Assessing regional structures in German residential apartment yields

Johann Götz Norbert Trenkler

Abstract

This study presents an extensive analysis of residential apartment yields across all 401 German counties in 2019. Spatial disparities are uncovered whereby the focus lies on differences between eastern (formerly belonging to the German Democratic Republic) and western German counties and to a lesser extent on differences between rural and urban areas. The aim is to uncover regional determinants of yields. Linear regression models are estimated using regional economic data. The findings reveal significant differences in yields between Eastern and Western, but also between northern and southern German states. Even 30 years after reunifying, eastern German states show yields considerably above the national average. At the same time, yields in Bavaria, certain coastal areas and major urban regions are particularly low. This is associated with different characteristics of regional drivers such as contrasting levels of household income. The findings of this study may inform institutional and private investors.

Colofon

Document	Master's Thesis Real Estate Studies
Title	Assessing regional structures in German residential apartment yields
Author	Johann Götz Norbert Trenkler
Student number	S4515749
E-mail	j.g.n.trenkler@student.rug.nl
Primary supervisor	Dr. M.N. (Michiel) Daams
Secondary supervisor	Dr. F.J. (Frans) Sijtsma
	University of Groningen Faculty of Spatial Sciences Master's Program in Real Estate Studies
Date	April 13, 2022

Contents

1. Introduction 5
   1.1. General Introduction 5
   1.2. Study area: Historical context to price-to-rent ratios in Germany 8

2. Materials and Methods 9
   2.1. Methodology 9
   2.2. Data 11

3. Results 15
   3.1. Differences in yields across space 15
   3.2. Differences in drivers of yields 24

4. Discussion 26

5. Conclusion 28

6. References 29

A. Further statistical analyses 32

B. Tables 34

C. Figures 37

D. Stata Codes 40

List of Tables

1. Grouped Variables . . . 12
2. Summary statistics of all variables . . . 15
3. Bottom and top 20 yields . . . 18
4. Means of net rents, dwelling prices and yields for old and new states . . 19
5. Descriptive statistics for apartment yields grouped by city size . . . 21
6. Estimation results involving `newold`, `lvacancy` and `laverageincome` . . 26
A.7. Estimation output for model (3) . . . 32
B.1. Two-sample t-test for equal means of yields in old and new states . . . 34
B.2. One-way ANOVA test for equal yields in different cities . . . 35
B.3. Variance inflation factors resulting from estimating (3) . . . 35
B.4. Breusch-Pagan test for constant error variances after estimating (3), . . 36
B.5. Re-estimating (3) using robust standard errors. . . . 36
B.6. Ramsey's RESET test for omitted variables after estimating (3) . . . 36
B.7. Shapiro-Wilk test for normality of disturbances . . . 36

List of Figures

1. Macro variables, residential property and rent index 2000–2019 . . . 8
2. Residential apartment yields in Germany . . . 16
3. Logarithm of yields distinguished by federal state. . . . 17
4. Yields sorted by size of the city. . . . 20
5. Yields in Urban versus Rural areas. . . . 21
6. Pairwise scatter plots of log variables and log yields . . . 23
7. Pairwise scatter plots and boxplots of exogenous variables . . . 25
C.1. Pairwise scatter plots . . . 37
C.2. Differences in logarithm of yields in old versus new federal states . . . 38
C.3. Distribution of yields in old versus new states . . . 39
C.4. Residual vs fit plot . . . 39
C.5. Histogram of OLS residuals . . . 40

1. Introduction

1.1. General Introduction

Residential real estate yields are used by institutional and private investors to assess the profitability of their investments. Yields are a reflection of both rents and housing prices and provide rent-to-price ratios (Wheaton and Nechayev, 2005). According to Baum (2015, p. 128), yields are defined as net operating income divided by the current value or purchase price. As for Germany, these two drivers are thoroughly scrutinized by current research, with public interest being high as changes in these drivers affect the livelihoods of many people. At the same time, however, when it comes to assessing spatial differences in yield structures and how investors value influential drivers the literature available appears to be scarce. It begs the question to what extent characteristics of different regions are priced into yields.

