different economic and capital market phases. Capital markets typically process information rapidly. A lead-lag relationship between stocks could indicate market imperfections or fundamental differences. If the latter holds, leading stocks can be a precursor for the lagging stock (Myer & Webb, 1993; Barkham & Geltner, 1995; Okunev et al., 2000). Consistent time sequences can inform investment strategies even without a direct causal relationship. Such patterns are usually investigated using VAR or Granger Causality Tests. (Yavas & Yildirim, 2011; Hoesli et al., 2015). Glascock et al. (2000) used cointegration and VAR models to assess relationships over different periods. Granger causality was also used by Sumer and Özorhon (2020) to study REITs in a foreign exchange rate context. Drawing from these approaches, our study delves into how the performance might differ across real estate stock companies with varying business activities.

3.3 Clustering of the real estate stock companies' revenue structure across sectors and activities along the value chain

3.3.1 Methodology and data

We model real estate stock companies's activities along two defining dimensions: value chain business segments and the real estate market sectors they are active on. The companies' revenue structure across these dimensions is a proxy for the intensity of a company's value process. However, there is limited data for a stock company's revenue by underlying activities, and this might account for the gap in the literature. We try to overcome this data limitation in four steps: First, we use the constituents of the DIMAX and IATX, which include all relevant listed real estate stock companies in the German speaking market to identify the relevant investment universe and added companies only active in Germany. Second, we collect data for those 64 real estate stock companies in Austria, Germany and Switzerland, all focused on the German real estate market. This yields a data-set on revenue structures across both dimensions for a total of 48 relevant companies (71% of all companies and 88% of market capitalisation). Third, we recode the categorical data to metrical percentage values for each revenue category and conduct a cluster analysis to group real estate stock companies with comparable revenue structures. Finally, we construct indices for the identified clusters.

Data retrieval and survey

We gathered data on the company's revenue structures in terms of market segments and asset classes from their annual reports. Where this data was unavailable, we collected information through surveys and personal interviews with company representatives. We contacted all companies for data validation.[14]

We then split up the average revenue of each company from 1 January 2016 to 31 December 2022 into revenue shares across the categories, once across the business and again across sector segments. For this purpose, we categorised the revenue structures of the real estate companies into six business segments along the value chain according to German Society for Real Estate Research (2013) and Bone-Winkel et al. (2016): development, construction, transaction, rental, services, other[15] and six different sectors (residential, office, retail, hotel, logistics[16] and special use[17]). The selection of sector categories considers both the typological aspects of the German real estate market (as outlined by Ahrens, 2016) and the market volume of specific sectors as reported by industry experts (see for example CBRE, 2023).

We exclude financiers, as they primarily generate revenue from financing activities rather than directly participating in the value chain of the real estate industry.

The categorization aims to offer a comprehensive coverage of the real estate value chain. Additionally, it considers associated services that companies may provide. We argue that these segments represent the fundamental building blocks of the real estate value chain, collectively contributing to the industry's overall functioning and the available investment opportunities. Table 3.1 provides a brief definition of each business segment.

---

[14] The surveys can be provided on request.

[15] Please note that REITs solely engage in this the renting category. To the best of our knowledge four REITs exist in German speaking countries. All are encompassed in our sample.

[16] Logistics aggregates industrial, warehouse and factory use.

[17] Special purpose property aggregates parking, plots/land, as well as special uses (e.g., media production studios).

Table 3.1: Description of business segments.

Business segment	Description
Renting	Targets portfolio holders with a long-term perspective on property ownership and renting of properties.
Development	Addresses development, focusing on the acquisition of land and consequent sell or lease of the developed properties for a profit
Transaction	Addresses the activity of short-term acquisitions, as well as the classic privatization activity.
Service	Addresses the activity as an investment manager for third parties, as well as asset and property management.
Construction	Addresses the classic construction industry
Other	Addresses all other business segments that are not among the first five segments

We do not distinguish our categorisation by year. We assign categorical values for each category to quartile as follows:

(1) Main activity: 75%-100%

(2) Predominant activity: 50%-75%

(3) Balanced activity: 25%-50

(4) Marginal activity: >0-25%

In order to rule out breaks in the business models, this was queried during data validation. In addition, the analysis period was limited in order to rule out changes in the business model.

In the second step, we converted the data for revenue shares into an ordinal scale (1=100-75%; 2=75-50%; 3=50-25%; 4=25-0%). This enables the application of clustering methods with metric distance measures. Our clustering emphasises differences in revenue structures between companies. As differences increase with increased revenue polarisation, we use minimum polarisation assumptions in the coding to ensure robust results. This means that the assumed revenue shares could be more polarised in reality, but not less. Consequently, resulting fragmentation indices serve as the upper limits for the actual distribution of revenue structures among the real estate stock companies in the data set.[18]

[18] We measure polarisation across shares, with the fragmentation index F as a variant of the Hirschmann-Herfindahl-Index (HHI). The fragmentation index F is defined as: $F = 1 - HHI = 1 - \sum_i^k s_i^2$ with s_i as the share of k

Hierarchical clustering

We use Ward's algorithm (Ward, 1963) based on Euclidean distances to cluster revenue shares across business segments and sectors. Thus, we define separate groups of real estate stock companies in the German-speaking region. We also ran k-means as a fusion algorithm with no meaningful differences. The results are available upon request.

Ward's method is a widely used clustering technique that seeks to create clusters with similar revenue structures inside a cluster while maximizing the heterogeneity between clusters. Essentially, this hierarchical cluster algorithm starts with each object as a cluster of one company. Then, more companies are progressively added until all objects belong to a best suiting cluster. A plot of this merging process results in a tree-like structure, which helps to determine a meaningful number of clusters.

Index construction

After the clustering, we construct weighted indices for each cluster. The cluster analysis accounts for company heterogeneity, allowing the indices to be aggregated according to the average market capitalisation of the included companies throughout the sample period. New entrants are rebalanced to the previous index value upon introduction to prevent structural breaks during the observation period. This approach avoids structural breaks in an expanding public real estate universe.[19] Financial market data for the relevant companies is sourced from Refinitiv (Thompson Reuters, 2023).[20]

3.3.2 Results

Resulting data set

We retrieved revenue data for 64 real estate stock companies in Germany, Austria and Switzerland. We reduced the sample to a total of 48 companies with a total market capitalisation

---

different categories. The fragmentation index F has a minimum value of zero when a company's revenue is completely polarised, i.e., all company revenue is derived from a single business segment or sector, to a maximum of 0.83 when company revenue is evenly distributed across the six business segments and the six sectors considered here, respectively. See Appendix B1 for further details of the numerical coding of the revenue data.

[19] We recursively tested for the intergroup robustness of the index compositions via the methods employed in the second part of this paper.

[20] Two companies included in the sample were delisted in Q4 of 2022. An omission of these companies would have a negligible effect on the presented empirical results.

of €89bn due to data availability of reliable information on the revenue structure. This still represents approximately 87% of the initial universe's combined average market capitalisation. We included one company listed in Switzerland but active in Germany only. The descriptive statistics of the revenue data are presented in Table 3.2.

Table 3.2: Descriptive statistics of the data retrieval process.

	Category	N	Mean	Std. deviation	Median	Min	Max
Business segment	Renting	48	69%	37%	75%	0%	100%
	Transaction	48	7%	18%	0%	0%	75%
	Development	48	14%	26%	0%	0%	100%
	Service	48	10%	21%	0%	0%	100%
	Construction	48	0%	1%	0%	0%	8%
	Other	48	1%	3%	0%	0%	12%
Fragmentation index business segment		48	0.21	0.23	0.19	0.00	0.66
Sector	Residential	48	36%	41%	25%	0%	100%
	Office	48	29%	29%	27%	0%	100%
	Hotel	48	4%	8%	0%	0%	38%
	Logistics/ Factory/ Warehouse	48	5%	12%	0%	0%	62%
	Retail	48	20%	29%	12%	0%	100%
	Special purpose	48	5%	13%	0%	0%	62%
Fragmentation index sector		48	0.37	0.30	0.44	0.00	0.79

The data shows a differentiated picture when considering the two dimensions of business segments and sectors. In business segments, renting plays the most important role, constituting 69% of all the activities. Several companies solely engage in rental activities, while others focus exclusively on development. In fact, 49% of companies concentrate on one business segment, reflected by an average fragmentation index of 0.21 for business segments.

The majority of companies engages in the asset classes residential or office. Moreover, many firms adopt pure play strategies in residential, office and retail. Some 37% of companies focus on a single sector, resulting in a fragmentation index of 0.37. This suggests that diversifying activities across sectors may be simpler than diversifying along the value chain.

Clustering results

Figure 3.1 displays the clustering structure as a phylogram,[21] where individual merging steps appear as branches. Companies closer to these branches have more similar revenue structures across both dimensions than companies that are far from a branch. By means of Silhouette algorithm we choose four clusters as best fit. The four groups are highlighted and labelled.

The clusters can be described and labelled based on their represented companies as follows:

(1) **Mixed portfolio holder:** Real estate stock companies predominantly focused on '*renting*' real estate, but their investments are not solely confined to the '*residential'* sector. This group is nominally the largest, with 27 companies representing a total market capitalisation of €26.16 billion.

(2) **Residential portfolio holder:** Real estate stock companies that generate revenue exclusively from the '*renting'* business segment of real estate in the '*residential*' sector. With only eight companies, this group has the largest market capitalisation at €56.42 billion.

(3) **Investor/asset manager:** Real estate stock companies primarily involved in the '*transaction'* and '*service'* business segments. This group comprises of only five members and has a market capitalisation of €2.43 billion.

(4) **Developer:** Real estate stock companies whose revenues are mainly generated in the '*development'* business segment. This group consists of eight companies and has a market capitalisation of €1.8 billion.

[21] The phylogram represents the similarity structure of the analysed companies as generated by hierarchical clustering according to Ward's criterion. Ward's criterion selects, at each step of the agglomerative process, the pair of clusters whose merger results in the smallest increase in within-cluster variance. The terminal nodes (leaves) of the tree correspond to individual observational units, while the internal nodes represent the successive fusions of clusters. Edges connect a cluster to the superordinate cluster formed by its merger with another cluster. The resulting tree structure is hierarchical, connected, and acyclic, such that exactly one unique path exists between any two nodes. The vertical position of an internal node corresponds to the increase in within-cluster variance induced by the respective fusion and thus defines the distance threshold at which the merged clusters can be treated as a single unit. For more information on phylogram drawing, see, for example, Pavlopoulos et al. (2010).

Figure 3.1: Phylogram of the clustering result.

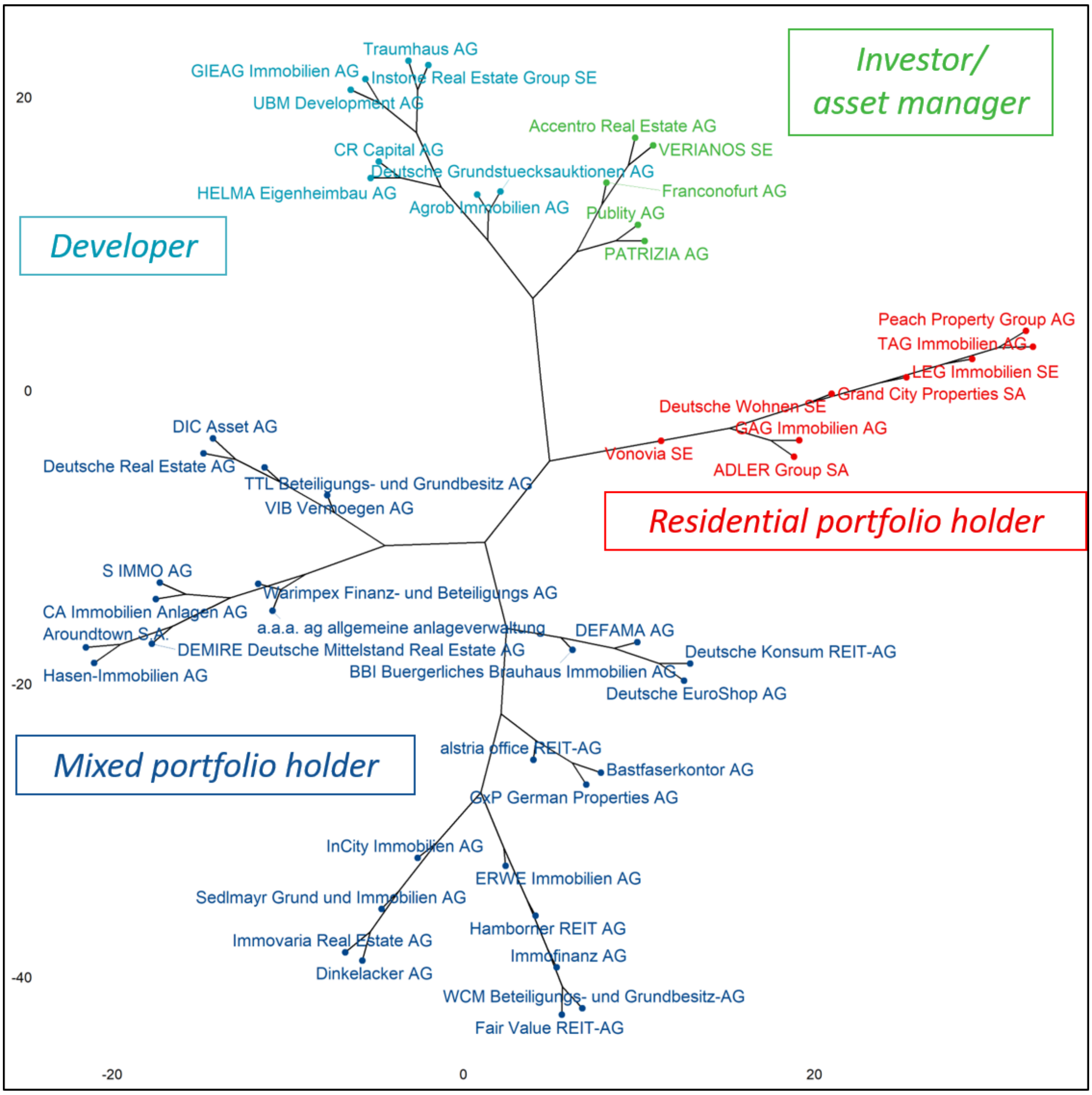
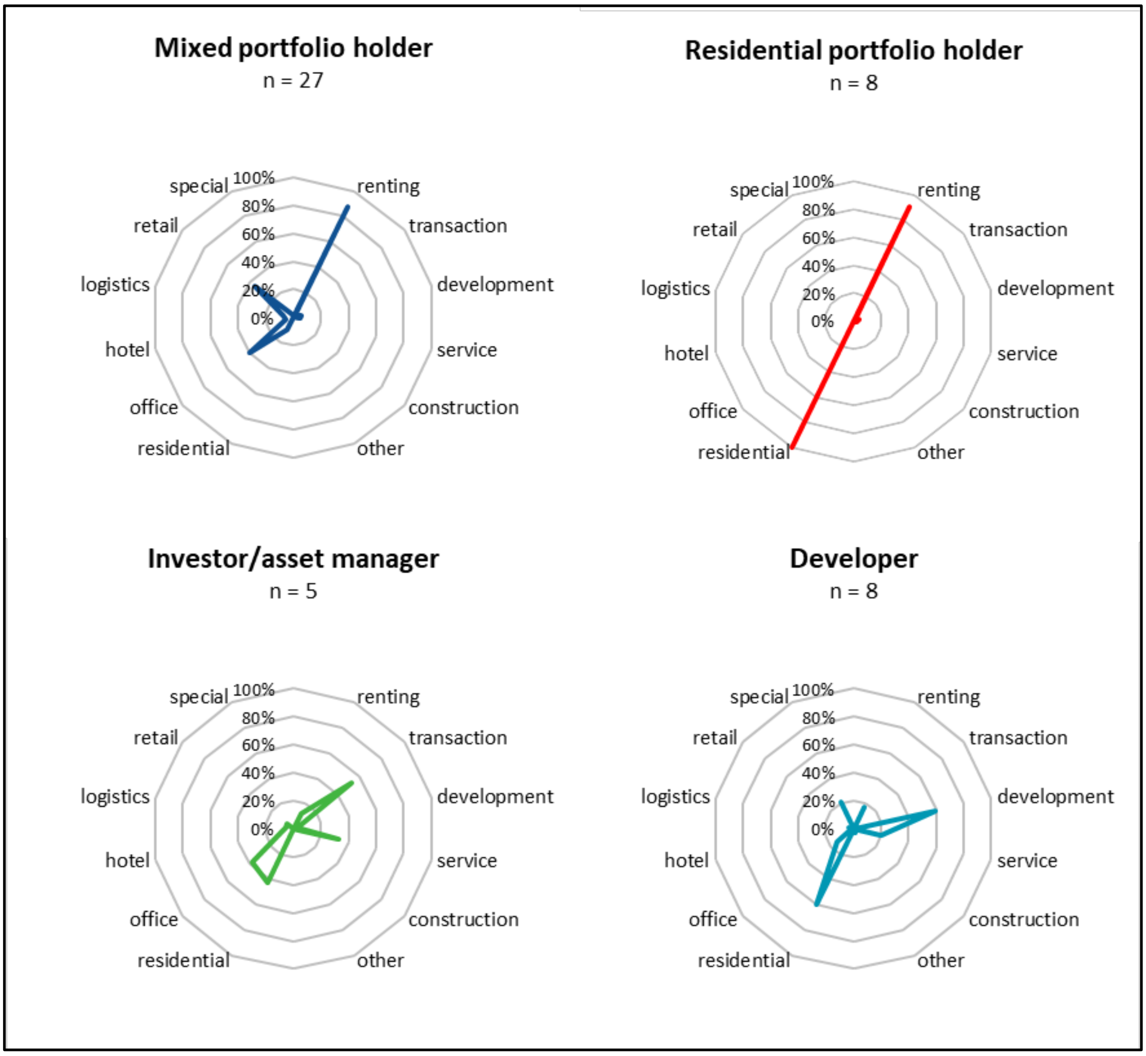


Figure 3.2 presents the descriptive statistics of the revenue structures of the business models. The sectors are depicted on the left-hand side of the web diagrams, while the activities along the value chain are shown on the right.

Figure 3.2: Revenue structures of the four business models.

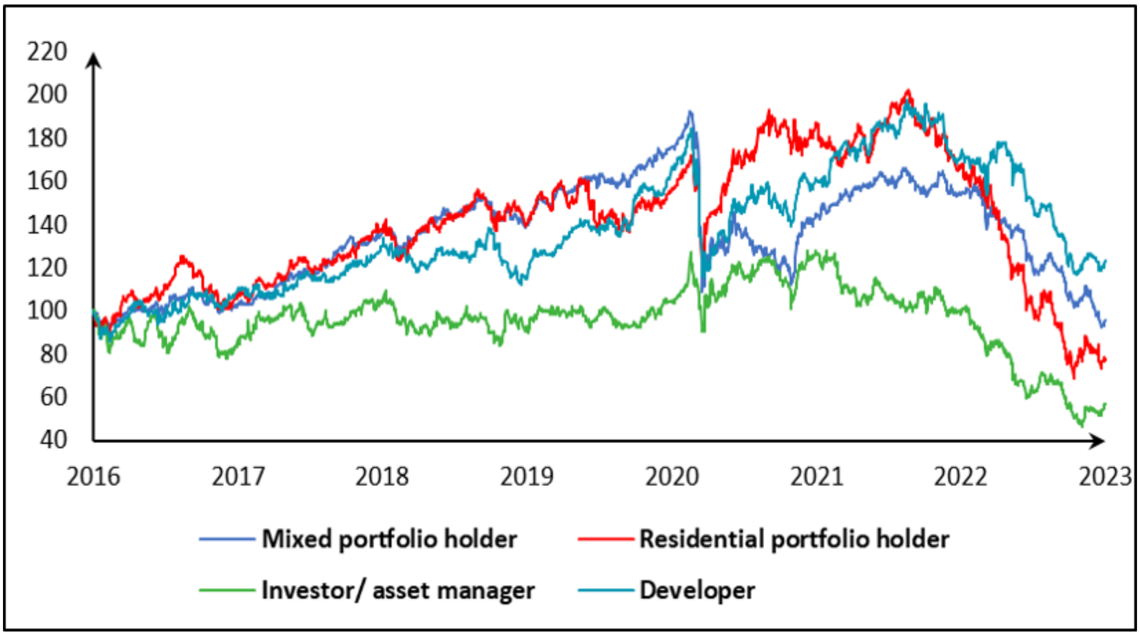


Business model indices

Based on our four sets of activities, which we consider core to underlying business models, we construct four corresponding stock market indices for our observation period.