First, regions vary in their level of economic development. Some regions are more prosperous than others. Belke and Keil (2018) find that there is considerable variation in housing prices between eastern and western Germany as well as between cities and rural areas. This has implications for yields, which will be discovered later in this analysis. Likewise, Möbert (2021) point out that for the most part of eastern Germany (except for the state of Brandenburg in direct proximity to Berlin), population has declined over the decades since reunification. At the same time, there has been a positive net migration towards the economically prosperous regions. This may lead to lower yields, because higher demand implies higher prices for apartments.

Regarding the scientific literature, numerous publications are assessing the impact of drivers of house prices and rents in Germany over time. Algieri (2013) analyzes the main drivers of housing prices in Germany and several other European countries from 1970 to 2010. Moreover, Belke and Keil (2018) uncover determinants of real estate prices covering a huge selection of about 100 cities. Likewise, Kholodilin et al. (2018) analyze house price increases in Germany during the time period 1990–2013. On the other hand, Egner and Grabietz (2018) apply time series analysis to assess drivers of rents and find that housing markets differ to a great extent across the country. These papers, whilst pointing out the staggering increases in rents and prices over time, do not conduct analyses on a regional, county level. At the same time, research has been conducted assessing yields in its own respect. Chichernea et al. (2008), whose research is a key component of this paper, conduct seminal work on cap rates (yields). They analyze U.S. housing markets aiming at identifying the main underlying drivers. Similarly, Wheaton and Nechayev (2005) scrutinize fundamentals of properties in the U.S. and assess how these are incorporated in prices.

This paper aims at providing a better understanding of cross-sectional differences in yields across German regions in a particular year. Possible determinants are primarily drawn from the literature , and a wide selection of county level data has been made accessible. The central research question is as follows:

Does the East-West divide in yields remain after controlling for yield drivers?

Apartment yields for the year 2019 with a uniform apartment size of 70 square meters are observed. These equally sized apartments ensure the observed properties to be of relative constant quality. However, the comparable literature analyzes house prices and yields mostly over time (longitudinally) whereas this work focuses on the cross section (Chichernea et al., 2008; Belke and Keil, 2018). Likewise, other studies focus predominantly on estimating house prices while on the contrary this thesis is centered at apartment yields. Hence, the underlying drivers may vary. German housing markets are at least partially touched upon in most scientific literature that was assessed in this paper.

Next, rural-urban disparities are addressed. Kajuth et al. (2013) assert that apartment price determinants differ greatly across regions. Firstly, rural-urban migration and urbanization play an important role (Zhang et al., 2012). While rural areas across the country have seen a decline in population, the opposite can be said for urban areas, where population and economic growth is concentrated (World Bank Group, 2017). In fact, the rural-urban divide is large compared to the OECD average where Germany ranks 7th in terms of GDP per capita disparities among small regions (OECD, 2019). Following this, during the 2000s an increasing migration of low-income individuals from rural to more expensive urban areas was observed (Dustmann et al., 2018). Driven by increased demand, yields may decrease in the more densely populated regions, and vice versa.

Moreover, the East-West economic disparity that persists more than three decades after Germany's reunification is possibly very important in understanding differences in housing yields. Evidently, what followed the downfall of the wall was large scale migration from the former East to Western Germany (Redding and Sturm, 2005). Although, as Dustmann et al. (2018) point out, housing prices did increase substantially driven by large investments to modernize the housing stock, to the present moment they have not caught up to similar levels observed in Western Germany. Berlin played a salient role in that process in that investors appear to have overestimated its potential. According to Holtemöller and Schulz (2010), property price increases during the period 1980–2004 were the result of misjudgment and not backed by fundamentals.

Furthermore, in Section 2.1, certain macroeconomic indicators are compared graphically to a rent and house price index over a period of 19 years. However, the key element of this research comprises of a cross-sectional analysis of the state of the housing market in 2019 in which several variables for all 401 German counties are compiled to compare regions. These include, but are not limited to population, net rent, population density and unemployment.

Likewise, this analysis revolves around the year 2019, as it resembles the most current state of the housing market before the Covid-19 pandemic impacted real estate markets in Germany and across the world. As the pandemic is still progressing to date, its effect on housing markets cannot be properly assessed. The time span between 2000 and 2019 is of particular interest, because unlike other major European countries, the German housing market did not experience a boom at the start of the 2000s and until after the Great Financial Crisis of 2008–2009 when housing prices rapidly rose, particularly in urban areas (Kholodilin et al., 2018). However, the central research of this thesis comprises of several OLS estimations of numerous regional level determinants on a county level. These are collected from public and private sources to provide detailed answers to the central research question of this study. It must be highlighted that throughout this thesis the term yield refers to gross yields, which is in accordance with the data at disposal. A distinction between gross and net yields, as in Hargreaves (2005), was not made. The findings aim to add to the literature by unveiling differences in yield structures between eastern and western Germany.

The remainder of this thesis is organized as follows: Section 1.2 provides historical context to current residential apartment yields in Germany. Section 2 outlines methodologies used for this research and corresponding data. In Section 3, results regarding the distribution of yields as well as regression results examining the underlying drivers of yields are presented. Section 4 provides a discussion of the results and positions the findings in the broader literature. Finally, in Section 5 a conclusion is presented bringing together the main insights of this research.

1.2. Study area: Historical context to price-to-rent ratios in Germany

To understand how yield structures for the year 2019 came about, this study provides a brief historical context. Kholodilin et al. (2018) point out that in recent years price-to-rent ratios have increased substantially, which begs the need for briefly assessing the underlying macro conditions over time. Figure 1 depicts mean annual base interest rate as stipulated by the ECB (Arestis and González, 2014; Kholodilin et al., 2018), development of the DAX score as a proxy for stock market activity (Algieri, 2013; Belke and Keil, 2018) and per capita GDP as a modification of GDP (Igan and Loungani, 2012; Zhang et al., 2012; Goodhart and Hofmann, 2008). These are compared to a national residential property price index as a proxy for apartment price developments (Igan and Loungani, 2012; Zhang et al., 2012; Goodhart and Hofmann, 2008) as well as a national rent index (Kholodilin et al., 2018; Zhang et al., 2012).

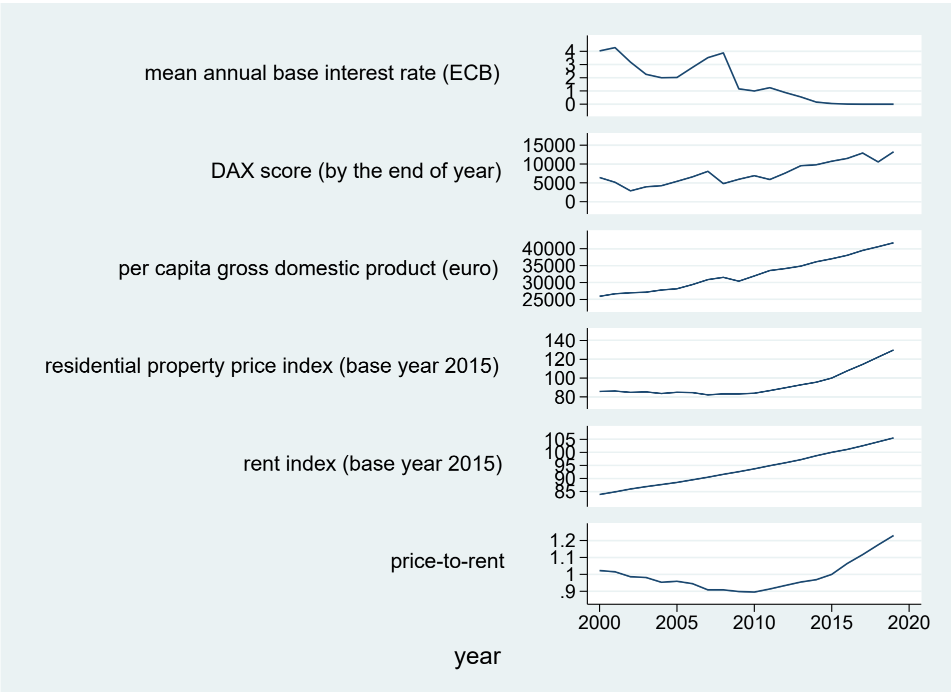


Figure 1: Macro variables, residential property and rent index 2000–2019. Stata code: `mtjt01.do`, page 40.