Figure 3.3 highlights significant differences between the four indices, as well as notable shocks within the sample period. Key events, such as the beginning of lockdowns (March 2020) and the shift in the interest rate regime (interest rate reversal) are clearly seen. A comparison of the overall performance and volatility across the four groups is provided in Table 3.3.

Figure 3.3: Index performance of business models.



The two aforementioned shocks elicited different reactions from the four index groups, which had previously developed in tandem up until March 2020. This indicates different resilience levels of the underlying business models when faced with exogenous shocks. In the next chapter, we elaborate on the resilience potential across business models during the two shock periods.

Table 3.3: Descriptive statistics of the business model indices.

	Average	Std. Dev.	Median	Min	Max
Portfolio holders mixed	136.0	21.5	138.5	93.6	189.5
Portfolio holders residential	144.3	27.0	144.7	88.2	202.3
Investor/ asset manager	99.4	12.0	98.3	58.5	129.4
Developers	135.1	26.0	129.3	85.7	193.4

Note: Base 100 = 31.12.2015; Data until 31.12.2022.

3.4 Stock market performance of business models among real estate stock companies

3.4.1 Methodology

To analyse the lead-lag structures inherent within the four activity-based clusters, our analysis unfolds in two steps:

First, we analyse the relative dynamics of the four clusters. Our hypothesis posits that companies focusing on different value chain activities exhibit distinct resilience levels to shocks. Additionally, we hypothesise, that more complex business models elude rigorous analysis of capital market actors compared to simpler business models. Consequently, the capital market performance of the clusters with inherent complexities mirror this information asymmetry, potentially giving rise to a lead-lag dynamic in the real estate industry's capital market. We apply Granger Causality tests to evaluate lead-lag structures.

If a time series systematically leads a second series, then it is said to be Granger-causal for the second (Granger, 1969; Sims, 1972). This ability is typically evaluated using vector autoregressive models (VAR, see e.g. Myer & Webb, 1993, and Hamilton, 1994). To investigate the lead-lag relationships between our four indices, we estimate a VAR model based on the stationary return series.[22] The optimal lag length is determined via Akaike-Information-Criteria (AIC). We then use the Granger Causality test to check if exogenous variables jointly hold significance and thus possess the ability to "forecast" endogenous variables (Hamilton, 1994). We execute pairwise testing for each index against the others, and a comprehensive examination involving all three exogenous indices in a multivariate framework, using both Chi-square-tests and F-tests (Bao, 2022).

Secondly, we analyse the differences in volatility structures of the four indices by calculating three standard volatility measures: standard deviation (as a proxy of general risk), maximum drawdown to account for event risk (Hamelink & Hoesli, 2003; Moss et al., 2015), and the Sharpe Ratio (Sharpe, 1966) to encapsulate overall stock market performance. These volatility

[22] We construct a stationary return series from the four stock indices as the log difference of two consecutive days. To test stationarity of the return series, we conduct Augmented-Dickey-Fuller tests. Results are available upon request.

measures allow us to compare both the static volatility measures and the dynamic analysis of different market phases.[23]

We then employ the PELT (Pruned Exact Linear Times) method to identify potential volatility changepoints within the four-return series. By applying this method to the indices' volatility and return series, we isolate possible structural breaks within the time series. This technique is suitable for high-frequency data, such as daily returns series, and results in relatively stable volatility phases that effectively indicate distinct market phases. Identifying structural breaks across the four clusters could support the lead-lag structures we previously analysed through Granger Causality tests.

This method can provide accurate results with a reasonable computational effort. Starting from an ordered series $y_{1:n} = (y_1, \dots, y_n)$, a changepoint is defined as a point $\tau \in \{1, \dots, n-1\}$ at which the statistical property of interest, here the volatility between two segments of the series $\{y_1, \dots, y_\tau\}$ and $\{y_{\tau+1}, \dots, y_n\}$, differs.

To estimate these points, the sum of a segment's cost function and a penalty for too many changepoints is usually jointly minimised. The PELT algorithm builds on the optimal partitioning method of Jackson et al. (2005). This can be illustrated by the formula in (3.1):

$$\sum_{i=1}^{m+1}\left[C\left(y_{(\tau_{i-1}+1):\tau_i}\right) + \beta\right] \tag{3.1}$$

(1) is minimised (Killick et al., 2012; Killick and Eckley, 2014), where C is a cost function and β is a penalty term that is independent of the location and number of changepoints. The search for the optimal partitioning then occurs recursively, starting from the last changepoint and spanning the entire dataset. During this process, any possibilities that are not optimal are excluded to reduce computational effort, which is called pruning. Finally, the optimal partitioning of the data $F(n)$ is obtained as shown in (3.2):

$$F(n) = \min_{\tau_m}\left\{F(\tau_m) + C\left(y_{\tau_m+1}, \dots, y_n\right)\right\} \tag{3.2}$$

[23] The standard deviation and $Sharpe\ Ratio$ for the dynamic analysis are annualised. The maximum drawdown is phase-specific and cannot be annualised. See Appendix B2 for further details of the selected volatility measures.

We examine changepoints in each return series twice to identify short and more extended periods of constant volatility. Initially, we set a minimum segment length of two days (short period) and, subsequently, 60 days (extended period).

3.4.2 Results

Granger Causality

Using a VAR(4) model, the Granger Causality tests indicate that the index for mixed portfolio holders within the chosen analysis window, appears to be ahead of the other return series. This is evident in models where the tests yield statistically significant results, as displayed in Table 3.4. This signals to a potential leading role of mixed portfolio holders in the stock market performance of real estate stock companies.

Table 3.4: Granger Causality Test results based on a VAR (4) model.

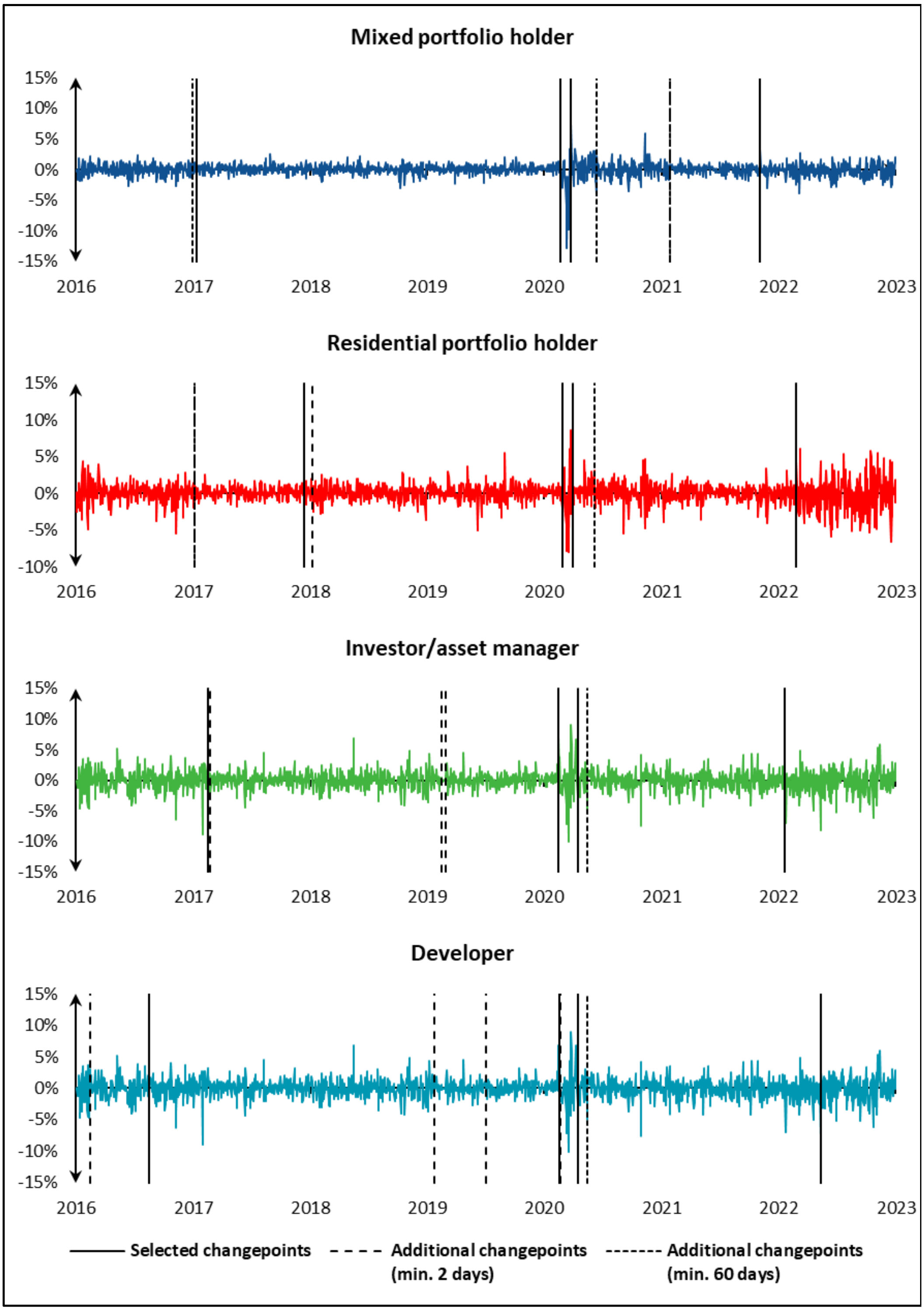
	F	df_1	df_2	*p*		χ^2	*df*	*p*	
Mixed portfolio holder ← Residential portfolio holder	2.30	4	1767	.057	.	9.18	4	.057	.
Mixed portfolio holder ← Investor/asset manager	1.44	4	1767	.219		5.75	4	.218	
Mixed portfolio holder ← Developer	1.80	4	1767	.126		7.20	4	.126	
Mixed portfolio holder ← ALL	1.69	12	1767	.062	.	20.33	12	.061	.
Residential portfolio holder ← Mixed portfolio holder	2.40	4	1767	.048	*	9.61	4	.047	*
Residential portfolio holder ← Investor/asset manager	2.44	4	1767	.045	*	9.76	4	.045	*
Residential portfolio holder ← Developer	1.01	4	1767	.400		4.05	4	.400	
Residential portfolio holder ← ALL	1.85	12	1767	.037	*	22.15	12	.036	*
Investor/asset manager ←Mixed portfolio holder	3.01	4	1767	.017	*	12.02	4	.017	*
Investor/asset manager ← Residential portfolio holder	1.31	4	1767	.265		5.23	4	.264	
Investor/asset manager ← Developer	0.90	4	1767	.464		3.60	4	.463	
Investor/asset manager ← ALL	2.84	12	1767	<.001	***	34.09	12	<.001	***
Developer ← Mixed portfolio holder	8.76	4	1767	<.001	***	35.03	4	<.001	***
Developer ← Residential portfolio holder	0.50	4	1767	.737		1.99	4	.737	
Developer ← Investor/asset manager	1.27	4	1767	.281		5.06	4	.281	
Developer ← ALL	5.74	12	1767	<.001	***	68.94	12	<.001	***

Note: * p > 0.05. ** p > 0.01. *** p>.001.

Changepoint analysis

In Figure 3.4, we show the daily log returns, together with the derived changepoints for the two selected lag lengths resulting from the PELT algorithm.

Figure 3.4: Index returns and estimated changepoints.



From the comprehensive set of changepoints identified, we discern meaningful and robust changepoints that enable the demarcation of comparable market phases for the different return series during the analysed period.[24]

We only use those computed changepoints when both minimum segment lengths (two days and 60 days, respectively) yield the same changepoint day. If they yield close but different changepoints, we opt for the longer minimum segment length. Based on these assumptions, we identify four distinct and mostly homogenous starting points to general market phases: (1) reference period, (2) COVID-19 shock, (3) pandemic phase and (4) interest rate reversal, presented in Table 3.5. Note that, although the phases are similar for the four clusters, their exact starting and ending points differ. Clusters seem to react differently to shocks. The results of the changepoint analyses also largely confirm the results of the Granger Causality test, spotlighting the mixed portfolio holder as the leading index.

Our analysis identifies a long period of constant volatility preceding the COVID-19 shock across all four indices. This period coincides with the phase where the indices exhibit parallel development, which we previously identified in Figure 3.3. We designated this as the reference period to the other market phases. The succeeding phase was marked by the COVID-19 shock.

All indices show a phase of constant volatility, which concludes towards the end of March 2020 for the portfolio holders and by mid-April for the other two indices. After this period of high volatility characterised by pandemic-induced uncertainties, the stock market stabilised. The shift in interest rates affected the volatility of mixed portfolio holders in 2021, while the other three clusters did not react before 2022, with developers being the last to be impacted in May of 2022. The latter part of 2021 witnessed a surge in markets uncertainty, with signs of inflationary pressures coming around the end of 2021. This appears to have predominantly affected mixed portfolio holders. The risk stemming from complexity seems to have been valued more important than even the developer risk. Residential portfolio holders, on the other hand, registered a volatility shift on the day of the Ukraine assault in February 2022 (see also Farelly and Moss (2023), who selected similar changepoints by hand).

[24] We checked the estimated changepoints to determine whether they coincide with changes in the number of constituents of the indices. We do not find any coincidences.

Table 3.5: Changepoints of individual market phases (PELT method).

	Reference period	COVID-19 shock	Pandemic phase	Interest rate reversal
Mixed portfolio holders	12.01.2017 (2)	22.02.2020 (3)	25.03.2020 (1)	04.11.2021 (1)
Residential portfolio holders	15.12.2017 (4)	27.02.2020 (4)	31.03.2020 (2)	24.02.2022 (3)
Investor/ asset manager	18.02.2017 (3)	14.02.2020 (1)	16.04.2020 (3)	22.01.2022 (2)
Developers	17.08.2016 (1)	19.02.2020 (2)	16.04.2020 (3)	14.05.2022 (4)

Notes: Chronological order of dates presented in brackets (.).

The results of Tables 3.4 and 3.5 reveal that mixed portfolio holders did not lead the others in inflection points until the COVID-19 shock. However, after this phase, the mixed-portfolio holders' cluster emerges as the first index/business model to show a change in volatility patterns, particularly during the stabilising period of the ongoing pandemic. These relationships suggest that portfolio management complements traditional regional or sector-specific portfolio strategies.

Index performances

We consequently use these market phases to compare the phase-specific performance of the business model indices in the next step, as depicted in Figure 3.5.

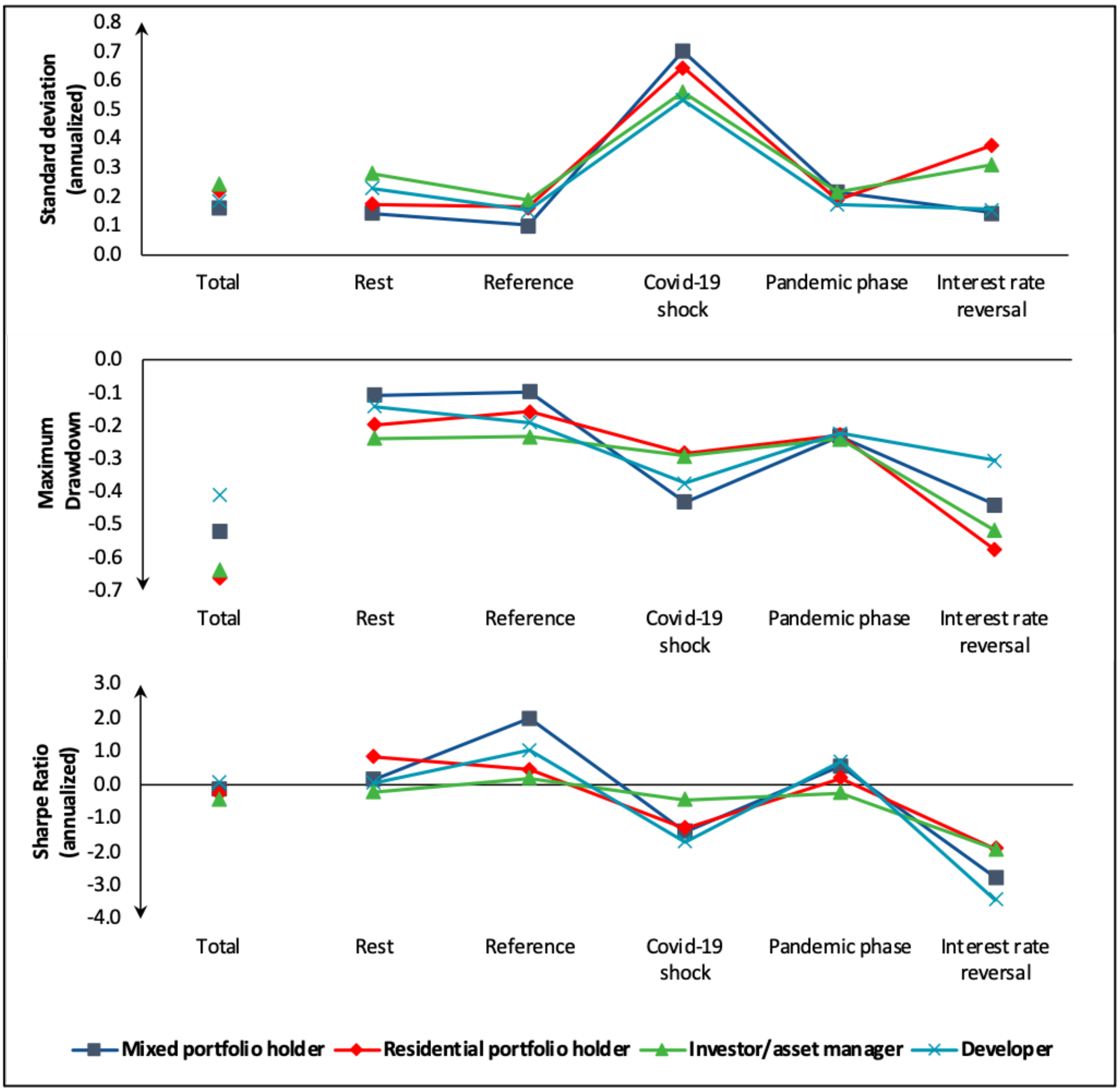
We can deduce two findings from Figure 3.5:[25] First, the mixed portfolio holders' index displays significantly less volatility than the residential portfolio holders and investors/asset managers indices over the total observation period. This contradicts the notion that residential real estate (in stock companies) offers less volatility and more security than commercial real estate. Regarding the presented volatility measures, mixed portfolio holders have shown a commendable performance. Developers perform well throughout the entire sample period for all volatility measures, suggesting their capacity to deliver risk-adjusted positive returns, as

[25] Please note that the difference in standard deviation is due to this table reporting the return volatility dynamics, whereas Table 3.3 reports on index values.

indicated by a positive Sharpe Ratio. These results were not expected due to the developments towards the end of the observation period (see Figure 3.3).

Second, investors/asset managers and residential portfolio holders display pronounced maximum drawdowns compared to mixed portfolio holders and developers. This potentially indicates a higher level of resilience inherent in the business models of the latter groups. On the other hand, residential portfolio holders exhibited greater resilience, experiencing only a 28% decline, mirroring the performance of the investors/asset managers. This underscores their relative stability in comparison to other business models.

Figure 3.5: Performance measures of business model indices in different market phases.



Turning the perspective to the performance of the individual business models across the different market phases, it is essential to note that while these phases occur within the same timeframe, they possess individual starting and endpoints. Compared to the reference phase, the COVID-19 shock triggered a substantial but temporarily limited spike in market volatility. Mixed portfolio holders and developers reacted most strongly and with the steepest maximum drawdowns of -43% and -37%, respectively. Residential portfolio holders exhibited greater resilience with only a 28% decline, similar to investors/asset managers. Further underscoring the relative stability of these business models in comparison to the others.