As depicted in Figure 1, rents have increased steadily over time whereas property prices only started seeing a steep upswing around the year 2010. Remarkably, this roughly concurred with the ECB's gradual lowering of the base interest rate to its historically low level of 0% in 2016 where it stabilized for the remaining years, driving up investments into alternative assets such as Real Estate. Likewise, the Financial Stability Review issued by the ECB (European Central Bank, 2021) point out that investors react strongly to changes in monetary policy and adapt their behavior. House prices rising more rapidly than rents imply that yields are decreasing which leads to riskier parts of the fund sector expanding. Announcements of expansionary monetary policy (as happened after the GFC) leads to asset managers reducing their cash holdings but to invest more into alternative assets such as apartments, which are generating higher returns at the expense of coming alongside higher risks (European Central Bank, 2021).

Moreover, German share values (measured by DAX) as well as per capita GDP rose considerably during the period under review. The literature agrees that strong growth in those variables drives up house and apartment prices. More broadly speaking, low base interest rates and higher purchasing power accelerate investments into Real Estate and hence drive up their prices. This can be assessed as one reason why apartment prices across Germany rose substantially during the last decade.

2. Materials and Methods

2.1. Methodology

The primary goal of this papers' analysis is to better understand regional differences in yields across all German counties by analyzing their underlying determinants. The first part revolves around univariate analyses of the yield structures across space in Germany. In a second step, several hedonic models closely related to Chichernea et al. (2008) are derived from the literature and from available data. Their general form is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1} + \cdots + \beta_K x_{iK} + \varepsilon_i\,, \quad i = 1, \ldots, 401\,, \tag{1}$$

relating the yield y_i of city or county i to influential exogenous variables $x_{i1}, \ldots, x_{iK}$ such as unemployment, GDP per person, homeownership rate and so forth. Furthermore, β_0 is the intercept, $\beta_1, \ldots, \beta_K$ measure the influence of the variables on the yield and ε_i is the error term.

Variables to be included are mostly derived from the literature in order to assess the most suitable drivers. They are extracted from papers assessing rents and prices in Germany (Belke and Keil, 2018; Algieri, 2013) but also international papers concerned with similar analyses (Zhang et al., 2012; Chichernea et al., 2008), Furthermore, in order to assess the most promising variables to be included, scatter plots are carefully scrutinized. Apartment specific characteristics are not covered, which is a limitation of the model.

The preliminary model used in the analysis will be as follows:

$$\begin{aligned}\texttt{lyield}_i = \beta_0 + \beta_1\texttt{laverageincome}_i + \beta_2\texttt{lpop2019}_i + \beta_3\texttt{larea}_i \\ + \beta_4\texttt{lhomeownershiprate}_i + \beta_5\texttt{lvacancy}_i + \beta_6\texttt{newold}_i + \varepsilon_i\,, \\ i = 1, \ldots, 401\,. \quad (2)\end{aligned}$$

As described in equation (2), the model is composed of six exogenous short run variables, each of which refer to the year 2019 and are briefly described here. The first exogenous variable to be included is `laverageincome`, that is the logarithm of the average disposable household income. It highlights differences between economically prosperous and unprosperous counties. Furthermore, the second variable included is the logarithm of population size in 2019 (`lpop2019`), which assigns the number of inhabitants to a given county to illustrate spatial density. In addition, the third variable logarithm of area (`larea`) depicts the total area of a given county in square kilometers to account for spatial differences. Likewise, home ownership rates (`homeownershiprate`) are a crucial element of the analysis and are defined by Empirica Institute in the following fashion: The ownership rate is the share of owner-occupied dwellings in all dwellings in residential and nonresidential buildings. The remaining share is accounted for by rented apartments. In the model, however, logarithmic transformations are used for almost all variables to mitigate the effects of skewness and to ease interpretation of the beta coefficients as elasticities (Heij et al., 2004, p. 296). Furthermore, the variable `lvacancy` depicts logarithmic vacancy rates, i.e. the share of apartments that are unoccupied. Lastly, a binary variable named `newold` is added with the aim of pinpointing differences between federal states existing before (old states) and states that became part of Germany (new states) after its reunification in 1990.

The set of exogenous variables entering the model is primarily guided by Chichernea et al. (2008) and Wheaton and Nechayev (2005), whose research are centered around similar questions examining housing markets in the United States. Likewise, and to a lesser extent variables were selected based on the work of Egner and Grabietz (2018), Belke and Keil (2018) as well as Voigtländer (2009) who are focusing on German residential housing markets. With the aim of estimating yields across Germany, the variables in the model are based on the literature to account for the predefined groups which are further specified in Section 2.2. These cover supply and demand, population as well as socioeconomical aspects and geographical context.

As outlined above, rents and prices are key components of yields. They are driven by regional and economic characteristics (captured in the exogenous variables) and are assumed to vary profoundly across space. Therefore, the conjecture is that scarcely populated areas with weak economic performance are linked to lower housing values driving yields up, and vice versa. In addition, to get a better understanding of yields and in preparation for the model building, numerous pairwise scatter plots are created regressing variables to yields in which they are distinguished between Eastern and Western federal states. It is expected that economically prosperous regions with higher demand exhibit lower yields, and vice versa.

2.2. Data

First of all, a rich set of data has been compiled. However, only a small subsample of relevant variables is considered throughout the analysis. In the first part of the results section, data visualizations are made to showcase the regional nature of the data. These include mapping the data, scatter plots as well as box and whisker plots. On top of that, subsamples will be estimated to specifically point out regional differences in the yields. These are primarily differences between old and new federal states, and to a lesser extent urban rural divide. Furthermore, cities are grouped by size.

This study builds upon a set of housing price data for the year 2019, including all 401 German counties. The dataset is not publicly available but has been made accessible to the author directly by a German financial services institute called *Deutsche Postbank*. *Deutsche Postbank*, in its turn, relies on data from the *Federal German Statistical Office*, *Michael Bauer Research GmbH*, *Empirica* and *Hamburgisches WeltWirtschaftsInstitut (HWWI)*. Furthermore, most of the data concerning influential variables used for the hedonic models are extracted from a publicly available data source named *Regionaldatenbank Deutschland* which is part of the federal and state statistical offices in Germany. However, some of the data was made available by empirica regio GmbH under the condition of being treated confidentially. The collected variables are then merged into one dataset that is thoroughly examined using `STATA`. The corresponding data for the timeline in Figure 1, p. 8, is extracted from Statista using a paid for premium access. The data obtained allow for a thorough analysis at the regional depth of counties and independent cities.

For simplicity, the drivers for houses and apartments are assumed to be very similar in this research. Furthermore, this paper is based solely upon applicable data. Several determinants such as building regulations (Belke and Keil, 2018) or whether apartments are owner-occupied or rented out (Hill and Syed, 2016) are unavailable and hence not subject to further research. All variables available to this research are first grouped (Table 1), whereby variables included in the model are underlined. This paper builds up on seminal work conducted by (Egner and Grabietz, 2018; Belke and Keil, 2018). The former suggest a fourfold distinction into supply side, population, socio-economic attributes and city context variables whereas the latter propose a twofold distinction into supply as well as demand related variables. On this basis, the two approaches are merged to partition the following five groups: supply, demand, population as well as socio-economic variables and a fifth group. The fifth group is reformulated into geographical context. That is because the dummy variable aimed at highlighting differences between old and new federal states is an integral part of this study.

Table 1: Grouped Variables

Supply side	Demand side	Population	Socioeconomic attributes	Geographical context
area, construction activities, building completions, building permissions	**vacancy**, **homeownershiprate**, real purchase price increase, netrent, estimated cost of the structure, net business registrations, buildinglandprices	**population in 2019**, net migration, population density	**average income**, gdp per person, unemployment	**new and old states**, federal states

In addition, the model subject to this research covers one (for the supply side variables two) selected variable(s) from each of the five predefined groups. Choosing the variables, the author carefully scrutinized pairwise scatter plots to assess the structure in yields. At the same time, a dummy variable is created for new and old states (0 = old state, 1 = new state) to pinpoint regional differences between eastern and western federal states. Although the grouping is thematically exhaustive, the groups are not mutually exclusive. For example, the share of students depicts the population, but it could also belong to socioeconomic attributes. Attention is also drawn to making sure the variables are not interrelated to avoid multicollinearity and that only the most promising variable for each group is chosen. With the aim of the model to be straightforward and comprehensible, the author deliberately chose to include few variables with high explanatory power. However, it needs to be pointed out that the chosen model makes no claim to completeness. More elaborate models could unveil more promising results but are beyond the scope of this research.