Market volatility subsided during the pandemic phase, as evidenced by the standard deviation and maximum drawdown metrics aligning across all business models. Still, volatility remained elevated compared to pre-pandemic levels. The pronounced maximum drawdowns, nearing -23%, likely reflect persistently heightened uncertainties. Notably, the returns of the business model indices largely moved in tandem up until the end of the pandemic phase. However, a stark asymmetric performance became evident during the interest rate reversal phase. Mixed portfolio holders and developers saw their volatility levels revert close to pre-pandemic levels. In contrast, residential portfolio holders and investors/asset managers experienced a negative development of both volatility measures. In the case of maximum drawdown, the deterioration surpassed the level of the COVID-19 shock and is partially due to rising valuations at the end of the pandemic phase.

With regards to the Sharpe Ratio, the mixed portfolio holders' index eclipses the other three during the reference period, and displays a negative risk-adjusted return during the COVID-19 shock. The mixed portfolio holders were only outperformed by developers and then overtaken in terms of negative development during the interest rate reversal. In contrast, residential portfolio holders and investors/asset managers exhibit a much smaller deviation of the Sharpe ratio across the different market phases.

Ultimately, while the pandemic posed a severe shock for real estate companies, it did not fundamentally alter market dynamics or volatility. The responses across the four business models were relatively uniform during this phase. However, reactions to the interest rate shift exhibited variations among the business models. Specifically, companies focused on residential portfolios and those with an investment-centric approach show increased volatility compared to diversified portfolio holders and developers. These findings show very clearly that using an activity-based business model perspective can enhance the analysis of stock market performance for Real estate stock companies.

3.5 Conclusion

In this study, we answered three research questions. First, we categorised the real estate stock companies in the German-speaking region (primarily active in Germany) into activity clusters based on their core business models by clustering revenues in two dimensions: the activities across sectors (asset classes) and activities along the value chain. Through a formal cluster procedure, we successfully identified four distinct activity-based company clusters: mixed portfolio holders, residential portfolio holders, investors/asset managers and developers. This diverges from the traditional perspective on REITs which are only encompassed among the former two business models and are less relevant to the German speaking region where there are only four active REITs (ZIA, 2023).[26]

Secondly, we observed measurable differences in both the performance and the volatility of the stock price indices, which were created based on the four identified clusters. Intriguingly, the clusters with the lowest volatility also had the lowest performance. Moreover, it was unexpected to find that residential portfolio holders showed the highest volatility.

Third, using Granger Causality tests and a series of changepoint analyses (PELT), our findings highlight that during the two distinct shocks within the study period, mixed portfolio holders responded most quickly to the change in information and the accompanying heightened uncertainty. This rapid response could be due to the stock market's punishing complexity quickly. Notably, this early entry into a new phase of increased volatility did not coincide with a particularly pronounced drop in performance.

Investors can conclude three insights from our analyses. First, there is merit in distinguishing business models more precisely beyond mere sectors (or regions). Future analyses could extend the activity model presented here by integrating other aspects of the business model literature, such as resources, networks, other structural features and value commitment, according to typical business model canvases.

Second, these insights could be used for a trading strategy based on a potential sequence of share price dynamics in response to shocks. Crucially, any such analysis should not be limited to REITs, allowing for a more expansive exploration of the business models. Thirdly, results from the PELT analysis show that while there have been longer phases of similar volatilities, these volatilities differ based on the identified clusters. Such insights can be invaluable for

[26] Deutsche Industrie REIT-AG was delisted in January 2022.

formulating trading strategies, too. The presented results indicate that the intervals between the different phases are inconsistent.

Future research might build on our research by stressing some of our modelling assumptions, e.g. regarding the Granger Causality Test, the selected lag length of the VAR model is not arbitrary but may differ. Similarly, the minimum segment length of the PELT algorithm could be varied.

This analysis raises several potential follow-up questions that remain outside the scope of this study. Firstly, the question arises as to whether the regional component has a significant influence on the differentiation of the business models and whether the clusters identified can provide additional information to the traditional stock allocations. Furthermore, while this paper focused on volatility dynamics, exploring other critical indicators, such as individual items on the balance sheet or income statement like operational efficiency, could provide insights. This approach has been applied to REITs by Beracha et al. (2019). Similarly, the overarching macroeconomic conditions that govern the structural changes in the broader economy, ultimately leading to different market phases, warrant a closer look. Studies on macroeconomic determinants reveal consistent patterns in commercial property market returns. These patterns are often driven by macroeconomic factors such as interest rates, bond yields, capital markets, credit spreads, inflation and unemployment (Macregor & Schwann, 2003; Conner & Liang, 2005; Fei et al., 2010). Stevenson (2016), in his analysis of the factors influencing correlation in global listed real estate markets, suggests financial and macroeconomic risk factors to affect the degree of co-movement in real estate markets worldwide.

Finally, the analysed timeframe is short, focusing primarily on German-speaking countries, raising the question whether an expanded time horizon would identify other clusters or alternative patterns in volatility or sequences. The proposed approach may be scalable to European markets as 44% of market capitalisation of the FTSE EPRA NAREIT EMEA Developed market is constituted by non-REITS. Transitioning this study to other regions, such as the USA or other European countries, would be a logical next step, since it can be assumed that different structures can be found in other countries according to business model.

These questions cannot be answered in this study, which was primarily concerned with the fundamental inclusion of the business model in a comprehensive framework, their composition and stock market performance.

Funding details

The Accentro Real Estate AG partly supported this research.

Disclosure statement

The authors state that there are no competing interests to declare.

Data availability statement

The data supporting the findings of this study are available from the corresponding author, K. F. K., upon reasonable request.

References

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, *22*(6–7), 493–520.

Arens, J. (2016). Differentiation according to property types. In K. W. Schulte, S. Bone-Winkel & W. Schäfers (Eds.), *Business fundamentals* (pp. 84–106). Walter de Gruyter GmbH & Co. KG.

Bao, H. W. S. (2022). bruceR: Broadly useful convenient and efficient R functions. *R package version 0.8. x*.

Barkham, R., & Geltner, D. (1995). Price discovery in American and British property markets. *Real Estate Economics*, *23*(1), 21–44.

Beracha, E., Feng, Z., & Hardin, W. G. (2019). REIT operational efficiency: Performance, risk, and return. *Journal of Real Estate Finance and Economics*, *58*(3), 408–437.

CBRE. (2023). German real estate investment market. [online] https://news.cbre.de/en/german-real-estate-investment-market-transaction-volume-still-hugging-low-levels-despite-much-stronger-momentum-in-the-third-quarter/#:~:text=Multi%2Dfamily%20housing%20(upward%20of,respectively%20contributing%20to%20the%20strong [13 December 2023]

Cherny, V., & Obłój, J. (2013). Portfolio optimisation under non-linear drawdown constraints in a semimartingale financial model. *Finance and Stochastics*, *17*(4), 771–800.

Chiang, K. C. H. (2010). On the comovement of REIT prices. *Journal of Real Estate Research*, *32*(2), 187–200.

Conner, P., & Liang, Y. (2005). The complex interaction between real estate cap rates and interest rates. *Briefings in Real Estate Finance*, *4*(3), 185–197.

Cvitanic, J., & Karatzas, I. (1997). On portfolio optimization under drawdown constraints. *IMA lecture notes in mathematics and applications*, *65*.

Drucker, P. F. (2006). What executives should remember. Harvard Business Review, *84*(2).

Eichholtz, P. M. A., Hoesli, M., MacGregor, B. D., & Nanthakumaran, N. (1995). Real estate portfolio diversification by property type and region. *Journal of Property Finance*, *6*(3), 39–59.

Falkenbach, H., Niskanen, J., & Kiehelä, S. (2013). Development and performance of the public real estate investment sector in Finland. *International Journal of Strategic Property Management*, *17*(3), 233–247.

Fama, E. F. (1970). Efficient Capital Markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383–417.

Farrelly, K., & Moss, A. (2023). *Identifying a set of general equity style and factor guidelines for European listed real estate. Working paper ERES 2023*.

Fei, P., Ding, L., & Deng, Y. (2010). Correlation and volatility dynamics in REIT returns: Performance and portfolio considerations. *Journal of Portfolio Management*, *36*(2), 113–125.

Geltner, D., Kumar, A., & Van de Minne, A. M. (2023). Is there super-normal profit in real estate development? *Journal of Real Estate Research*, *45*(2), 160–187.

Glascock, J. L., Lu, C., & So, R. W. (2000). Further evidence on the integration of REIT, bond, and stock returns. *Journal of Real Estate Finance and Economics*, *20*(2), 177–194.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Hamelink, F., & Hoesli, M. E. R. (2003). The maximum drawdown as a risk measure: The role of real estate in the optimal portfolio revisited. FAME Research Paper no. 87. *SSRN Electronic Journal*.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hartzell, J. C., Sun, L., & Titman, S. (2014). Institutional investors as monitors of corporate diversification decisions: Evidence from real estate investment trusts. *Journal of Corporate Finance*, *25*(C), 61–72.

Hoesli, M., & Malle, R. (2022). Commercial real estate prices and COVID-19. *Journal of European Real Estate Research*, *15*(2), 295–306.

Hoesli, M., & Oikarinen, E. (2012). Are REITs real estate? Evidence from international sector level data. *Journal of International Money and Finance*, *31*(7), 1823–1850.

Hoesli, M., & Reka, K. (2013). Volatility spillovers, comovements and contagion in securitized real estate markets. *Journal of Real Estate Finance and Economics*, *47*(1), 1–35.

Hoesli, M., Oikarinen, E., & Serrano, C. (2015). Do public real estate returns really lead private returns? *Journal of Portfolio Management*, *41*(6), 105–117.

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., San, P., Tan, L., & Tun Tao Tsai. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, *12*(2), 105–108.

Kallberg, J. G., Liu, C. H., & Pasquariello, P. (2002). Regime shifts in Asian equity and real estate markets. *Real Estate Economics*, *30*(2), 263–291.

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.

Killick, R., & Eckley, I. A. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software*, *58*(3), 1–19.

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598.

Lee, S. (2010). The changing benefit of REITs to the mixed-asset portfolio. *Journal of Real Estate Portfolio Management*, *16*(3), 201–215.

Lee, S., & Stevenson, S. (2005). The case for REITs in the mixed-asset portfolio in the short and long run. *Journal of Real Estate Portfolio Management*, *11*(1), 55–80.

Liow, K. H., Liao, W.-C., & Huang, Y. (2018). Dynamics of international spillovers and interaction: Evidence from financial market stress and economic policy uncertainty. *Economic Modelling*, *68*(C), 96–116.

Liu, G., Li, K., Shrestha, A., Martek, I., & Zhou, Y. (2018). Strategic business model typologies evident in the Chinese real-estate industry. *International Journal of Strategic Property Management*, *22*(6), 501–515.

Macgregor, B. D., & Schwann, G. M. (2003). Common features in UK commercial real estate returns. *Journal of Property Research*, *20*(1), 23–48.

Moss, A., & Farrelly, K. (2015). The performance of a blended real estate portfolio for UK DC investors. *Journal of Property Investment and Finance*, *33*(2), 156–168.

Moss, A., Clare, A., Thomas, S., & Seaton, J. (2015). Trend following and momentum strategies for global REITs. *Journal of Real Estate Portfolio Management*, *21*(1), 21–31.

Neil Myer, F. C. N., & Webb, J. R. (1993). Return properties of equity REITs, common stocks, and commercial real estate: A comparison. *Journal of Real Estate Research*, *8*(1), 87–106.

Okunev, J., Wilson, P., & Zurbruegg, R. (2000). The causal relationship between real estate and stock markets. *Journal of Real Estate Finance and Economics*, *21*(3), 251–261.

Osterwalder, A., Pigneur, Y., & Tucci, C. L. (2005). Clarifying business models: Origins, present, and future of the concept. *Communications of the Association for Information Systems*,

Ovans, A. (2015). What is a business model? *Harvard Business Review*. 23, January.

Pagliari, J. L., Scherer, K. A., & Monopoli, R. T. (2005). Public versus private real estate equities: A more refined, long-term comparison. *Real Estate Economics*, *33*(1), 147–187.

Pavlopoulos, G. A., Soldatos, T. G., Barbosa-Silva, A., & Schneider, R. (2010). A reference guide for tree analysis and visualization. *BioData mining*, *3*(1), 1.

Peric, M., Durkin, J., & Vitezic, V. (2017). The constructs of a business model redefined: A half-century journey. *SAGE Open*, *7*(3).

REITG (REIT-Gesetz) (2007). [online] https://www.gesetze-im-internet.de/reitg/BJNR091410007.html [21 January 2026]

Riddiough, T. J., Moriarty, M., & Yeatman, P. J. (2005). Privately versus publicly held asset investment performance. *Real Estate Economics*, *33*(1), 121–146.

Russell, F. (2023). FTSE EPRA Nareit developed REITs and non-REITs indices. [online] https://research.ftserussell.com/Analytics/Factsheets/Home/DownloadSingleIssue?issueName=ERGLANDENRG&IsManual=false [31 December 2023]

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, *39*(S1), 119–138.

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

Stevenson, S. (2016). Macroeconomic and financial determinants of comovement across global real estate security markets. *Journal of Real Estate Research*, *38*(4), 595–624.

Sumer, L., & Özorhon, B. (2020). The exchange rate effect on housing price index and reit index return rates. *Finansal Araştırmalar ve Çalışmalar Dergisi*, *12*(22), 249–266.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Webb, J. R., Curcio, R. J., & Rubens, J. H. (1988). Diversification gains from including real estate in mixed-asset portfolios. *Decision Sciences*, *19*(2), 434–452.

Wirtz, B. W., Pistoia, A., Ullrich, S., & Göttel, V. (2016). Business models: Origin, development and future research perspectives. *Long Range Planning*, *49*(1), 36–54.

Yavas, A., & Yildirim, Y. (2011). Price discovery in real estate markets: A dynamic analysis. *Journal of Real Estate Finance and Economics*, *42*(1), 1–29.

Yunus, N., Hansz, J. A., & Kennedy, P. J. (2012). Dynamic interactions between private and public real estate markets: Some international evidence. *Journal of Real Estate Finance and Economics*, *45*(4), 1021–1040.

ZIA. (2023). REITs. [online] https://zia-deutschland.de/project/reits/ [21 January 2026]

B. Appendix

B1 Details of the numerical coding of the revenue data

We used four assumptions to recode the categorical values into numerical percentage values:

(1) The revenue shares across business segments and sectors must each add up to 100%.
(2) The fragmentation index for each dimension should be maximised (minimum polarization).
(3) Weighting is retained where possible, and category averages used where appropriate.
(4) Finally, differences that cannot be resolved are equated.

For example, if (0;0;0;2;3;4) is specified as the resulting revenue structure in one dimension, the two is first assigned 50% of the total. To maintain the weighting between the categorical values of three and four, category averages are used; therefore, 37.5% is assigned to three and 12.5% to four last (0;0;0;50%;37,5%;12,5%).

The result of the data cleansing and individual analysis is presented in Table 3.6.

Table 3.6: Quality of classification overview.

Market capitalization	Number	Share of companies	Validation type	Included
50.081.495.316€	23	35,94%	Company	Yes
12.188.398.894€	5	7,81%	Expert	Yes
18.509.126.747€	12	18,75%	Clear weighting possible	Yes
6.086.582.718€	8	12,50%	Weight based off market assumptions	Yes
915.491.642€	13	20,31%	No classification / insolvency	No
11.635.250.859€	3	4,69%	Construction companies	No

B2 Selected volatility measures

The volatility measures allow us to achieve a comparison of both the static volatility measures, as well as a dynamic analysis of different market phases. Details of the volatility measures are presented in Table 3.7.

Table 3.7: Selected volatility measures.

Description	Formula
As the analysis of the information content of public real estate returns has shifted from the mean analysis of returns to the associated variances, general risk levels are commonly measured through the volatility as expressed by the standard deviation (σ_p). Where n is the total number of values y_i and μ is the mean.	$\sigma_p = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \mu)^2}{n}}$
We employ the maximum drawdown as an asymmetric measure of risk regarding the event risk over a specific time period, we. This measure, focused on capital preservation, a key concern of investors, assesses the relative riskiness of investment strategies (Cvitanic & Karatzas, 1995; Hamelink & Hoesli, 2003; Cherny & Obloj, 2011; Moss et al., 2015). Where $X(.)$ signifies the stock market price at time t and $X(\tau)$ the highest price up until that point in time within the time frame τ. In the context of this paper the maximum drawdown serves as a proxy of resilience to an event risk over a specified time period.	$DD(T) = \max_{\tau \in (0,T)} D(\tau)$ $= \max_{\tau \in (0,T)} D(\tau) \left[\max_{t \in (0,\tau)} X(t) - X(\tau) \right]$
Sharpe Ratio (Sharpe, 1966) provides us with an additional component in the analysis by allowing for the incorporation of return for each additional unit of risk. Where R_p is the return over the reference period, R_f is the risk-free rate proxied by the return of the 10yr German government bond, σ_p is the standard deviation.	$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$

4 Anpassung qualifizierter Mietspiegel: methodische Hinweise zur statistischen Absicherung stichprobenbasierter Verfahren

Abstract Qualifizierte Mietspiegel müssen zwei Jahre nach Veröffentlichung an die Marktentwicklung angepasst werden. Aber weder im Gesetz noch in der Mietspiegelverordnung wird ausgeführt, wie eine Anpassung eines Mietspiegels nach anerkannten wissenschaftlichen Grundsätzen genau umgesetzt werden soll. Der vorliegende Beitrag soll diese Lücke schließen. Hierbei sollen unterschiedliche methodische Ansätze zur Anpassung von qualifizierten Mietspiegeln mit einem besonderen Fokus auf stichprobenbasierte Verfahren dargestellt und diskutiert werden. Grundsätzlich gelten diese gleichermaßen für Mietspiegel, die nach der Tabellen- oder nach der Regressionsanalyse erstellt wurden. Aus stichprobentheoretischen Gründen empfehlen wir die Verwendung eines Querschnittsdatensatzes. Außerdem zeigen wir einen neuen Ansatz auf, der mit wenigen normativen Setzungen (zu Signifikanzniveau, Effektstärke und Teststärke) die Berechnung des benötigten Stichprobenumfangs erlaubt.

Keywords: Mietspiegel, Anpassung, Stichprobengröße

4.1 Einführung

Ein qualifizierter Mietspiegel ist nach § 558d Abs. 2 BGB im Abstand von zwei Jahren an die Marktentwicklung anzupassen. Dabei darf der Stichtag der Anpassung[27] nicht mehr als zwei Jahre nach dem Stichtag der Erhebung liegen. Zudem muss auch die Veröffentlichung der Anpassung spätestens zwei Jahre nach Veröffentlichung des neu erstellten Mietspiegels erfolgen. Eine Anpassung kann nur ein einziges Mal erfolgen. Nach Ablauf von vier Jahren seit der letzten Neuerstellung muss der Mietspiegel vollständig neu erstellt werden.[28]

Die Anpassung kann anhand der Entwicklung des vom Statistischen Bundesamt ermittelten Preisindexes für die Lebenshaltung aller privaten Haushalte in Deutschland (im Folgenden: Verbraucherpreisindex) oder anhand einer Stichprobe erfolgen. Weitere Anforderungen an die Anpassung inklusive deren Veröffentlichung und Dokumentation werden in der Mietspiegelverordnung (MsV) in § 22 (Anpassung mittels Index) und § 23 (Anpassung mittels Stichprobe) geregelt. Eine Stichprobe kann dabei den gleichen Umfang wie bei der Neuerstellung des Mietspiegels haben, nach § 23 Abs. 2 Satz 2 MsV ist jedoch auch ein davon abweichender Stichprobenumfang, also auch eine kleinere Stichprobe zulässig.

Nach § 558d Abs. 1 BGB sind qualifizierte Mietspiegel nach anerkannten wissenschaftlichen Grundsätzen zu erstellen. In Abs. 2 ist ein gleichlautendes Erfordernis für die Anpassung nicht enthalten. Ungeachtet der möglichen rechtlichen Anforderungen an die Anpassung eines qualifizierten Mietspiegels orientieren sich die nachfolgenden Ausführungen an einem Standard, der anerkannten wissenschaftlichen Grundsätzen entspricht.

Weder im Gesetz noch in der Mietspiegelverordnung wird ausgeführt, wie eine Anpassung nach anerkannten wissenschaftlichen Grundsätzen genau umgesetzt werden soll. Auch in den vom Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) herausgegebenen „Handlungsempfehlungen zur Erstellung von Mietspiegeln“ fehlen detaillierte Angaben zur Anpassung von Mietspiegeln.[29] Der vorliegende Beitrag soll diese Lücke schließen. Hierbei sollen unterschiedliche methodische Ansätze zur Anpassung von qualifizierten Mietspiegeln mit einem besonderen Fokus auf stichprobenbasierte Verfahren dargestellt und diskutiert

[27] In der Praxis hat sich auch der Begriff „Fortschreibung“ etabliert. In diesem Aufsatz wird entsprechend der Formulierung in § 558d Abs. 2 BGB nur der Begriff der "Anpassung“ verwendet.

[28] Grundsätzlich besteht auch die Möglichkeit, anstelle einer Anpassung den Mietspiegel alle zwei Jahre neu zu erstellen.

[29] Vgl. Sebastian et al. (2024, S. 68 f).

werden. Grundsätzlich gelten diese gleichermaßen für Mietspiegel, die nach der Tabellen- oder nach der Regressionsanalyse erstellt wurden.

4.2 Möglichkeiten der Anpassung und ihre Bewertung

Die grundlegende Idee einer Anpassung besteht darin, die Mietwerte M des Mietspiegels mit Hilfe eines Preisindexes P als Anpassungsfaktor an die Entwicklung des Mietmarktes anzupassen:

$$M_{neu} = M_{alt}P \tag{4.1}$$

mit

$$P = \frac{P_{neu}}{P_{alt}} = 1 + \frac{P_{neu} - P_{alt}}{P_{alt}} \tag{4.2}$$

Bei einem Tabellenmietspiegel müssen dabei alle Werte in der Tabelle entsprechend angepasst werden. Bei einem Regressionsmietspiegel richtet sich die konkrete Form der Anpassung dagegen nach dem gewählten Regressionsmodell, konkret ob es sich um ein additives oder um ein multiplikatives Modell handelt. Bei additiven Modellen mit absoluten Zu- und Abschlägen müssen sämtliche Zu- und Abschläge und damit sämtliche Regressionskoeffizienten angepasst werden. Bei multiplikativen Modellen mit prozentualen Zu- und Abschlägen ist nur die Basis anzupassen. Bewertungshilfen zur Spanneneinordnung sind analog anzupassen.

Hierbei wird zwischen der Anpassung nach § 22 MsV (Anpassung mittels Index) und nach § 23 MsV (Anpassung mittels Stichprobe) unterschieden. Die Überschriften sind insofern irreführend, als dass in beiden Fällen anhand einer stichprobenbasierten Erhebung ein Index errechnet wird, der dann zur Berechnung des Anpassungsfaktors verwendet wird. Diese Anpassungsvarianten unterscheiden sich unter anderem darin, welcher Preisindex herangezogen wird bzw. inwieweit er nach Teilmärkten ausdifferenziert ist bzw. sein darf.

Die Anpassung an die Marktentwicklung anhand der Entwicklung des Verbraucherpreisindexes nach § 22 MsV stützt sich auf den bundesweiten Verbraucherpreisindex des Statistischen Bundesamtes. Dieser beschreibt die durchschnittliche Preisentwicklung aller Waren und Dienstleistungen, die private Haushalte für Konsumzwecke kaufen und die in einem statistischen Warenkorb enthalten sind. Maßgebend ist dabei der Indexstand zum Stichtag der Datenerhebung. Die Anwendung eines anderen Indexes – beispielsweise des

Nettokaltmietenindexes eines Bundeslandes – ist bei qualifizierten Mietspiegeln rechtlich unzulässig.[30]

Demgegenüber wird bei einer Anpassung anhand einer Stichprobe nach § 23 MsV ein Preisindex eigens für die betreffende Mietspiegelkommune berechnet, der auf Basis einer dafür durchgeführten stichprobenbasierten Erhebung ermittelt wird. Zu beachten ist hierbei, dass für diese Erhebung analoge Grundgesamtheiten definiert werden, eine analoge fragebogenbasierte Filterung mietspiegelrelevanter Fälle erfolgt, nachträglich analoge Datenbereinigungs- bzw. -plausibilisierungsschritte durchgeführt werden und eine erneute Gewichtung des Datensatzes zum Ausgleich von ggf. unterschiedlichen Ziehungs- und Teilnahmewahrscheinlichkeiten vorgenommen wird. Eine Ausdifferenzierung des so ermittelten Preisindexes nach Teilmärkten ist dabei anders als bei einem Rückgriff auf den Verbraucherpreisindex zulässig, aber nicht verpflichtend.

Jedes der beiden Anpassungsverfahren ist mit spezifischen Vor- und Nachteilen verbunden. Die vergleichende Bewertung der Verfahren muss sich dabei auf geeignete Bewertungskriterien stützen. Ein in der Praxis nicht unerhebliches Bewertungskriterium ist der höchst unterschiedliche Zeit- und Kostenaufwand der verschiedenen Anpassungsverfahren. Während die Anpassung anhand der Entwicklung des Verbraucherpreisindexes grundsätzlich eine sehr zeitsparende und äußerst kostengünstige Möglichkeit der Anpassung an die Marktentwicklung darstellt, gehen Stichprobenerhebungen mit sich anschließender Ermittlung von Preisindizes mit einem deutlich größeren Zeit- und Kostenaufwand einher.

Ein weiteres Bewertungskriterium stellt die statistische Anpassungsgüte (goodness of fit) dar. Im Idealfall bildet ein Mietpreisindex die tatsächliche mittlere Marktentwicklung ab. Letztere ist jedoch unbekannt und sie im Mietspiegelkontext wahrheitsgetreu ermitteln zu wollen, wäre nur auf Vollerhebungsbasis möglich. Aus Zeit- und Kostengründen werden Mietspiegel jedoch in aller Regel auf Stichprobenbasis berechnet.

Wenn nun aber die wahre Marktentwicklung faktisch unbekannt ist, kann die Entscheidung für eine der beiden Anpassungsmethoden nicht darauf beruhen, wie stark der jeweils herangezogene Preisindex von der tatsächlichen Entwicklung abweicht. Aus diesen grundlegenden Überlegungen folgt, dass eine Anpassung mittels des Verbraucherpreisindexes aus rein

[30] Im Regierungsentwurf zur Mietspiegelverordnung vom 27.02.2020 war die Möglichkeit der Anpassung unter Rückgriff auf einen bundeslandspezifischen Nettokaltmietenindex oder einen noch stärker regionalisierten Index noch vorgesehen (vgl. S. 5 und 21).

methodischer Sicht abzulehnen wäre – und zwar nicht deshalb, weil der Verbraucherpreisindex die tatsächliche Mietpreisentwicklung in einer Mietspiegelkommune notgedrungen falsch wiedergeben würde, sondern allein deshalb, weil es keinerlei stichhaltige Anhaltspunkte dafür gibt, wie verlässlich er im Einzelfall ist.[31]

Im Gegensatz zu einer Anpassung mittels Verbraucherpreisindex ermöglicht eine stichprobenbasierte Indexberechnung dagegen eine Bewertung der Anpassungsgüte (goodness of fit). Denn gemäß § 8 Abs. 1 Satz 3 MsV muss für einen qualifizierten Mietspiegel die Stichprobe „repräsentativ" sein, was bedeutet, dass von der Stichprobe ein (zumindest approximativ) erwartungstreuer bzw. unverzerrter Rückschluss auf die Gesamtheit aller mietspiegelrelevanter Wohnungen[32] im Geltungsbereich des Mietspiegels möglich sein muss.[33] Erwartungstreue bedeutet dabei nicht notwendigerweise, dass ein auf einer solchen Stichprobe ermitteltes Mietpreisniveau exakt dem wahren, aber unbekannten Mietpreisniveau in der Grundgesamtheit aller mietspiegelrelevanten Mietverhältnisse entspricht. Erwartungstreue bedeutet vielmehr, dass gemittelt über alle Stichproben, die unter Zugrundelegung desselben Ziehungsverfahrens gezogen werden könnten, das wahre, aber unbekannte Mietpreisniveau in der Grundgesamtheit getroffen wird.[34] Auf Basis der konkret gezogenen Stichprobe wird daher zwar möglicherweise ein Mietpreisniveau abgeleitet, das in unbekanntem Ausmaß über oder unter dem wahren Mietpreisniveau liegt. Und folgerichtig kann eine stichprobenbasierte Mietspiegelanpassung gemessen an der (unbekannten) Realität genauso richtig und falsch wie eine Anpassung mittels Verbraucherpreisindex sein. Im Unterschied zur Anpassung via Verbraucherpreisindex ist eine stichprobenbasierte Anpassung aber wenigstens im Erwartungswert richtig.[35] Und darüber hinaus hält die Stichprobentheorie mit dem Standardfehler eine Kennziffer für den Grad an Sicherheit bereit, mit der von der Stichprobe auf

---

[31] Im Ergebnis wird durch den Rückgriff auf den Verbraucherpreisindex allerdings eine Geldwertsicherung für die Vermieterseite sichergestellt (vgl. Börstinghaus/Clar, 2023, S. 301).

[32] Zur Abgrenzung der mietspiegelrelevanten Wohnungssegmente im Detail vgl. Sebastian et al. (2024, S. 29-32).

[33] Vgl. allgemein zum Begriff der Repräsentativität sowie zu den Voraussetzungen zu deren Herstellung Sebastian et al. (2024, S. 32 f).

[34] Bei dieser Mittelung ist jede Stichprobe mit ihrer Eintrittswahrscheinlichkeit zu gewichten (vgl. Särndal et al. 1992, S. 40).

[35] Da ein auf Stichprobenbasis berechneter Preisindex gemäß Gleichung (2) zwei Schätzer ins Verhältnis zueinander setzt, handelt es sich beim Anpassungsfaktor um einen sog. Verhältnisschätzer. Obwohl – wie im vorliegenden Fall – die Schätzer im Zähler und Nenner jeweils erwartungstreu sind, ist der Verhältnisschätzer nur approximativ erwartungstreu, wobei das Ausmaß an Verzerrung aber gegen Null konvergiert (vgl. Särndal et al., 1992, S. 176 ff.; Stenger, 1986, S. 65 ff).

die unbekannten Verhältnisse der zugrundeliegenden Grundgesamtheit geschlossen werden kann.[36]

Bei der Wahl der Anpassungsmethode ist grundsätzlich willkürfrei zwischen dem Zeit- und Kostenaufwand einerseits und der (ggf. unbekannten) Anpassungsgüte abzuwägen. Da eine derartige Abwägung mit einem erheblichen Beurteilungsspielraum verbunden ist, entzieht sich die Entscheidung für eine Anpassungsmethode im Regelfall einer rechtlichen Beurteilung. Daher wird eine Anpassung mittels Verbraucherpreisindex immer eine rechtlich zulässige Methode sein. Entsprechend bedarf es auch keiner Begründung bei der Wahl der Anpassungsmethode oder auch des Wechsels einer einmal gewählten Methode. Hingegen ist es kritisch zu sehen, wenn ein Wechsel der Anpassungsmethode mit der Aussicht auf eine möglichst niedrige Erhöhung der im Mietspiegel auszuweisenden ortsüblichen Vergleichsmiete begründet wird. Auch wird es nicht zulässig sein, erst die Anpassungsfaktoren mittels Stichprobe nach § 23 MsV zu berechnen, eine Anpassung dann aber doch mittels des Verbraucherpreisindexes durchzuführen.

Zum Stand Januar 2022 wird in nahezu allen mietspiegelpflichtigen Städten, die einen qualifizierten Mietspiegel erstellen, die Möglichkeit einer Anpassung gewählt. Nur in drei Städten wird stattdessen der Mietspiegel alle zwei Jahre neu erstellt. Dabei werden etwa 72 % der Anpassungen mittels des Verbraucherpreisindexes und entsprechend 28 % mittels einer Stichprobe durchgeführt. Hiervon wiederum nehmen 63 % die Möglichkeit der Anpassung mittels einer umfangsreduzierten Stichprobe in Anspruch. Entsprechend entfallen 37 % der Anpassung mittels Stichprobe auf Neuerhebungen, d. h. Umfragen mit Stichproben im gleichen Umfang wie für die Neuerstellung des Mietspiegels.[37]

---

[36] Konkret beschreibt der Standardfehler das Ausmaß, mit der ein auf Basis einer Stichprobe abgeleiteter Parameter – hier insbesondere die mittlere Nettoquadratmetermiete mietspiegelrelevanter Wohnungen im Geltungsbereich des Mietspiegels – von Stichprobe zu Stichprobe schwankt. Der Standardfehler misst somit die Interstichprobenvariabilität. Auch wenn er exakt nur auf Vollerhebungsbasis beziffert werden kann, ist es möglich, aus der gezogenen Stichprobe heraus einen als Näherungswert fungierenden Schätzwert abzuleiten (vgl. Cischinsky et al. 2014, S. 240).

[37] Vgl. Sebastian/Range (2022, S. 33.).

4.3 Anforderungen an eine stichprobenbasierte Anpassung

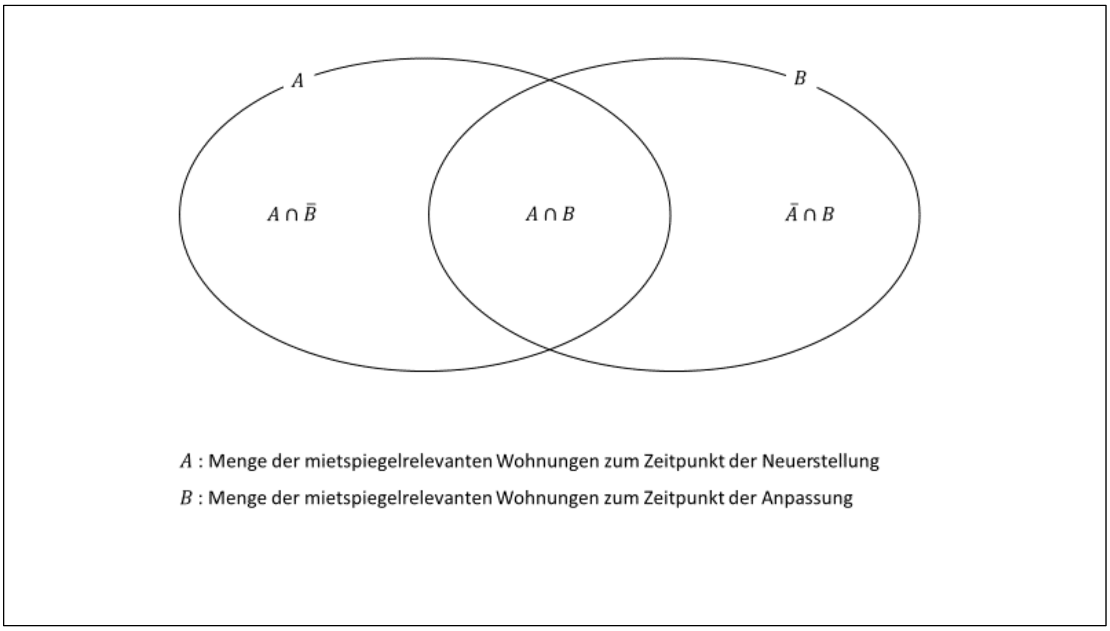
Fällt nun die kommunale Entscheidung gegen einen Rückgriff auf den Verbraucherpreisindex und zugunsten einer stichprobenbasierten Mietspiegelanpassung, ergeben sich eine Reihe von Fragen, die nachfolgend erörtert werden.

4.3.1 Querschnitts- oder Längsschnittdaten?

Eine Frage von hoher inhaltlicher, methodischer und erhebungspraktischer Bedeutung ist diejenige, ob der Preisindex gemäß Gleichung (2) auf der Grundlage von Längsschnitt- oder Querschnittsdaten ermittelt werden sollte. Im ersten Fall werden die (Quadratmeter-) Mieten von identischen (Stichproben-) Wohnungen zu zwei unterschiedlichen Zeitpunkten – dem zwei Jahre zurückliegenden Zeitpunkt der Neuerstellung und dem Anpassungszeitpunkt – herangezogen, während sich im anderen Fall die Ermittlung des Preisindexes auf die Mieten von unabhängig gezogenen Stichproben mietspiegelrelevanter Wohnungen zum Neuerstellungs- bzw. zum Anpassungszeitpunkt stützt.

Was die Vorgehensweise auf Basis von Längsschnittdaten betrifft, können aus Perspektive der Erstellungspraxis zwei grundlegende Herangehensweisen unterschieden werden. Deren nachfolgende Erläuterung und Bewertung stützt sich dabei auf das in Abbildung 4.1 enthaltene Venn-Diagramm, das den Umstand graphisch veranschaulicht, dass die Grundgesamtheiten mietspiegelrelevanter Wohnungen im Zeitablauf nicht notwendigerweise stabil sind. Vielmehr können vormals mietspiegelrelevante Wohnungen zum Zeitpunkt der Mietspiegelanpassung ihre Mietspiegelrelevanz verloren haben ($A \cap \bar{B}$). Umgekehrt wird es in der Mietspiegelpraxis andere Wohnungen geben, die im betrachteten Zweijahreszeitraum zwischen Neuerstellung und Anpassung zum Kreis der mietspiegelrelevanten Wohnungen dazugestoßen sind ($\bar{A} \cap B$). Und schließlich werden Wohnungen vorhanden sein, die zu beiden Zeitpunkten mietspiegelrelevant waren bzw. sind ($A \cap B$).

Abbildung 4.1: Mengenmäßige Beziehungen zwischen den Grundgesamtheiten mietspiegelrelevanter Wohnungen zum Zeitpunkt der Neuerstellung und zum Zeitpunkt der Anpassung



Im Sinne eines Panelansatzes könnte man sich auf die mietspiegelrelevanten Wohnungen des Neuerstellungszeitpunkts (Menge A) stützen, was erhebungspraktisch darauf hinausläuft, für die der bereinigten Nettostichprobe zugehörigen Mietwohnungen der Neuerstellung nach zwei Jahren erneut die Miete zu erheben und den dabei auf Stichprobenbasis geschätzten Mittelwert ins Verhältnis zu demjenigen zu setzen, der zum Neuerstellungszeitpunkt – ebenfalls auf Stichprobenbasis – ermittelt wurde. Hierbei stellt sich ein inhaltliches Problem in Bezug auf den Umgang mit inzwischen aus der Mietspiegelrelevanz herausgefallen Wohnungen ($A \cap \bar{B}$). Nach Ansicht der Autoren ist davon auszugehen, dass gemäß der gesetzlichen Vorgabe, dass die ortsübliche Vergleichsmiete auf Basis von mietspiegelrelevanten Wohnungen zu bilden ist, nur Wohnungen berücksichtigt werden dürfen, die beim Stichtag der Anpassung mietspiegelrelevant sind. Eine Nichtberücksichtigung dieser Vorgabe würde dazu führen, dass bei der Berechnung der angepassten ortsüblichen Vergleichsmiete insbesondere Wohnungen einfließen, die aufgrund ausbleibender Mieterhöhungen aus der 6-Jahres-Frist fallen, und dadurch das durchschnittliche Mietniveau unterschätzt wird. Doch auch die Begrenzung auf ($A \cap B$) ist inhaltlich problematisch. Denn da es sich bei den zu beiden Zeitpunkten mietspiegelrelevanten Wohnungen tendenziell um solche handeln dürfte, bei denen regelmäßig die Miete angehoben wird, kann dieser Panelansatz in letzter Konsequenz dazu führen, dass durch die Beschränkung auf solche Wohnungen der Mietpreisanstieg in einer Kommune überzeichnet wird.