To begin with, vacancy rates constitute an important demand side variable. Wheaton and Nechayev (2005) draw attention to their effect on attainable rents and Chichernea et al. (2008) point out their relevance in assessing cap rates (which are a proxy for yields). If the number of supplied apartments exceed demand, vacancies increase, implying lower demand. Lower demand in its turn reduces rents and apartment prices. If vacancies increase, rents are expected to decline (Wheaton and Nechayev, 2005) but presumably to a lesser extent than market prices. On the other hand, changes in supply and demand (expressed through vacancies) have a more profound impact on prices (Wheaton, 1990). Therefore, a positive sign of the estimated coefficient can be expected. *Empirica Institute* provides data for vacancy rates, which they define as follows: Market-active vacancies of apartments in multifamily buildings that are immediately disposable, as well as vacant apartments that are not currently available for rent due to defects but could be capitalized within six months.

Furthermore, as key demand related variable, a variable capturing home ownership rate has been added. Home ownership rate is the proportion of owner-occupied housing which is strikingly low in Germany compared to other European countries (Voigtländer, 2009). Data for home ownership rates have been collected on a county level. Including them into the model may provide meaningful insights, as they affect rents and thereby yields (Voigtländer, 2009; Egner and Grabietz, 2018). A positive sign of the estimated coefficient can be expected, because the higher homeownership, the more saturated a market becomes and the less demand there is for houses, slowing down price increases.

Regarding the population group, the variable population density is approximated by including its main underlying components, that are population size and area into the model. The use of population is comprehensively underpinned in the literature. This allows for higher flexibility when interpreting the results, because the two parameters are estimated separately. When population increases, the demand for apartments goes up. According to the literature, population size positively affects both apartment prices (Kholodilin et al., 2018; Kajuth et al., 2013) as well as rents (Egner and Grabietz, 2018). When it comes to yields, however, Chichernea et al. (2008) determine the effect of population growth (as a proxy for population size) on yields and are expecting a negative sign, which implies that the effect of housing prices would predominate over rents. For area, a positive sign is expected, because the higher the underlying area, the less densely it is usually populated implying weaker demand and hence lower underlying prices.

Moreover, pertaining to the fourth group of socioeconomic variables, average disposable household income is added. The literature agrees that this constitutes a key metric in assessing housing prices (Algieri, 2013; Belke and Keil, 2018; Arestis and González, 2014; Igan and Loungani, 2012; Zhang et al., 2012) and rents (Egner and Grabietz, 2018). The expected sign of the estimated coefficient is negative. Higher income leads to higher purchasing power and hence higher demand for apartments. Not only do apartment values increase, also land prices go up. For economically prosperous German regions this means that lower yields are to be expected, and vice versa.

The remaining variable `newold` represents a dummy variable for new versus old federal states, as explained above (new federal eastern German states added after the reunification are assigned the value `1`, preexisting federal states are given the value `0`). Following Chichernea et al. (2008) and Möbert (2021), it was created by the author to be able to better account for regional differences in yields between the east and the west. It is expected that even after more than three decades following the reunification eastern German states would exhibit significantly higher yields than their western German counterparts due to lower underlying house prices (and an overall development that trails western German states). Therefore, a positive sign is expected, because on average new states are expected to exhibit higher yields. Table 2 summarizes all variables included in the model.

Likewise, Figure C.1, p. 37 in the appendix provides further graphical insights into mutual dependencies of the variables included in the model, especially to unveil causes of potential multicollinearity which may be faced in estimating equation (2). In the following section 3 the estimation results will be scrutinized by means of interpretation of coefficients, variance inflation factors, RESET tests and so forth. However, one must bear in mind that the model only estimates yields of apartments with a certain uniform size, which is clearly a limitation of the model. Nonetheless, uniformity also means that the apartments under investigation are similar in their characteristics which means that they can be easily compared with one another. Furthermore, Chichernea et al. (2008) point out that yields are majorly impacted by

Table 2: (Edited) summary statistics of all variables with corresponding Stata-code.

```
run C:\mtjt\Stata\mtjt.do\ProvideData.do
summarize yield averageincome pop2019 area homeownershiprate vacancy newold
```

Variable	Obs	Mean	Std. Dev.	Min	Max
yield	401	.0443963	.0116222	.0139111	.0886919
averageincome	401	47865.92	6721.364	34629.4	74673.07
pop2019	401	207030.5	243880.3	34209	3644826
area	401	891.7051	724.168	35.7	5495.6
homeownershiprate	401	49.10798	13.3144	12	72.5
vacancy	401	3.588529	2.591065	.2	13.3
newold	401	.1895262	.3924156	0	1

supply constraints. Unfortunately, no data was available to reflect on stringency of building regulation as well as household composition on a county level in Germany.