Erhebungspraktisch stellt sich das für Panelerhebungen typische Problem von Panelmortalität aufgrund von Nichterreichbarkeit. Hier ist also der Ausfall von Wohnungen der bereinigten Nettostichprobe der Mietspiegelneuerstellung gemeint, die zwar zwei Jahre später noch vorhanden sind, für die aber eine Datenerhebung wegen Nichterreichbarkeit der Ansprechperson nicht zustande kommt. Panelmortalität wirkt dann verzerrend, wenn der Grund des Ausscheidens mit dem Untersuchungsgegenstand korreliert. Gerade bei Mieterbefragungen erscheint ein solcher Zusammenhang nicht ausgeschlossen, nämlich dann, wenn Haushalte in kleineren Wohnungen mit höheren Quadratmeterpreisen häufiger umziehen und umzugsbedingt nicht mehr befragt werden können. Erschwerend kommt hinzu, dass Umzüge im Mietwohnungsmarkt verhältnismäßig häufig vorkommen, so dass der beschriebene Panelansatz faktisch nur im Rahmen einer Vermieterbefragung denkbar ist. Allerdings ist auch bei Vermietern als Befragte ein gewisses Maß an Fluktuation und damit an Panelmortalität nicht ausgeschlossen, wobei hier etwa Kontaktierungsprobleme aufgrund veränderter Wohnadressen privater Vermieter und veränderter Eigentumsverhältnisse bzw. Verantwortlichkeiten für die Vermietung von Relevanz sein dürften. Entsprechende Ausfälle wären wiederum hinsichtlich einer potenziell verzerrenden Selektivität zu überprüfen, da das Panelsterben bei Privatpersonen, die Wohnraum vermieten, von größerer Bedeutung sein dürfte als bei institutionellen Vermietern.

Das Gegenstück zum bisher beschriebenen Ansatz stützt sich auf Wohnungen (bzw. eine Stichprobe daraus), die zum Anpassungszeitpunkt mietspiegelrelevant sind (Menge B). Dabei wird bei der Erhebung neben der Miete zum Anpassungszeitpunkt gleich auch die Miete von vor zwei Jahren erfragt oder alternativ geschätzt.[38] Anders als beim zuerst beschriebenen Panelansatz kann die bereinigte Nettostichprobe der zwei Jahre zurückliegenden Neuerstellung nicht wiederverwendet werden. Stattdessen muss eigens für die Anpassung eine Auswahlgrundlage erstellt werden, woraus sodann eine Bruttostichprobe zu ziehen und zu kontaktieren ist, um daraus wiederum eine bereinigte Nettostichprobe zu generieren. Viel kritischer und rechtlich eventuell sogar fragwürdig ist jedoch, dass auch bei diesem Ansatz das weiter oben beschriebene Problem besteht, dass in die spätere Indexbildung erhobene oder ggf. zugeschätzte Mieten von Wohnungen einfließen, die zum damaligen Zeitpunkt der Neuerstellung noch gar nicht mietspiegelrelevant waren. Sofern die erwähnte Schätzung auf

[38] Um eine Schätzung wird man in jedem Fall dann nicht umhin kommen, wenn sich die Befragungsperson nicht mehr an die Miethöhe zum Zeitpunkt der Mietspiegelneuerstellung vor zwei Jahren erinnern kann oder sie die Miete nicht kennt, weil sie damals noch nicht in der Wohnung lebte oder – im Falle einer Vermieterbefragung – damals noch kein Eigentum an der betreffenden Wohnung hatte.

Basis einer Schätzgleichung aus der vorangegangenen, also zwei Jahre zurückliegenden Mietspiegelerhebung erfolgt, entsteht ein weiteres Problem dadurch, dass diese Schätzgleichung auf eine abweichende Population übertragen wird, die in ihrer abweichenden Zusammensetzung zu einer anderen Schätzgleichung kommen würde.

Die beschriebenen Ansätze auf Basis von Längsschnittdaten sind somit inhaltlich wie auch erhebungspraktisch kritisch zu sehen, wodurch eine querschnittsdatenbasierte Vorgehensweise in den Fokus rückt. Bei einem solchen Ansatz wird die (auf Stichprobenbasis geschätzte) mittlere Miete zum Anpassungszeitpunkt (Menge B) ins Verhältnis zu derjenigen im Neuerstellungszeitpunkt (Menge A) gesetzt. Im Zähler und Nenner des Indexes stehen dadurch (geschätzte) Mittelwerte, denen im Regelfall unterschiedliche Grundgesamtheiten mietspiegelrelevanter Wohnungen zugrunde liegen.[39] Dies hat zur Folge, dass selbst für den Extremfall, in dem es in einer Mietspiegelkommune im Betrachtungszeitraum keinerlei Mieterhöhungen gab, allein durch Strukturverschiebungen zwischen den beiden Grundgesamtheiten sehr wahrscheinlich eine von Null abweichende Mietpreisdynamik gemessen werden kann. Die in der Praxis gemessene Mietpreisdynamik wird daher in unbekanntem, wenn auch vermutlich eher geringem Maße durch Populationsveränderungen mitbeeinflusst. Der auf Basis von Querschnittsdaten ermittelte Preisindex geht somit von der Annahme aus, dass solche Populationsveränderungen vernachlässigbar sind. Da es sich bei jeder indexbasierten Mietspiegelanpassung wesensbedingt jedoch stets um ein vereinfachtes Verfahren handelt, erscheint die Zugrundelegung einer solchen Annahme vertretbar, zumal die strukturellen Veränderungen der Grundgesamtheit innerhalb eines nur zwei Jahre umfassenden Zeitraums im Regelfall gering bleiben dürften.[40]

In der Gesamtbetrachtung ist ein Ansatz auf der Grundlage von Querschnittsdaten einem längsschnittdatenbasierten u.E. vorzuziehen, auch wenn Ersterer auf vereinfachenden, mit der Idee einer Anpassung allerdings vereinbaren Vorgehensweise beruht. Denn das mit einer längsschnittdatenbasierten Vorgehensweise üblicherweise erfolgte Ziel der Messung eines

[39] Nur im Ausnahmefall von $A = B = A \cap B$ sind die zugrunde liegenden Grundgesamtheiten identisch.

[40] Theoretisch ließe sich der Einfluss von Populationsveränderungen durch eine Standardisierung ausschalten. Dabei wird für den Anpassungszeitpunkt ein Mittelwert unter Zugrundelegung der Populationsstrukturen von vor zwei Jahren berechnet. In der Anwendungspraxis stellt sich hierbei jedoch das Problem, entlang welcher Merkmale die Populationsstruktur festzumachen ist. Die Vorgabe einer Standardpopulation impliziert daher faktisch eine kritisch zu bewertende Priorisierung von Wohnwertmerkmalen, die im Übrigen auch ein wesentliches Argument gegen eine nach Teilmärkten differenzierende Mietspiegelanpassung darstellt (vgl. Abschnitt 4.3.5)

reinen Preiseffekts ist nur unter Inkaufnahme erhebungspraktischer und inhaltlicher Nachteile zu erreichen.

4.3.2 Kann die Bruttostichprobe der letzten Neuerstellung erneut verwendet werden?

Auch wenn aus Sicht der Autoren ein Ansatz vorzuziehen ist, der sich auf Querschnittsdaten stützt, stellt sich die Frage, ob zumindest nicht die Bruttostichprobe der letzten Neuerstellung wiederverwendet werden kann, um dadurch Einspareffekte durch die entfallende Aufbereitung der Auswahlgrundlage und die dann redundante Stichprobenziehung zu realisieren. Allerdings sind wie auch beim längsschnittdatenbasierten Vorgehen die beiden zugrunde liegenden Stichproben dann nicht mehr länger unverbunden. Dies hat zur Folge, dass die in den folgenden Kapiteln diskutierten Fragen, welcher Mittelwerttest anzuwenden ist, nach welcher Formel ein reduzierter Stichprobenumfang für die Erhebung zur Anpassung des Mietspiegels festzulegen ist und was bei einer teilmarktspezifischen Anpassung zu beachten ist, im Wesentlichen anders zu beantworten wären. Wie nachfolgend herausgearbeitet wird, ist ein Rückgriff auf die „alte“ Bruttostichprobe jedoch mit methodischen Schwächen und nachteiligen erhebungspraktischen Konsequenzen verbunden, so dass er u. E. nicht erwogen werden sollte.

Zuerst ist sich vor Augen zu halten, dass neu gebaute Wohnungen, die seit der letzten Mietspiegelneuerstellung auf den Mietwohnungsmarkt gekommen sind, nicht über die Bruttostichprobe der letzten Neuerstellung erreichbar sind.[41] Sie müssen daher gesondert erhoben werden, wobei hierfür auf eine separate Datenbasis zugegriffen werden muss. Sodann stellt sich die Herausforderung, dass die so zusätzlich erhobenen Mietverhältnisse in Relation zu denjenigen mietspiegelrelevanten Mietverhältnissen gesetzt werden müssen, die über die alte Bruttostichprobe erreicht werden können. Dies macht ein Gewichtungsverfahren erforderlich, das Annahmen über den Anteil neu gebauter Wohnungen in der Grundgesamtheit aller mietspiegelrelevanter Wohnungen zum Anpassungszeitpunkt voraussetzt. Im Ergebnis läuft eine solche ergänzende Stichprobenerhebung dem eigentlichen Ziel des Rückgriffs auf die alte Bruttostichprobe, nämlich die Reduzierung des Erhebungsaufwands, zuwider.

Ein zweiter Grund, der gegen den Rückgriff auf die Bruttostichprobe der letzten Neuerstellung spricht, stellt die bereits unter 4.3.1 erläuterte Panelmortalität dar.

[41] Je nach verwendeter Auswahlgrundlage gilt dies auch für vormals gewerbliche und inzwischen zu Wohnzwecken umgenutzte Wohnungen.

Entscheidet man sich trotz der geschilderten Nachteile und der methodischen Vorbehalte für einen solchen Rückgriff, sollte die Auswahlgrundlage der vorangegangenen Mietspiegelerhebung nicht allzu stark um (zum damaligen Zeitpunkt) mietspiegelirrelevante Fälle bereinigt worden sein, also etwa um Wohnungen, für die damals aufgrund einer öffentlichen Förderung noch eine Mietpreisbindung bestand oder um vom Eigentümer selbstgenutzte Wohnungen. Denn anderweitig könnten solche Wohnungen auch dann nicht berücksichtigt werden, wenn sie etwa bedingt durch Bindungsauslauf oder die Vermietung einer zuvor selbstgenutzten Wohnung zwischenzeitlich mietspiegelrelevant geworden sind. Die mit Blick auf die Anpassungserhebung aufzustellende Forderung nach einem möglichst weitgehenden Verzicht auf die Bereinigung der Auswahlgrundlage der vorangegangenen Mietspiegelerhebung steht jedoch dem berechtigten Anliegen dieser vorangegangenen Mietspiegelerhebung diametral entgegen, aus der damaligen Auswahlgrundlage (damals) irrelevante Wohnungen zur Reduktion des Erhebungsaufwandes und damit zur Steigerung der Stichprobeneffizienz soweit möglich herauszunehmen.

4.3.3 Wann ist anzupassen bzw. wann kann tatsächlich eine Mietniveauveränderung festgestellt werden?

Da ein stichprobenbasiert abgeleitetes Mietpreisniveau stets unsicherheitsbehaftet ist, ist es nicht ausgeschlossen, dass ein gemessener Unterschied zwischen dem mittleren Mietniveau des anzupassenden Mietspiegels und demjenigen, das im Zuge der Anpassung ermittelt wurde, in Wirklichkeit überhaupt nicht besteht. Ob eine Anpassung eines qualifizierten Mietspiegels angezeigt ist, lässt sich zweckmäßig mittels eines Mittelwerttests bestimmen. Dieser prüft, ob sich zwei stichprobenbasierte Mittelwerte statistisch signifikant voneinander unterscheiden.[42] Statistisch signifikant bedeutet dabei unter Inkaufnahme einer vorzugebenden maximalen Irrtumswahrscheinlichkeit (Signifikanzniveau), die üblicherweise auf 5 % festgelegt wird. Neben dem vorzugebenden Signifikanzniveau hängt das Ergebnis eines solchen Mittelwerttests und damit die Feststellung statistisch signifikanter Unterschiede von den beiden zu vergleichenden (Stichproben-)Mittelwerten und von den jeweils dahinterstehenden Standardfehlern ab. Je stärker sich die beiden Mittelwerte voneinander unterscheiden und je kleiner die zugehörigen Standardfehler sind, desto eher führt dies dazu, gemessene Unterschiede als statistisch signifikant zu qualifizieren.

[42] Technisch gesprochen überprüft ein solcher Mittelwerttest, ob sich die Nullhypothese identischer Mittelwerte in den beiden zugrunde liegenden Grundgesamtheiten anhand von Stichprobendaten unter Zugrundelegung eines vorzugebenden Signifikanzniveaus verwerfen lässt.

Schlussendlich hängt das Testergebnis auch von dem herangezogenen Testverfahren ab. Hierfür bietet sich im vorliegenden Fall ein zweiseitiger Welch´s-t-Test für unabhängige Stichproben an. Der Rückgriff auf die zweiseitige Version ist deshalb angezeigt, weil eine Mietpreissteigerung gegenüber dem bisherigen Mietpreisniveau und damit eine Anpassung mit einem positiven Anpassungsfaktor nicht notwendigerweise vorausgesetzt werden kann, auch wenn in der Praxis fallende Mietpreise die Ausnahme sein dürften. Im Gegensatz zu einem „konventionellen" t-Test setzt Welch´s-t-Test keine Varianzhomogenität, d. h. keine identische Streuung des betrachteten Merkmals (hier: die Nettoquadratmetermiete) in den beiden jeweils zugrunde liegenden Grundgesamtheiten voraus. Der genannte Mittelwerttest unterstellt zwar streng genommen normalverteilte Grundgesamtheiten – im vorliegenden Fall normalverteilte Nettoquadratmetermieten. Allerdings reagiert er wenig empfindlich auf eine Verletzung der Normalverteilungsannahme (vgl. z. B. Kubinger et al., 2009; Sawilowsky/Blair, 1992). U. E. kann daher auf eine Überprüfung der Normalverteilungsannahme verzichtet werden, zumal empirische Verteilungen der Nettoquadratmetermieten üblicherweise einen der Normalverteilung ähnelnden, allerdings rechtsschiefen glockenform-ähnlichen Verlauf aufweisen, ohne jedoch durch eine extreme Schiefe gekennzeichnet zu sein. Und abgesehen davon sind die üblichen Stichprobenumfänge im Kontext von Mietspiegelerhebungen ohnehin in einer Größenordnung, bei der die Verteilung der Testgröße mit der Standardnormalverteilung faktisch zusammenfällt. Ein sich auf den Zentralen Grenzwertsatz stützender Gauß-Test würde daher im Regelfall dasselbe Testergebnis liefern, d. h. zu derselben Aussage im Hinblick auf statistische Signifikanz kommen (vgl. Anderson et al., 1997, S. 182; Schlittgen, 2012, S. 358 f.).

4.3.4 Wie ist der Umfang der Stichprobe bei der Anpassung festzulegen?

Bei einer stichprobenbasierten Anpassung eines qualifizierten Mietspiegels sind nach § 23 Abs. 2 Satz 1 MsV grundsätzlich die Anforderungen für die Neuerstellung nach den §§ 7 bis 21 MsV entsprechend anwendbar. Ausdrücklich ist es aber nach Satz 2 zulässig, dass die Nettostichprobe von den in § 11 MsV aufgeführten Werten (nach unten) abweicht. Es ist also entsprechend der üblichen Praxis eine verkleinerte Stichprobe möglich, ohne dass explizit geregelt ist, wie stark diese Verkleinerung sein darf. Aufgrund dessen soll im Rahmen des vorliegenden Beitrags hierfür eine begründete Empfehlung für die Anpassungspraxis formuliert werden. Dabei sind die nachfolgenden Überlegungen leitend:

Die Höhe des (auf Stichprobenbasis geschätzten) Standardfehlers hängt neben der Streuung des betreffenden Merkmals – hier der Nettoquadratmetermiete – in der Stichprobe (sog. Stichprobenstandardabweichung) entscheidend vom (Netto-) Stichprobenumfang ab. Unter sonst gleichen Bedingungen gilt, dass der Standardfehler mit dem Stichprobenumfang negativ korreliert ist. Je größer der Stichprobenumfang ist, desto kleiner ist unter sonst gleichen Bedingungen der Standardfehler. Und je kleiner die Standardfehler der auf Stichprobenbasis bestimmten mittleren Nettoquadratmetermieten des anzupassenden Mietspiegels und der zur Anpassung durchgeführten Erhebung sind, desto eher schlägt der o. a. zweiseitige Welch´s-t-Test für unabhängige Stichproben an und qualifiziert gemessene Unterschiede als statistisch signifikant.

Welches Maß an Unsicherheit und damit welchen maximalen Standardfehler man zu akzeptieren bereit ist, bleibt dabei eine normative Frage. In der Mietspiegelverordnung werden hierzu keine Festlegungen getroffen.[43] Aufgrund fehlender rechtlicher Vorgaben, wie weit der Stichprobenumfang bei einer Anpassung reduziert werden darf, wird im Folgenden ein Vorschlag unterbreitet. Dem Vorschlag liegt die Idee zugrunde, dass ein „praktisch relevanter Unterschied" zwischen den mittleren Nettoquadratmetermieten in den beiden zu vergleichenden Grundgesamtheiten – derjenige der anzupassenden Mietspiegelerhebung und derjenige zum Anpassungszeitpunkt – durch den o.a. Mittelwerttest mit einer vorzugebenden Mindestwahrscheinlichkeit aufgedeckt werden sollte. Die zu beantwortende Frage lautet daher, wie groß der Umfang (netto) der im Rahmen der Anpassung zu ziehenden Stichprobe mindestens sein muss, so dass unter Vorgabe einer Mindestwahrscheinlichkeit ein praktisch relevanter Mittelwertunterschied durch einen zweiseitigen Mittelwerttest aufgedeckt, also als statistisch signifikant qualifiziert wird.

Die Beantwortung dieser Frage kommt aus statistischen Gründen nicht ohne normative Vorgaben aus, die allerdings aus praxisnahen Erwägungen heraus und unter Verweis auf übliche Konventionen der statistischen Testtheorie festgelegt werden können. Aufgrund des typischen Stichprobenumfangs von Mietspiegelerhebungen ist eine Messung von statistisch signifikanten Unterschieden der Mittelwerte in der Regel bei sehr kleinen Abständen der Mittelwerte nicht möglich. Daher erscheint für den Abstand der Mittelwerte der Nettoquadratmetermieten (sog. Effektstärke) die Vorgabe einer Veränderung um 0,20 Euro zugleich praktisch relevant als auch bei typischen Stichprobenumfängen statistisch absicherbar. Es handelt sich somit um eine

[43] Vorgaben bestehen nur in Form von Mindeststichprobenumfängen bei der Neuerstellung eines qualifizierten Mietspiegels nach § 11 MsV. Den vorstehenden Ausführungen zufolge ist dies nur bedingt sachgerecht. Statt Vorgaben zum Stichprobenumfang sollte der maximal zulässige Standardfehler (beispielsweise 5 %) vorgegeben werden, woraus dann quasi endogen und einzelfallbezogen erforderliche (Netto-) Stichprobenumfänge abzuleiten wären.

angemessene normative Vorgabe zur Bestimmung des erforderlichen Stichprobenumfangs, um eine reale Veränderung der Nettoquadratmetermiete in Höhe der vorgegebenen Größenordnung (also um 0,20 Euro) durch den Mittelwertest mit „ausreichend großer Sicherheit“ aufzudecken.[44] Bereits an dieser Stelle sei darauf hingewiesen, dass diese Festlegung in der Anwendungspraxis weder Anpassungen von unter 0,20 Euro ausschließt noch garantiert, dass bei Abständen zwischen den Stichprobenmittelwerten von mehr als 0,20 Euro stets eine Anpassung in Höhe des gemessenen Mittelwertunterschieds vorzunehmen ist. Ob am Ende ein gemessener Mittelwertunterschied eine entsprechende Mietspiegelanpassung zur Folge hat, ist allein dem weiter oben besprochenen Mittelwerttest vorbehalten und bleibt daher ex ante offen.

Die Quantifizierung der „ausreichend großen Sicherheit“ adressiert die Thematik der Teststärke eines statistischen Tests, welche die Wahrscheinlichkeit angibt, mit der ein tatsächlich vorhandener Unterschied durch einen statistischen Test aufgedeckt wird.[45] Einem Vorschlag von Cohen (1988, S. 56) folgend, wird diese Wahrscheinlichkeit üblicherweise auf 80 % festgelegt, was im Folgenden übernommen wird.

Bei Mittelwerttests im Allgemeinen wird als Signifikanzniveau, d. h. als Obergrenze für die Wahrscheinlichkeit einer irrtümlichen Verwerfung einer von einem fehlenden Unterschied ausgehenden Nullhypothese und damit als Obergrenze für die Wahrscheinlichkeit eines Fehlers 1. Art bzw. einer falsch-positiven Testentscheidung, üblicherweise und so auch hier ein Wert von 5 % angesetzt (siehe oben).

Die Herleitung einer allgemeinen Variante der nachfolgenden Formel (4.3) ohne die normativen Setzungen in Bezug auf die Effektstärke, die Trennschärfe und das Signifikanzniveau findet sich im Anhang, so dass sich die folgenden Ausführungen auf ein praxisnahes Beispiel beschränken können. Unter der zusätzlichen Annahme, dass sich die Streuung der Nettoquadratmetermieten in der Grundgesamtheit und damit sehr wahrscheinlich auch in der für die Anpassung zu ziehende Stichprobe gegenüber der Situation zwei Jahre zuvor nicht

---

[44] Sofern angestrebt wird, auch geringere Veränderungen der mittleren Nettoquadratmetermiete statistisch nachweisen zu können, ist ein größerer Stichprobenumfang erforderlich (vgl. hierzu Abbildung 4.2 und die Erläuterungen dazu).

[45] Im vorliegenden Fall ist das die Wahrscheinlichkeit, mit der eine falsche Nullhypothese eines im Anpassungszeitraum unveränderten Mietniveaus abgelehnt wird. Die Teststärke ist daher das Gegenstück für die Wahrscheinlichkeit eines Fehlers zweiter Art, also einer falsch-negativen Testentscheidung, bei der eine unzutreffende Nullhypothese nicht abgelehnt wird.

verändert hat, bestimmt sich der Mindeststichprobenumfang (netto) n_2 der zur Anpassung durchzuführenden Erhebung nach folgender Näherungsformel:

$$n_2 \approx \frac{n_1 S_1^2 \left(\frac{0{,}842+1{,}96}{0{,}2}\right)^2}{n_1 - S_1^2 \left(\frac{0{,}842+1{,}96}{0{,}2}\right)^2} \qquad \text{für } n_1 > S_1^2 \left(\frac{0{,}842+1{,}96}{0{,}2}\right)^2 \tag{4.3}$$

Dabei bezeichnen n_1 den Nettostichprobenumfang der anzupassenden und daher bereits erfolgten Mietspiegelerhebung und S_1^2 die für diese Erhebung berechenbare Stichprobenvarianz in Bezug auf die Nettoquadratmetermiete.[46]

Zu beachten ist, dass die Näherungsformel nur für $n_1 > S_1^2 \left(\frac{0{,}842+1{,}96}{0{,}2}\right)^2$ definiert ist und daher eine Anforderung an den Nettostichprobenumfang der vorangegangenen Mietspiegelerhebung stellt. Ist diese Anforderung nicht erfüllt, ist das der Näherungsformel zugrunde liegende Anliegen, eine tatsächliche Mietpreisveränderung von 0,20 Euro mit einer Wahrscheinlichkeit von 80 % durch einen Mittelwerttest zum Signifikanzniveau von 5 % aufzudecken, nicht erreichbar – selbst dann nicht, wenn die Anpassungserhebung auf Vollerhebungsbasis durchgeführt wird. In diesem Fall war die Schätzgenauigkeit bereits beim anzupassenden Mietspiegel zu niedrig, um den Formelvorgaben zu genügen. Als einziger Ausweg bleibt dann die Senkung der normativen Anforderungen, beispielsweise durch Erhöhung der Effektstärke auf 0,30 Euro.

Zu beachten ist auch, dass die Näherungsformel unter Umständen zu einem Nettostichprobenumfang n_2 führt, der über n_1 liegt. Sowohl aus statistischer Sicht als auch nach der Mietspiegelverordnung ist hiergegen zwar nichts einzuwenden. In der Erhebungspraxis wird es aber kaum darstellbar sein, für die Anpassungserhebung einen höheren Stichprobenumfang als für den anzupassenden Mietspiegel vorzusehen, zumal dadurch das mit einer Anpassung verfolgte Ziel, nämlich eine Kostenreduktion gegenüber einer Mietspiegelneuerstellung zu realisieren, konterkariert wird. In der Praxis werden für die Anpassungserhebung daher nur Stichprobenumfänge maximal in Höhe des Stichprobenumfangs n_1 des anzupassenden Mietspiegels in Erwägung gezogen werden. Es wird daher

$$n_2^* = min\{n_1; n_2\} \tag{4.4}$$

[46] Zur Berechnung der Stichprobenvarianz im Falle einer uneingeschränkten Zufallsauswahl (ohne Zurücklegen) vgl. z.B. Särndal et al. (1992, S. 47).

vorzugeben sein.

Als Beispiel sei angenommen, dass der qualifizierte Mietspiegel einer Beispielkommune eine mittlere Nettoquadratmetermiete in Höhe von 10,00 Euro aufweist, die auf Basis von $n_1 = 4.000$ mietspiegelrelevanten Wohnungen der bereinigten Nettostichprobe berechnet wurde. Weiter sei angenommen, dass sich in der damaligen Mietspiegelerhebung für die Nettoquadratmetermiete eine Stichprobenstandardabweichung in Höhe von 2,40 Euro ergab, woraus sich eine Stichprobenvarianz S_1^2 in Höhe des Quadrats ableitet, d. h. $S_1^2 = 5{,}76$. Durch Einsetzen dieser Parameter in Formel (4.3) errechnet sich ein (Mindest-) Stichprobenumfang n_2 von (näherungsweise) 1.576 Fällen. Können in der Anpassungserhebung somit in etwa 1.576 auswertbare Fragebögen zu mietspiegelrelevanten Wohnungen gewonnen werden, wird ein zweiseitiger Zwei-Stichproben-t-Test zum Signifikanzniveau 5 % mit einer Wahrscheinlichkeit von 80 % einen statistisch signifikanten Unterschied zwischen den Stichprobenmittelwerten der Nettoquadratmetermieten der Anpassungserhebung und der zwei Jahre zuvor durchgeführten Mietspiegelerhebung dann aufdecken, wenn sich die Miete im Betrachtungszeitraum um mindestens 0,20 Euro auf 10,20 Euro erhöht hat (oder um mindestens denselben Betrag auf 9,80 Euro gesunken ist).

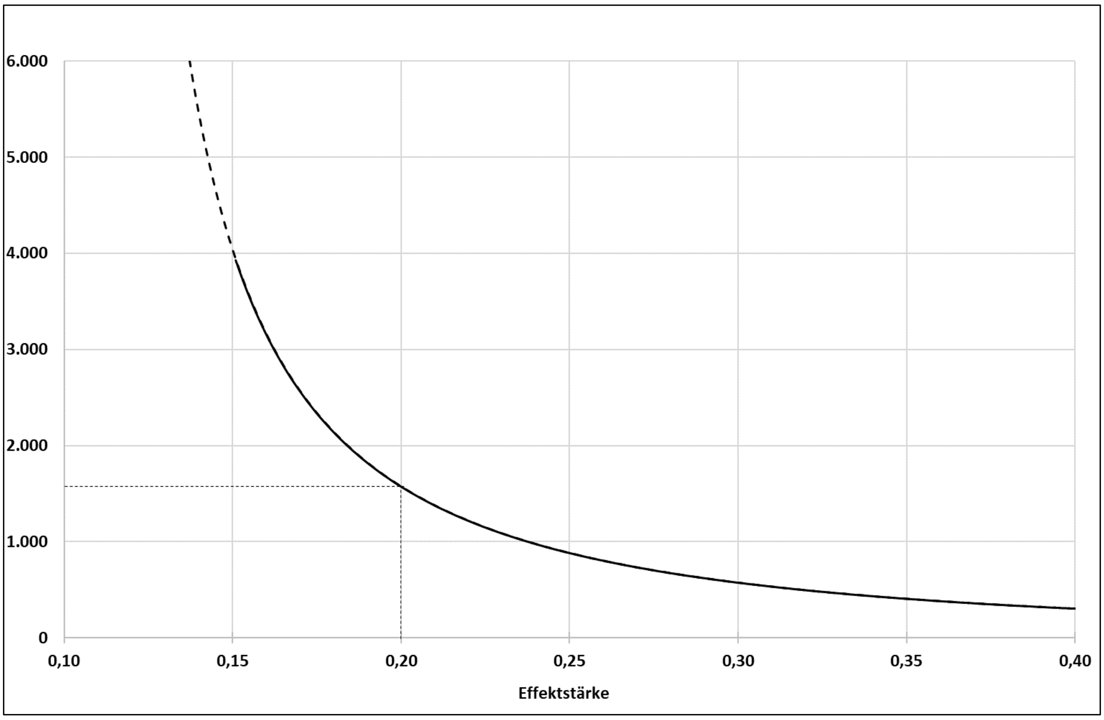
Die obige Formel reagiert sensitiv auf die Veränderung ihrer Einflussgrößen. Ist der Stichprobenumfang der vorherigen Mietspiegelerhebung (n_1) größer oder ist die für diese Erhebung berechnete Stichprobenvarianz bzw. -standardabweichung niedriger, sinkt der Mindeststichprobenumfang n_2 für die Anpassungserhebung. Gleiches gilt, wenn ein höheres Signifikanzniveau oder eine niedrigere Teststärke vorgegeben werden oder wenn die Effektstärke heraufgesetzt wird, beispielsweise auf 0,40 Euro.

Wie die Näherungsformel (4.3) bezogen auf das angeführte Beispiel bei einer Variation der normativen Vorgaben in Bezug auf die Effektstärke, das Signifikanzniveau und die Teststärke reagiert, ist den nachfolgenden drei Abbildungen zu entnehmen.

Abbildung 4.2 illustriert für die Beispielkommune den negativen Zusammenhang zwischen der vorzugebenden Effektstärke (hier gemessen in Euro) und dem gesuchten Mindeststichprobenumfang n_2. Kleinere Effektstärken gehen demnach mit größeren Werten für n_2 einher. Da Werte für $n_2 > 4.000$ die Vorgabe in Gleichung (4) verletzen, wonach der Stichprobenumfang für die Anpassungserhebung denjenigen des anzupassenden Mietspiegels nicht überschreiten darf, ist die Kurve für $n_2 > 4.000$ gestrichelt dargestellt. Effektstärken im gestrichelten Bereich der Kurve können daher dann nicht durch einen Mittelwerttest zum

Signifikanzniveau von 5 % mit einer Wahrscheinlichkeit von 80 % aufgedeckt werden, wenn der Stichprobenumfang n_2 auf 4.000 Fälle begrenzt wird.

Abbildung 4.2: Mindeststichprobenumfang n_2 in Abhängigkeit der Effektstärke*

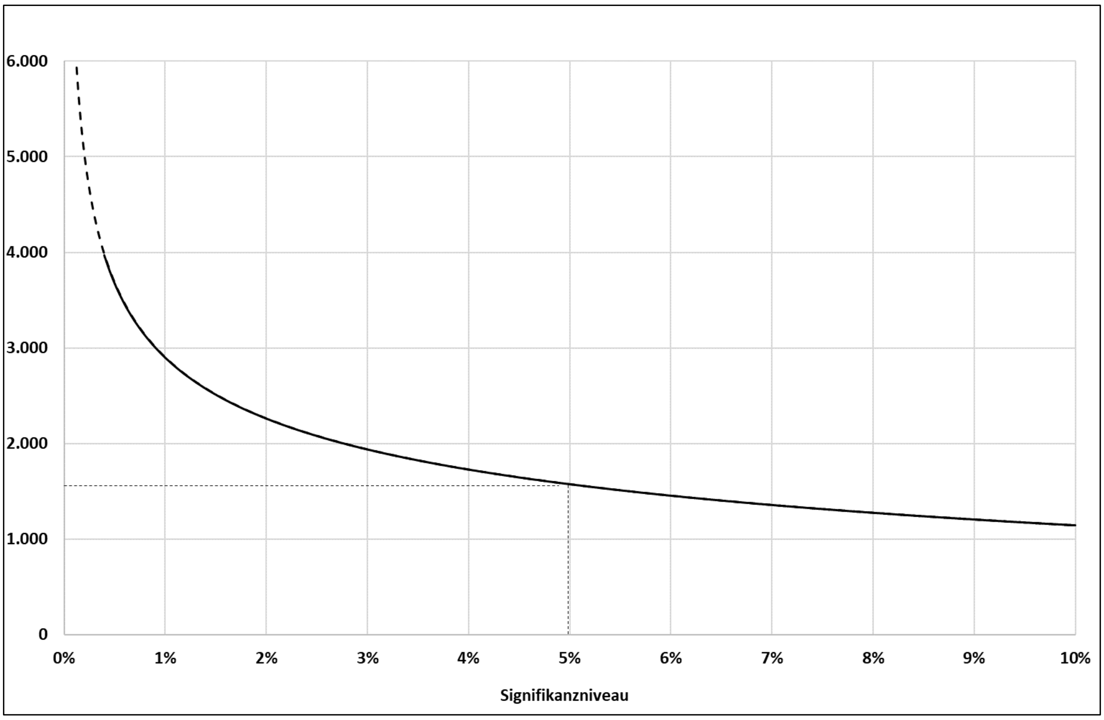


* *Vorgenommene Parametersetzungen:* $n_1 = 4.000$ *und* $S_1^2 = 5{,}76$

Die Kurve in Abbildung 4.2 nähert sich von rechts kommend asymptotisch an eine Effektstärke von 0,107 Euro an. Noch kleinere Effektstärken können mit einer Wahrscheinlichkeit von 80 % demnach selbst dann nicht statistisch signifikant nachgewiesen werden, wenn die Anpassungserhebung als Vollerhebung durchgeführt wird und das mittlere Mietpreisniveau zum Anpassungszeitpunkt nicht mehr geschätzt, sondern exakt beziffert wird. Ursächlich hierfür ist die Schätzungenauigkeit des anzupassenden Mietspiegels.

Abbildung 4.3 beschreibt den ebenfalls inversen Zusammenhang zwischen dem normativ festzulegenden Signifikanzniveau und n_2. Der „Preis" für ein immer weiter abgesenktes Signifikanzniveau ist ein immer größerer Stichprobenumfang.

Abbildung 4.3: Mindeststichprobenumfang n_2 in Abhängigkeit des Signifikanzniveaus*



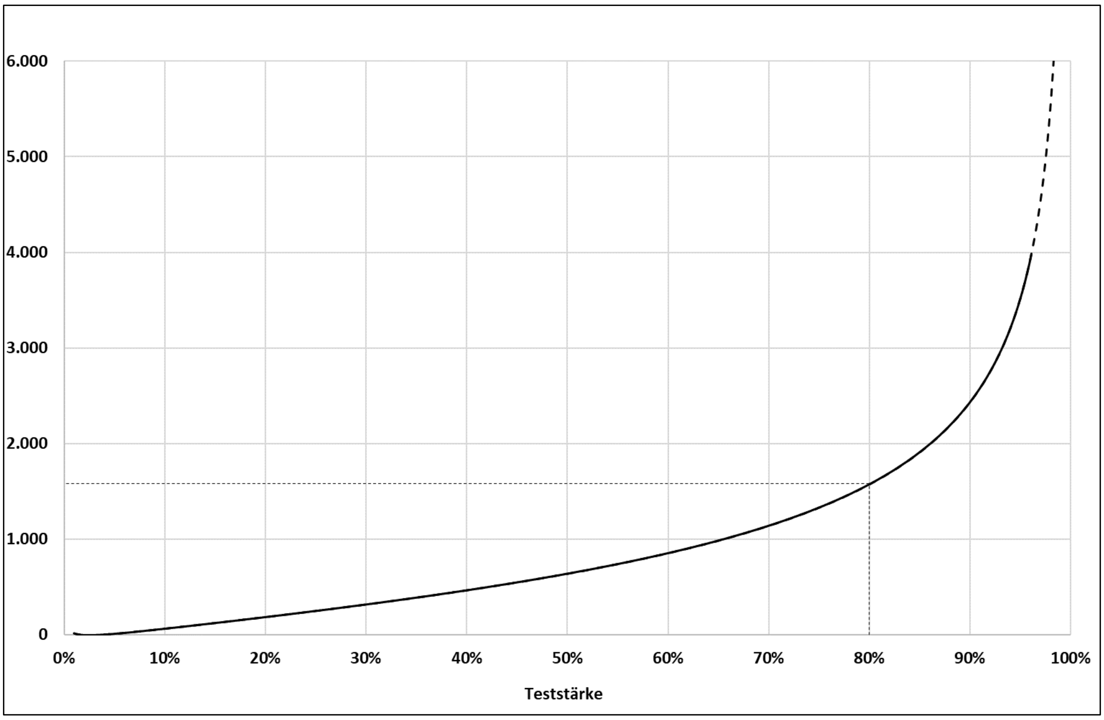
* *Vorgenommene Parametersetzungen:* $n_1 = 4.000$ *und* $S_1^2 = 5{,}76$

Ein Anstieg der Teststärke steht dagegen in einem positiven Zusammenhang mit dem Stichprobenumfang. Eine höhere Teststärke bedingt daher einen entsprechend größeren Stichprobenumfang (vgl. Abbildung 4.4). Dabei ist die Vorgabe einer Teststärke von 100 % genauso wenig sinnvoll wie die Vorgabe eines Signifikanzniveaus von 0 %.

Abschließend ist auf drei erhebungspraktische Aspekte hinzuweisen. Erstens ist zu beachten, dass die obige Formel auf den Mindeststichprobenumfang der bereinigten Nettostichprobe und nicht auf den der eigentlich zu ziehenden Bruttostichprobe abstellt. Um den korrespondierenden (Mindest-)Bruttostichprobenumfang für die Anpassungserhebung zu bestimmen, sind belastbare Annahmen zum Ausmaß von Ausfällen wegen Non-Response und Mietspiegelirrelevanz zu treffen, die sich im Idealfall auf Erfahrungen aus vorangegangenen Mietspiegelerhebungen stützen.[47] Darüber hinaus bietet es sich an, einen ausreichend bemessenen Sicherheitspuffer vorzusehen, etwa dafür, dass der Anteil mietspiegelirrelevanter Fälle gegenüber vorangegangenen Mietspiegelerhebungen größer ausfällt.

[47] Vgl. hierzu auch die Ausführungen in Sebastian et al. (2024, S. 36-38).

Abbildung 4.4: Mindeststichprobenumfang n_2 *in Abhängigkeit der Teststärke**



* *Vorgenommene Parametersetzungen:* $n_1 = 4.000$ *und* $S_1^2 = 5{,}76$

Ebenfalls zu beachten ist, dass die Bestimmung des für die Anpassungserhebung erforderlichen Stichprobenumfangs auf Ergebnisse des anzupassenden Mietspiegels angewiesen ist. Aufgrund dessen ist es zwingend erforderlich, dass die Dokumentation dieses Mietspiegels die entsprechenden Angaben ausweist, u. a. auch die Stichprobenstandardabweichung bzw. -varianz des Schätzwerts für die mittlere Nettoquadratmetermiete. Hierbei ist insbesondere aus Sicht der Auftraggebenden zu beachten, dass die Dokumentationspflichten des § 20 Abs. 2 MsV die explizite Ausweisung der Stichprobenstandardabweichung bzw. -varianz nicht beinhaltet. Um bei einem möglichen Wechsel des Mietspiegelerstellers nicht auf interne Informationen des früheren Erstellers angewiesen zu sein, sollte daher neben der mittleren Nettoquadratmetermiete auf die Ausweisung der zugehörigen Stichprobenstandardabweichung bzw. -varianz bestanden werden.

Nach § 23 Abs. 3 MsV ist ein abweichender Stichprobenumfang „in der Dokumentation darzulegen und zu begründen.“ Eine Begründung ist immer dann bereits implizit gegeben, wenn eine Differenz der für die Stichproben berechneten mittleren Nettomieten mit dem als akzeptabel zu definierenden Signifikanzniveau statistisch signifikant ist. Wenn der gemessene Mittelwertunterschied nicht statistisch signifikant ist, kann der geringere Stichprobenumfang nicht auf diesem Wege begründet werden und es stellt sich die Frage, warum der gemessene

Mittelwertunterschied nicht statistisch signifikant ist. In diesem Fall sind entweder die Umfänge der zu vergleichenden Stichproben zu klein oder der Mittelwertunterschied ist so gering, dass er ggf. auch dann nicht statistisch signifikant wäre, wenn für die Anpassungserhebung derselbe Stichprobenumfang wie für den anzupassenden Mietspiegel vorgesehen wurde. Sofern der Stichprobenumfang für die Anpassungserhebung nach Formel (4.3) festgelegt wurde, kann zumindest davon ausgegangen werden, dass Änderungen im berücksichtigten Umfang von 0,20 Euro mit einer Wahrscheinlichkeit von 80 % als statistisch signifikant qualifiziert werden können. Auf eine weitergehende Begründung kann dann u. E. verzichtet werden.

4.3.5 Einheitliche oder nach Teilmärkten differenzierte Anpassung?

Wie an früherer Stelle ausgeführt, ist eine nach Teilmärkten differenzierte stichprobenbasierte Anpassung nicht verpflichtend, aber zulässig. Bei einer solchen Anpassung stellen sich eng miteinander verwobene Fragen. Zunächst ist zu klären, zwischen welchen Teilmärkten unterschieden werden soll. Teilmärkte können beispielsweise für unterschiedliche Wohnlagen gebildet werden, aber auch nach jedem anderen Merkmal, etwa nach Baualtersklassen, Ausstattungsmerkmalen oder auch Kombinationen mehrerer Merkmale. Sinnvollerweise kommen als teilmarktspezifische Abgrenzungsmerkmale nur Wohnwertmerkmale des aktuell geltenden Mietspiegels in Betracht. Denn nur so ist eine Anpassung der Mietspiegelwerte in der bestehenden Struktur möglich, indem etwa die Basistabelle eines Mietspiegels sowie additive Zu- und Abschläge angepasst werden. Werden andere Abgrenzungsmerkmale hinzugezogen, muss die bisherige Struktur des Mietspiegels angepasst werden. Dies kann beispielsweise dadurch erfolgen, dass die Mietspiegelbroschüre für die verschiedenen Teilmärkte getrennt aufgesetzt werden muss. Es ist daher kritisch zu überprüfen, ob die zusätzliche Komplexität in der Anwendung gerechtfertigt erscheint.

Eine teilmarktspezifische Anpassung ist nur dann in Erwägung zu ziehen, wenn davon ausgegangen werden kann, dass die Mietpreisdynamik seit der letzten Mietspiegelneuerstellung innerhalb des räumlichen Geltungsbereichs des Mietspiegels nicht einheitlich gewesen ist. Darüber hinaus wird mit der Festlegung der Teilmärkte etwa anhand von Baualtersklassen oder Mietspiegellagen eine Priorisierung eines oder weniger Wohnwertmerkmale vorgenommen. Die übrigen Wohnwertmerkmale werden nicht berücksichtigt. Die Problematik der merkmalsspezifischen Priorisierung ist bei teilmarktspezifischen Fortschreibungen nicht zu vermeiden, aber sollte durch eine sachorientierte Abwägung der Teilmarktdefinition möglichst reduziert werden.

Entsprechende Abwägungen können sich zum einen auf Erfahrungswerte aus früheren Mietspiegelerstellungen stützen. So können etwa die bei der vorherigen Mietspiegelerstellung festgestellten Mietniveauunterschiede als Indikator für unterschiedliche Mietpreisdynamiken berücksichtigen werden, auch wenn es sich hierbei um einen zeitpunktbezogenen Vergleich und keinen direkten Indikator für unterschiedliche Dynamiken handelt. Auch aus dem Vergleich mehrerer Mietspiegel mit unterschiedlichen Stichtagen kann auf unterschiedliche Entwicklungen von Teilmärkten geschlossen werden. Hierbei ist allerdings der ggf. ausgeprägte Vergangenheitsbezug zu beachten, der nur bedingt Rückschlüsse auf die aktuellen Entwicklungen zulässt. Aktuelle Entwicklungen können aus anderen Datengrundlagen abgeleitet werden. Hierbei ist allerdings zu beachten, dass diese in der Regel den mietspiegelrelevanten Wohnungsmarkt nur sehr begrenzt abbilden.[48]

Zugleich ist die Größe des abgegrenzten Teilmarkts mit einzubeziehen.[49] Denn was an früherer Stelle zur Notwendigkeit rechtlicher Vorgaben in Bezug auf Mindeststichprobenumfänge im Falle eines – rechtlich zulässigen – reduzierten Gesamtstichprobenumfangs ausgeführt wurde, gilt in gleichem Maße auch für ein teilmarktspezifisches Vorgehen. In jedem Fall gibt es keinen Grund dafür, an eine teilmarktspezifische Anpassung andere statistische Anforderungen zu stellen als in Bezug auf eine einheitliche teilmarktübergreifende Anpassung. Solange keine entsprechenden rechtlichen Vorgaben bestehen, ist daher zu empfehlen, die vorgeschlagene Näherungsformel (Gleichung (4.3)) auch zur Bemessung der teilmarktspezifischen (Mindest-) Stichprobenumfänge anzuwenden. Sind nun aber Teilmärkte zu schwach besetzt, kommt für sie faktisch nur eine Vollerhebung infrage. Auch wenn dadurch das mittlere Mietniveau im Anpassungszeitpunkt exakt beziffert werden kann, ist es gleichwohl nicht nur möglich, sondern sogar wahrscheinlich, dass die der Näherungsformel zugrunde liegenden normativen Vorgaben nicht erfüllt werden können, zumindest dann nicht, wenn beim anzupassenden Mietspiegel der

[48] Beispielsweise sind Angebotsmietdatenbanken davon geprägt, dass nur online inserierte Wohnungsangebote und ggf. Inserate bei kooperierenden Zeitungen erfasst werden. Jedoch werden substanzielle Anteile des Mietwohnungsmarktes nicht online inseriert und es ist von einer systematischen Verzerrung der erfassten Angebote auszugehen. Zudem wird der mietspiegelrelevante Anteil der Mietverhältnisse, bei denen keine Neuvermietung vorliegt, sondern nur die Miete innerhalb der letzten sechs Jahre angepasst wurde, durch diese Datenquelle nicht abgebildet.

[49] So könnte es etwa sein, dass in dem vorherigen Mietspiegel der größte Mietniveauunterschied zwischen einem sehr selten auftretenden Substandard-Segment (z.B. mit unzeitgemäßem Heiz- oder Sanitärstandard) und den übrigen Wohnungen festzustellen ist. Häufig sind auch sehr große Mietniveauunterschiede zwischen Neubau- und Bestandswohnungen zu beobachten, wobei die Neubaufallzahlen nicht selten sehr gering sind.

entsprechende Teilmarkt nicht voll erhoben wurde und daher eine entsprechend große Schätzungenauigkeit in Bezug auf das mittlere Mietniveau besteht.

Eine weitere erhebungspraktische Schwierigkeit der teilmarktspezifischen Anpassung ergibt sich daraus, dass die Sicherstellung eines nach Gleichung (4.3) ermittelten teilmarktspezifischen (Mindest-) Stichprobenumfangs faktisch nur durch eine (Ex ante-) Schichtung nach Teilmärkten zu erreichen ist. Hierzu ist es allerdings erforderlich, die Auswahlgrundlage nach den betreffenden Teilmärkten aufzuteilen, was wiederum voraussetzt, dass für alle Fälle der Auswahlgrundlage die Teilmarkt- und damit Schichtenzugehörigkeit bekannt ist. Je nach dem zur Teilmarktbildung herangezogenen Kriterium wird diese Voraussetzung jedoch häufig nicht erfüllt sein.[50]

Schlussendlich sind zur teilmarktspezifischen Anwendung der Näherungsformel teilmarktspezifische Informationen zu den Nettostichprobenumfängen der anzupassenden Erhebung einschließlich teilmarktspezifischer Mittelwerte und Stichprobenvarianzen erforderlich. Aufgrund dessen ist anzuraten, bereits bei der Neuerstellung eines Mietspiegels eine später möglicherweise im Raum stehende teilmarktspezifische Anpassung mitzudenken und dafür Sorge zu tragen, dass teilmarktspezifische Fallzahlen, Mittelwerte und Stichprobenvarianzen für mögliche Teilmarktbildungen mitberechnet werden. Andernfalls muss die datentechnische Möglichkeit bestehen, die betreffenden Informationen nachträglich zu beschaffen.

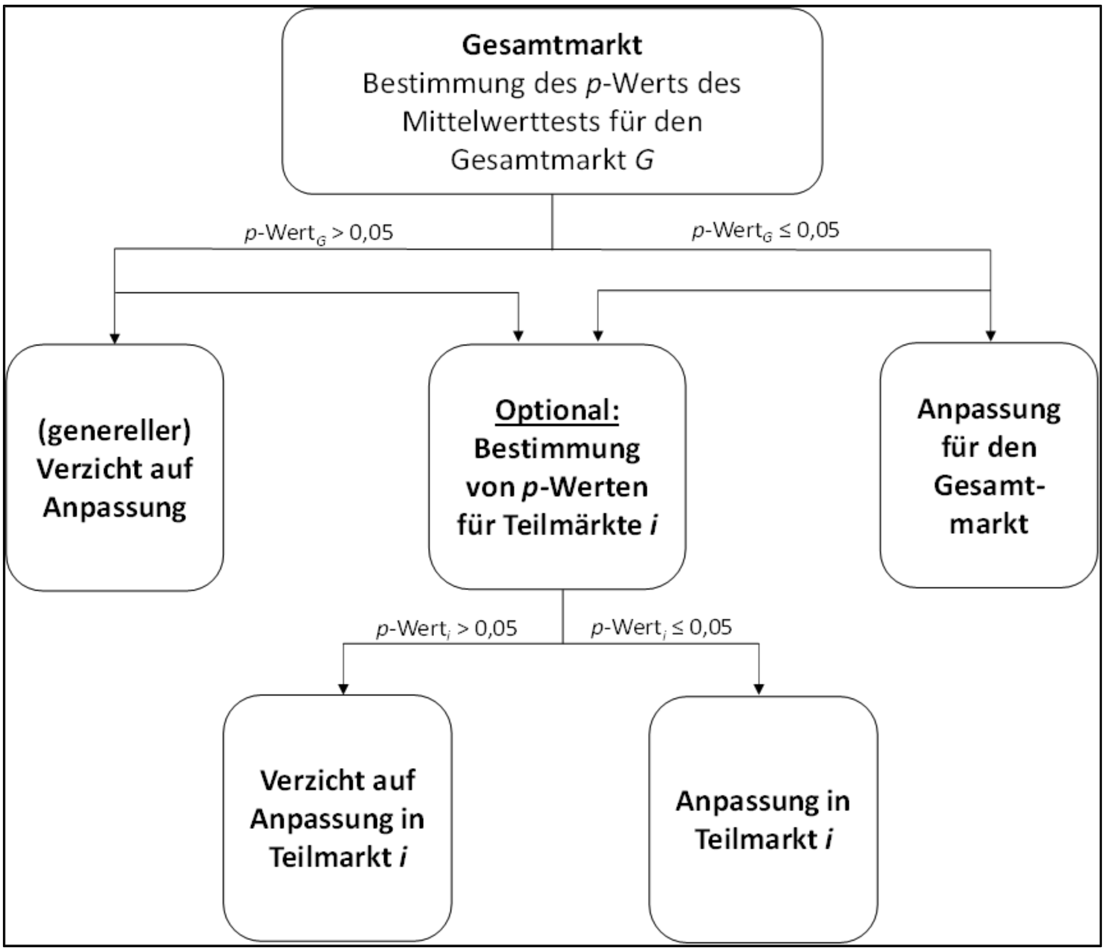
Zu beachten ist, dass die sich aus Formel (4.3) ergebenden teilmarktspezifischen Mindeststichprobenumfänge nicht nur für sich genommen, sondern sogar aufaddiert unter Umständen geringer sein können als der Mindeststichprobenumfang, der sich beim Verzicht auf eine teilmarktspezifische Herangehensweise ergibt. Denn sind die teilmarktspezifischen Stichprobenvarianzen aufgrund der Homogenität der gebildeten Teilmärkte sehr niedrig, wirkt sich dies unter sonst gleichen Bedingungen standardfehlersenkend aus, was wiederum die bei Mittelwerttests inhärenten Wahrscheinlichkeiten für Fehler 1. und 2. Art senkt bzw. die Teststärke erhöht. Diese Überlegungen stützen daher die bereits aus inhaltlichen Erwägungen abgeleitete Empfehlung, im Falle eines teilmarktspezifischen Vorgehens auf die Bildung möglichst homogener Teilmärkte zu achten.

[50] Bezogen auf das in der vorangegangenen Fußnote ausgeführte Beispiel des Substandard-Segments wird es in der Praxis mangels einschlägiger Informationen kaum möglich sein, die Auswahlgrundlage danach zu schichten, ob eine Zugehörigkeit zum Substandard-Segment besteht.

Wie bereits an früherer Stelle erläutert, gibt es keine Gewähr dafür, dass stichprobenbasierte Mittelwerttests anschlagen, also gemessene Mittelwertunterschiede als statistisch signifikant qualifizieren. Das gilt auch dann, wenn die Stichprobenumfänge nach Formel (4.3) bestimmt werden. Folglich kann sich die Feststellung statistisch signifikanter Unterschiede über Mittelwerttests danach unterscheiden, ob solche Tests für einen Teilmarkt oder für den Gesamtmarkt durchgeführt werden. Möglich ist, dass für den Gesamtmarkt statistisch signifikante Unterschiede im Vergleich zur vorangegangenen Erhebung festgestellt werden, diese sich bei einer teilmarktspezifischen Betrachtung aber nicht zeigen. Ein solches Ergebnis ist insbesondere dann zu erwarten, wenn der Gesamtmarkt nicht wesentlich heterogener als die Teilmärkte ist, denn dann „leiden" teilmarktspezifische Mittelwerttests unter ihren vergleichsweise kleinen Stichprobenfallzahlen, wodurch die Feststellung statistisch signifikanter Unterschiede erschwert wird. Denkbar ist aber auch der umgekehrte Fall, wonach nur bei teilmarktspezifischer Betrachtung die erwähnten Mittelwerttests anschlagen, bezogen auf den Gesamtmarkt dagegen kein statistisch signifikanter Unterschied zwischen der mittleren Nettoquadratmetermiete der letzten Mietspiegelneuerstellung und der aktuellen Erhebung festgestellt werden kann. Zu einem solchen Befund kann es bei teilmarktspezifisch unterschiedlichen Mietpreisdynamiken kommen, die sich so stark voneinander unterscheiden, dass der standardfehlererhöhende Effekt geringerer teilmarktspezifischer Stichprobenumfänge überkompensiert wird. Zu empfehlen ist eine Vorgehensweise nach dem folgenden Schema (vgl. Abbildung 4.5).

Gelingt es demnach nicht, über den Mittelwerttest zum (hier unterstellten) Signifikanzniveau von 5 % für den Gesamtmarkt statistisch signifikante Unterschiede zwischen den zu vergleichenden Stichprobenmittelwerten (also der mittleren Nettoquadratmetermiete des anzupassenden Mietspiegels und derjenigen, die im Rahmen der Anpassungserhebung ermittelt wurde) festzustellen, kann es sich neben einem generellen Verzicht auf eine Anpassung anbieten, für Teilmärkte entsprechende Tests durchzuführen. Schlägt für einen Teilmarkt ein solcher Test an, weil der entsprechende p-Wert das Signifikanzniveau (hier: 5 %) nicht übersteigt, stünde für diesen Teilmarkt eine Anpassung an. Für Teilmärkte ohne angeschlagenen Mittelwerttest bleibt es dagegen bei einem Verzicht auf eine Anpassung.

Abbildung 4.5: Empfohlenes Schema zur Ableitung, ob eine Anpassung für den Gesamtmarkt oder für Teilmärke erfolgt



Deckt dagegen der Mittelwerttest für den Gesamtmarkt statistisch signifikante Unterschiede auf, ist es dennoch möglich zu prüfen, ob auch für die Teilmärkte statistisch signifikante Testergebnisse erzielt werden können. Gelingt dies für einen Teilmarkt, wird eine Anpassung unter Anwendung des für diesen Teilmarkt ermittelten Anpassungsfaktors vorgenommen. Schlägt der Mittelwerttest dagegen nicht an, unterbleibt für diesen Teilmarkt eine Anpassung – es sei denn, man rückt vom teilmarktspezifischen Vorgehen generell ab und entscheidet sich am Ende dann doch für die Anwendung eines einheitlichen Anpassungsfaktors für den Gesamtmarkt. Aus statistischer Sicht nicht zulässig ist dagegen eine in gewisser Weise gemischte Vorgehensweise, bei der in Teilmärkten mit statistisch signifikanten Mittelwertunterschieden teilmarktspezifisch angepasst wird, während in anderen Teilmärkten ersatzweise auf den Anpassungsfaktor für den Gesamtmarkt zurückgegriffen wird.

4.3.6 Anpassung mittels stichprobenbasierten Indexes oder durch Schätzung einer aktualisierten Regressionsgleichung?

Die bisherigen Ausführungen zu stichprobenbasierten Vorgehensweisen gingen von einer indexbasierten Anpassung aus, die auf Basis eines reduzierten Stichprobenumfangs erfolgen kann. Theoretisch wäre es auch möglich, auf der Grundlage der neu gezogenen Stichprobe eine Regressionsgleichung unter Einbeziehung derjenigen exogenen Variablen zu schätzen, deren Koeffizienten sich beim „alten" (Regressions-) Mietspiegel als statistisch signifikant von Null verschieden herausgestellt haben.

Ein solches Vorgehen ist allerdings nicht zu empfehlen. Erstens bedingt es, dass nicht nur einfach die Miete pro Quadratmeter erhoben werden kann, sondern der Fragebogen der letzten Neuerstellung im Wesentlichen erneut vollständig abgefragt werden muss. Zweitens ist zu bedenken, dass aufgrund eines reduzierten Stichprobenumfangs einige der ehemals statistisch signifikanten Wohnwertmerkmale unter Umständen nicht mehr in die neue finale Schätzgleichung aufgenommen werden können, weil sich die neu geschätzten Regressionskoeffizienten fallzahlbedingt nicht mehr als statistisch signifikant erweisen. Drittens erscheint daher bei einem solchen Vorgehen auch hinsichtlich des Stichprobenumfangs die Einhaltung der Qualitätsstandards einer kompletten Mietspiegelneuerstellung geboten.

Bei einer derartigen Neuerhebung handelt es sich dennoch um eine Anpassung und nicht um eine Neuerstellung, da das Modell zur Datenauswertung nicht neu bestimmt wird, sondern nur die Parameter des bisherigen Modells neu geschätzt werden. Dennoch entsteht durch die umfassende Datenerhebung erheblicher Aufwand. Dies durchkreuzt jedoch das mit der Entscheidung zugunsten einer Anpassung gegenüber einer Neuerstellung angestrebte Ziel einer Zeit- und Kostenersparnis.

Analog zu obigen Ausführungen sollte bei Anwendung der Tabellenanalyse die Berechnung der Struktur der Mietspiegeltabelle anhand der Stichprobe der Neuerhebung und nicht der reduzierten Stichprobe der Anpassung erfolgen.

4.4 Zusammenfassung

Mietspiegel sind grundsätzlich nur zwei Jahre gültig; danach sind diese der Marktentwicklung anzupassen. Aufgrund des hohen Kosten- und Zeitaufwandes für die Erstellung eines qualifizierten Mietspiegels besteht einmalig die Möglichkeit, nach zwei Jahren statt einer

Neuerstellung eine vereinfachte Anpassung an die Marktentwicklung vorzunehmen. Dabei kann nach § 22 MsV die Anpassung entsprechend der Inflationsentwicklung berechnet werden. Dies stellt eine sehr kostengünstige Methode dar, hat aber mehrere Nachteile. Zum einen besteht nahezu kein Bezug zur Entwicklung der ortsüblichen Vergleichsmiete in der betroffenen Gemeinde. Zum anderen ist nicht bekannt, wie groß der hierdurch entstehende Fehler in der Anpassung der Mietspiegel ist.

Eine genauere Abbildung der Entwicklung der Mietpreise kann durch eine Anpassung nach § 23 MsV erreicht werden. Hierzu werden im Rahmen einer Umfrage erneut Mietdaten erhoben, wobei der Stichprobenumfang kleiner als bei einer Neuerstellung ausfallen kann. Auch ist üblicherweise der Umfang der für eine Wohnung erhobenen Informationen deutlich geringer. Hierdurch fallen insgesamt deutlich niedrigere Kosten als bei einer Neuerstellung an. Bei der Anpassung mittels einer reduzierten Stichprobe sind aber eine Reihe von statistischen Fragestellungen zu berücksichtigen. Dieser Beitrag soll hierzu Handlungsempfehlungen bieten.

Zunächst ist festzulegen, ob Querschnitt- oder Längsschnittdaten erhoben werden sollen. Diesbezüglich empfehlen wir aus stichprobentheoretischen Gründen die Verwendung eines Querschnittsdatensatzes. Zudem ist zu entscheiden, ob überhaupt anzupassen ist oder ob der alte Mietspiegel unverändert fortgeschrieben werden kann. Hierfür schlagen wir die Anwendung eines zweiseitigen Welch´s-t-Test vor. Wesentlich ist auch die angestrebte Größe des Umfangs der bereinigten Nettostichprobe. Wir zeigen hierzu einen neu-en Ansatz auf, der mit wenigen normativen Setzungen (zu Signifikanzniveau, Effektstärke und Teststärke) die Berechnung des benötigten Stichprobenumfangs erlaubt. Hierzu müssen allerdings der Stichprobenumfang der bereinigten Nettostichprobe des anzupassen-den Mietspiegels sowie hieraus der Mittelwert und die Standardabweichung der Miete in Euro pro Quadratmeter bekannt sein. Zu entscheiden ist auch, ob eine für den Gesamt-markt einheitliche Anpassung vorzunehmen ist oder ob für mehrere Teilmärkte eine je-weils unterschiedliche Anpassung berechnet werden soll. Auch wenn die Differenzierung nach Teilmärkten zunächst eine genauere Anpassung in Aussicht stellt, ist diese Vorgehensweise mit einer Reihe von zusätzlichen Problemen verbunden. Es sollte sorgfältig abgewogen werden sollte, ob eine einheitliche Anpassung ggf. nicht die bessere Alternative darstellt.

Letztlich zeigen wir auf, dass bei einer reduzierten Stichprobe die Anpassung mittels einer Neuschätzung der Parameter der Regressionsgleichung nicht sinnvoll ist. Falls dies gewünscht wäre, müsste eine Stichprobe im gleichen Umfang wie bei einer Neuerstellung erhoben werden.

Gleiches gilt analog bei der Verwendung der Tabellenanalyse für die Berechnung der Struktur der Mietspiegeltabelle.

Für die vergleichende Bewertung von Anpassungsverfahren auf Basis von Querschnittsdaten gegenüber solchen auf Basis von Längsschnittdaten wären unter anderem weiterführende empirische Erkenntnisse über die Stabilität der Grundgesamtheit mietspiegelrelevanter Wohnungen über die Zeit hilfreich, wobei allerdings zu berücksichtigen ist, dass die Frage der Stabilität der Grundgesamtheit in unterschiedlichen Mietspiegelkommunen womöglich unterschiedlich zu beantworten ist. Darauf aufbauend könnte die Sensitivität der eingesetzten Verfahren auf Populationsveränderungen zum Beispiel durch Simulationsstudien besser beurteilbar gemacht werden, als es der Rahmen dieses Beitrags ermöglichte.

Literaturverzeichnis

Anderson, O., Popp, W., Schaffranek, M., Steinmetz, D., & Stenger, H. (1997). *Schätzen und Testen. Eine Einführung in Wahrscheinlichkeitsrechnung und schließende Statistik* (2. Aufl.). Berlin et al.: Springer.

Börstinghaus, U., & Clar, M. (2023). *Mietspiegelrecht: MsV (Mietspiegel-VO), BGB, EGBGB: Kommentar*. München: C.H. Beck.

Cischinsky, H., von Malottki, C., Rodenfels, M., & Vaché, M. (2014). „Repräsentativität“ im Mietspiegel – Stichprobenmethodische Anforderungen an qualifizierte und grundsicherungsrelevante Mietspiegel. *Wohnungswirtschaft und Mietrecht, 67*(5), 239–250.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale et al.: Erlbaum.

Harrison, D. A., & Brady, A. R. (2004). Sample size and power calculations using the noncentral t-distribution. *The Stata Journal, 4*(2), 142–153.

Kubinger, K. D., Rasch, D., & Moder, K. (2009). Zur Legende der Voraussetzungen des t-Tests für unabhängige Stichproben. *Psychologische Rundschau, 60*(1), 26–27.

Liu, X. S. (2014). *Statistical Power Analysis for the Social and Behavioral Sciences. Basic and Advanced Techniques*. New York: Routledge.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t test to departure from population normality. *Psychological Bulletin, 111*(2), 352–360.

Schlittgen, R. (2012). *Einführung in die Statistik. Analyse und Modellierung von Daten* (12. Aufl.). München: Oldenbourg Verlag.

Sebastian, S., Cischinsky, H., Krapp, M.-C., & Eisfeld, R.K. (2024). *Handlungsempfehlungen zur Erstellung von Mietspiegeln*. Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR).

Sebastian, S., & Range, J. (2022). *gif-Mietspiegelreport 2022 – Auswertung der Mietspiegel der zweihundert größten Städte Deutschlands*. Wiesbaden: gif-Mietspiegelkommission.

Stenger, H. (1986). *Stichproben*. Heidelberg, Wien: Physica-Verlag.

C. Appendix

C1 Herleitung von Formel (4.3)

Seien μ_1 und μ_2 die unbekannten Mittelwerte der zu vergleichenden Grundgesamtheiten (hier: die mittlere Nettoquadratmetermiete zum Zeitpunkt der Erstellung des anzupassenden Mietspiegels und zwei Jahre danach) sowie $\bar{Y}_1$ bzw. $\bar{Y}_2$ die zu deren Schätzung auf Basis von n_1 bzw. n_2 Fällen berechneten Stichprobenmittel. Unter den Annahmen der Normalverteilung des zugrundeliegenden Merkmals (also der Nettoquadratmetermieten) und dessen identischer Streuung in beiden Grundgesamtheiten $(\sigma_1 = \sigma_2 = \sigma)$[51] ist die aus Stichprobendaten berechenbare Testgröße

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{4.5}$$

mit der gepoolten Stichprobenstandardabweichung

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \tag{4.6}$$

bei Gültigkeit der Nullhypothese $H_0\colon \mu_1 = \mu_2$ (Student-) t-verteilt mit $\nu = n_1 + n_2 - 2$ Freiheitsgraden, wobei S_1^2 und S_2^2 die Stichprobenvarianzen kennzeichnen (vgl. Liu 2014: 37; Harrison/Brady 2004: 144).

Trifft die Nullhypothese dagegen nicht zu und ist damit $\mu_2 - \mu_1 = \Delta \neq 0$, folgt die obige Testgröße nicht einer um den Erwartungswert 0 zentrierten t-Verteilung, sondern stattdessen einer um den Nichtzentralitätsparameter λ zentrierten t-Verteilung mit

$$\lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\frac{\Delta}{\sigma} \tag{4.7}$$

(ähnlich Harrison/Brady, 2004, S. 144).

Im Fall der Ungültigkeit der Nullhypothese (und damit im Fall der Gültigkeit der Alternativhypothese) gilt für die Teststärke $1 - \beta$ somit

---

[51] Die vereinfachende Annahme einer identischen Merkmalsstreuung kann hier getroffen werden, da im weiteren Verlauf aufgrund des Umstandes, dass die Stichprobe zur Mietspiegelanpassung ja noch nicht vorliegt, mangels anderer Informationen ohnehin sinnvollerweise unterstellt wird, dass die (noch nicht feststehende) Stichprobenvarianz S_2^2 derjenigen der vorangegangenen Erhebung (S_1^2) entspricht.

$$1-\beta = W\left(\left\{T' \geq t_{1-\propto/2,\nu}\right\} \cup \left\{T' \leq t_{\propto/2,\nu}\right\}\right), \tag{4.8}$$

wobei β die Wahrscheinlichkeit für einen Fehler 2. Art, T' die um den Nichtzentralitätsparameter λ zentrierte, t-verteilte Prüfgröße, α das für den zweiseitigen Zweistichproben-t-Test zugrunde gelegte Signifikanzniveau und $t_{1-\propto/2,\nu}$ sowie $t_{\propto/2,\nu}$ Parameter der (um 0 zentrierten) t-Verteilung sind (vgl. Liu, 2014, S. 38).

Ist die Effektstärke Δ positiv, ist auch $\lambda > 0$. Das hat zur Folge, dass $W\left(\left\{T' \leq t_{\propto/2,\nu}\right\}\right)$ vernachlässigbar klein ist[52], wodurch man zunächst zu

$$1-\beta \approx W\left(\left\{T' \geq t_{1-\propto/2,\nu}\right\}\right) = W\left(\left\{T' - \lambda \geq t_{1-\propto/2,\nu} - \lambda\right\}\right) = W\left(\left\{T \geq t_{1-\propto/2,\nu} - \lambda\right\}\right) \tag{4.9}$$

und wegen der Symmetrie der t-Verteilung um die Ordinate zu

$$1-\beta \approx F\left(\lambda - t_{1-\alpha/2,\nu}\right) \tag{4.10}$$

gelangt, wobei F die Verteilungsfunktion der (um 0 zentrierten) t-Verteilung bezeichnet.

Damit hat man

$$t_{1-\beta,\nu} = \lambda - t_{1-\alpha/2,\nu} \approx \sqrt{\frac{n_1 n_2}{n_1+n_2}}\frac{\Delta}{\hat{\sigma}} - t_{1-\alpha/2,\nu} \tag{4.11}$$

Da $\hat{\sigma}$ unter anderem von der Stichprobenvarianz S_2^2 der Anpassungserhebung abhängt und diese zum Zeitpunkt der Entscheidung hinsichtlich des Stichprobenumfangs n_2 naturgemäß noch unbekannt ist, gelangt man unter der (sich angesichts der Annahme $\sigma_1 = \sigma_2 = \sigma$ aufdrängenden und angesichts der noch nicht durchgeführten Anpassungserhebung ohnehin „unausweichlichen“) Setzung $S_1^2 = S_2^2$ und nach Auflösung nach n_2 schließlich zur allgemeinen Variante von Formel (4.3), d.h. zu

$$n_2 \approx \frac{n_1 S_1^2\left(\frac{t_{1-\beta,\nu}+t_{1-\alpha/2,\nu}}{\Delta}\right)^2}{n_1 - S_1^2\left(\frac{t_{1-\beta,\nu}+t_{1-\alpha/2,\nu}}{\Delta}\right)^2} \tag{4.12}$$

Da nun $t_{1-\beta,\nu} \xrightarrow[\nu\to\infty]{} 0{,}842$ für $\beta = 0{,}2$ und da $t_{1-\propto/2,\nu} \xrightarrow[\nu\to\infty]{} 1{,}96$ für $\propto = 0{,}05$ ist, gelangt man durch Einsetzen dieser Werte und für $\Delta = 0{,}2$ zu Formel (4.3) im Haupttext. Zu erwähnen bleibt noch, dass bereits bei vergleichsweise geringen Freiheitsgraden ν Werte für $t_{0{,}8;\nu}$ und $t_{0{,}975;\nu}$

[52] Ist die Effektstärke Δ dagegen negativ, ist $W\left(\left\{T' \geq t_{1-\propto/2,\nu}\right\}\right)$ vernachlässigbar gering, wodurch sich an der nachfolgenden Darstellung im Ergebnis nichts ändert.

erreicht werden, die sehr nahe an 0,842 bzw. 1,96 heranreichen. Ist beispielsweise $\nu = 400$, hat man $t_{0,8;400} = 0{,}843$ und $t_{0,975;400} = 1{,}966$.

5 Conclusion

Operating real estate during the 2010s in Germany was a story of stability, growth and success. Linear trend extrapolation and risk blindness came to an end, when new global crises emerged at the beginning of the 2020s, namely the COVID-19 pandemic and the war in Ukraine. This thesis deals with how those crises affect real estate practice in respect to cities, real estate companies and policy making.

Most real estate in Germany is concentrated in urban areas. A persistent threat posed by a respiratory virus would ultimately challenge the operation of these properties. Chapter 2 analyses regional mortality rates with respect to settlement density in order to decompose the pandemic-related risks for urban development. Is there a positive correlation between settlement density and mortality during the pandemic? Paradoxically, mortality rates and settlement density are negatively correlated. Thus, in less densely populated areas, relatively more people died after an infection with SARS-CoV-2. However, this correlation becomes significantly positive once we control for health-related indicators. We therefore conclude that density is indeed a risk factor during pandemics.

How large is this potential density effect on mortality relative to a broad range of covariates, notably geographic location within Germany? Our analysis shows that this effect is relatively small compared with individual health factors, and amounts to roughly one third of the combined effect of longitude and latitude one year after the start of the vaccination campaign. But do these correlations between density and mortality vary across time and space? Our research indicates that densely populated places in Germany experienced more rapidly rising COVID-19 mortality than less densely populated places during periods when non-pharmaceutical interventions were introduced either too late or inefficiently. Additionally, the regions most affected by the novel coronavirus in Germany are those in the south, from where the virus initially entered. This pattern of regional dispersion remains evident even almost two years after the first wave of infections.

Our paper summarises much of the scientific discussion on urban density during the COVID-19 pandemic and provides guidance for researchers and policy-makers in comparable future scenarios. Cities are more vulnerable during pandemics than peripheral regions. Nevertheless, mortality during the pandemic was higher in rural areas than in urban ones. Therefore, the advantages of dense settlement structures may not be outweighed by regional sprawl, even under contemporary pandemic scenarios, once the necessity of rapid intervention is understood.

However, long-term effects of the COVID-19 pandemic onto urban development keeps scope for further research. Moreover, our results show that the fight against a pandemic cannot stop at administrative borders, as regional dispersion occurs readily in densely populated areas such as Central Europe. In retrospect, stronger international cooperation offers potential for improvement and could prevent deaths in the event of a future epidemic.

Just as the pandemic was considered to be over in spring 2022, real estate companies in the German-speaking countries were struck by a second shock: the Russian Federation launched a full-scale invasion of Ukraine, resulting in the displacement of millions of people, large-scale destruction there, and sharply rising consumer prices, particularly for energy. The European central bank reacted with an unprecedented rise of key reference interest rates, ending the real estate prices boom across Europe. Drawing on a primary survey of companies' revenue structures, Chapter 3 seeks to understand the price reactions of the stocks of listed real estate companies to this shock and to the earlier shock of the pandemic. For example, companies operating predominantly office buildings may be affected differently by the pandemic, which altered the ways in which people work together, then companies engaged in the logistics sector. We hypothesised that two dimensions can represent a company's business model and that these business models are affected differently by different kinds of shocks.

Can listed real estate companies be meaningfully segmented into clusters based on two pivotal activity dimensions that characterise a business model? We cluster the listed real estate industry along two dimensions—sector (asset class) and position in the value chain—into four distinct business models: mixed-portfolio holder, residential-portfolio holder, investor/asset manager, and developer. We then examine whether these isolated clusters exhibit discernible differences in their stock-price volatilities and how they reacted, in terms of volatility and price performance, during the two most recent shocks—namely, the COVID-19 pandemic and the interest-rate reversal. Using share prices and market capitalisation, we construct stock-price indices for the four groups of real estate stock companies and analyse these indices in terms of performance, resilience, and the timing of market reactions.

We observe meaningful differences between the stock market performance of the indexes and our results challenge some commonly held assumptions about the real estate industry. For instance, the group corresponding to the mixed-portfolio-holder business model exhibited the lowest volatility, whereas the residential-portfolio-holder cluster experienced the largest maximum drawdown among all groups. In English-speaking countries, these business models are most closely associated with REITs. However, the listed real estate industry in German-speaking countries is more diverse in terms of company type, with REITs representing only a minor share. Our research suggests that investors would benefit from examining the various business models of real estate stock companies in European real estate markets more closely, and that such an examination can be structured by combining the two dimensions—a company's sectoral (asset-class) focus and its position along the value chain.

In our approach, we first cluster similar companies and then analyse differences in stock-market performance. Based on our newly derived data on the revenue structures of real estate stock companies, future researchers might instead proceed in the opposite direction: they could first cluster companies by similarity in stock-market performance and then relate those clusters to companies' revenue structures. This approach could help validate and strengthen the initial grouping. Still, the availability of disaggregated data on company revenues remains a challenge.

Chapter 4 concludes with a focus on a topic specific to Germany: the sample-based updating of qualified *Mietspiegel*. It is crucial to emphasise that the choice of method for updating a *Mietspiegel* should not be driven by which approach produces the lowest rent increase, but rather by whether a more precise measurement of changes in the local comparative rent, obtained through a survey-based method, justifies the additional costs. This consideration became particularly salient in the wake of the inflationary shock following the outbreak of the war in Ukraine in early 2022, when rapid increases in consumer prices raised concerns about the adequacy of CPI-based adjustments in reflecting actual developments of local comparative rents.

A key methodological decision for practitioners is whether cross-sectional or longitudinal data should be used for sample-based *Mietspiegel* adjustments. From a sampling-theoretical perspective, we advocate for the use of a cross-sectional dataset, as it allows for a more reliable representation of current market conditions. Another question concerns whether the gross sample from the most recent full survey can be reused. Our analysis shows that, while the gross sample may inform initial expectations, applying an adjustment based on a reduced sample requires careful consideration of the net sample size and the statistical properties of the underlying data.

Another key issue is determining whether an adjustment is necessary at all or whether the previous *Mietspiegel* can be carried forward unchanged. To address this, we propose the application of a two-sided Welch's t-test, which provides a statistically grounded method to determine whether observed changes in the *Mietspiegel* warrant adjustment or whether existing values remain sufficiently accurate. Closely related is the question of how the sample size for the adjustment should be defined. We introduce a new approach that, with only a few normative assumptions regarding significance level, effect size, and statistical power, enables researchers to calculate the required sample size. The calculation requires knowledge of both the net sample size of the *Mietspiegel* being updated and, based on it, the (weighted) mean and standard deviation of rents in euros per square metre.

Further methodological considerations include whether adjustments should apply uniformly or vary according to submarkets. While differentiation by submarkets might initially appear to yield more precise adjustments, it introduces additional challenges, including increased sampling complexity and potential instability in submarket estimates. Practitioners should carefully weigh whether a uniform adjustment, despite its simplicity, may offer a more robust and practical solution.

Finally, the question arises whether adjustments should rely on a sample-based index or an updated regression equation. Our findings indicate that a reduced sample does not permit meaningful parameter re-estimation for a regression-based *Mietspiegel*. If a regression update is desired, researchers must collect a sample of comparable size to that required for constructing a new *Mietspiegel*. The same limitation applies to tabular analyses aimed at (re-)estimating the structure of a rent table (*Tabellenmietspiegel*).

In summary, careful methodological choices are essential to ensure that sample-based *Mietspiegel* adjustments provide a valid reflection of local market conditions. Cross-sectional data collection, a clear rationale for sample size determination, critical consideration of submarket differentiation, and the appropriate selection between index-based and sample-based adjustments collectively form the foundation for reliable updates. By applying these principles, practitioners can ensure that *Mietspiegel* adjustments remain statistically robust, cost-efficient, and legally defensible, even in the context of macroeconomic shocks such as inflation following the outbreak of the Ukraine war.

However, further research is required to gain a better understanding of how stable the population of dwellings relevant for *Mietspiegel* remains over time. Such knowledge could make two important contributions. First, it could advance the methodological debate about how best to adjust qualified *Mietspiegel*. Second, it could help to assess the impact of regulatory changes—for instance, changes in the reference period for *Mietspiegel* relevance. This task is challenging. The conditions under which a Mietspiegel is created—under a statutory duty to provide information—cannot easily be transferred to experimental or empirical studies. Because of these limitations, a simulation-based approach may be the most promising route.

Ultimately, real estate is designed for long-term horizons that will encompass periods of relative stability as well as historically diverse crises. Accordingly, the management of real estate assets must at all times take both conditions into account. This implies that potential risks should be considered well before the onset of any crisis. Especially during times of crisis, strategic and immediate action becomes essential. The speed at which households, companies, and policymakers respond to changing conditions often determines the severity of a crisis's impact. Early implementation of non-pharmaceutical interventions is important in this context, as is the resilience of business models in the real estate sector to sharply rising interest rates, along with the presence of methodologically sound regulatory frameworks for real estate markets in general and rental markets in particular. Consequently, the lessons learned from the crises of recent years may also help strengthen the resilience of cities, real estate companies, and the regulatory framework governing the rental housing market.