Lastly, it is important to mention that a select few yield values appear to be erroneous. For example, the underlying dwelling prices obtained for the towns Aurich and Leer at 4239 €/sqm and 3703 €/sqm which are situated in the (structurally weak) northwestern German region of Ostfriesland seem unreasonably high and do not appear plausible. Indeed, reexamining the values on German real estate portal *Immowelt* provides different square meter values for dwellings of similar sizes, valued at 2660 €/sqm and 2200 €/sqm respectively. Outliers observed in that region must therefore be assessed with caution. However, as these erroneous observations are very limited to a few observations this does not imply an overall problem in the data set.

3. Results

3.1. Differences in yields across space

In this section, yield differences across regions are unveiled. Figure 2 provides a yield map for all German counties whereby a fourfold distinction of yields by quartiles is applied. The former border separating old and new states has been inserted (black color). Moreover, the map might visually suggest that yields are typically higher in new compared to old federal states (except for Berlin). These findings are undermined by the box and whisker plot provided in Figure C.2, p. 38, which makes a distinction between old and new federal states and suggests that yields in eastern Germany are significantly higher on average. It implies that the underlying purchase prices of houses in eastern Germany may be lower.

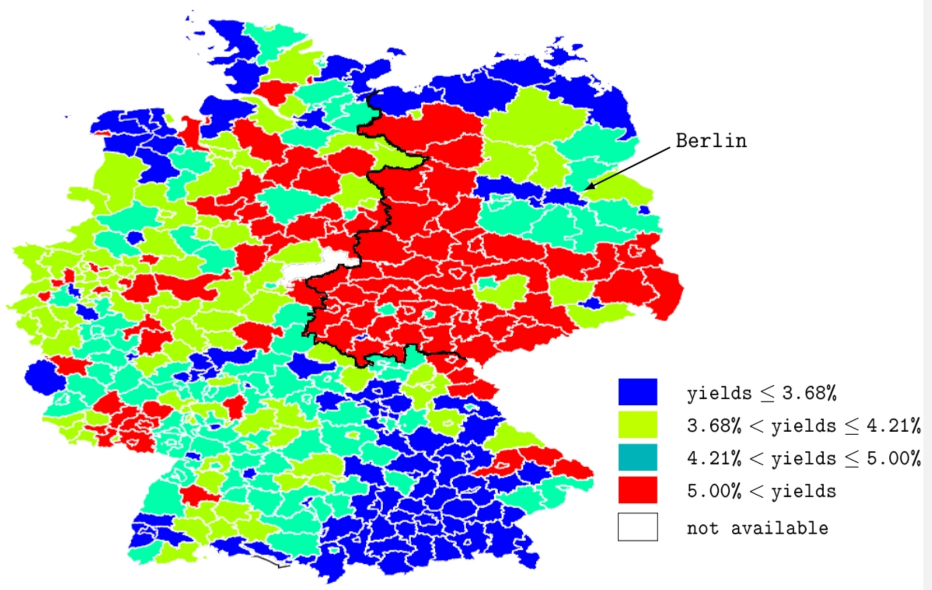


Figure 2: Residential apartment yields in Germany in 2019 for apartments with a uniform size of 70 sqm. Stata code: `mtjt03.do`, page 42 (output edited).

Figures 2 and C.2 suggest that for the most part yields are different in old versus new federal states. To assess this impression statistically a two-sample t-test for equal expectations is performed assuming that both data sets come from normal distributions. Furthermore, a Welch version is applied to counter the effect of possible unequal variances, see Table B.1, page 34. With a p-value of approximately 0.00, the null hypothesis of equal expectations is rejected at every usual level of significance. This underlines the assumption that yield values in eastern and western Germany are different.

Furthermore, Figure 3 applies a yield distinction by each federal state. The five right most box plots represent all eastern German states. Aside from Mecklenburg Western Pomerania and Brandenburg all new federal states provide yields significantly above the national average. For the former, this is counterintuitive and it is difficult to provide a justification. One explanation may be that it is a popular holiday destination. However, for the latter this may likely have to do with agglomeration effects of Berlin to suburban areas in Brandenburg, driving up demand for apartments, hence lowering yields. Contrastingly, and as expected, the majority of old states exhibit significantly lower yields.

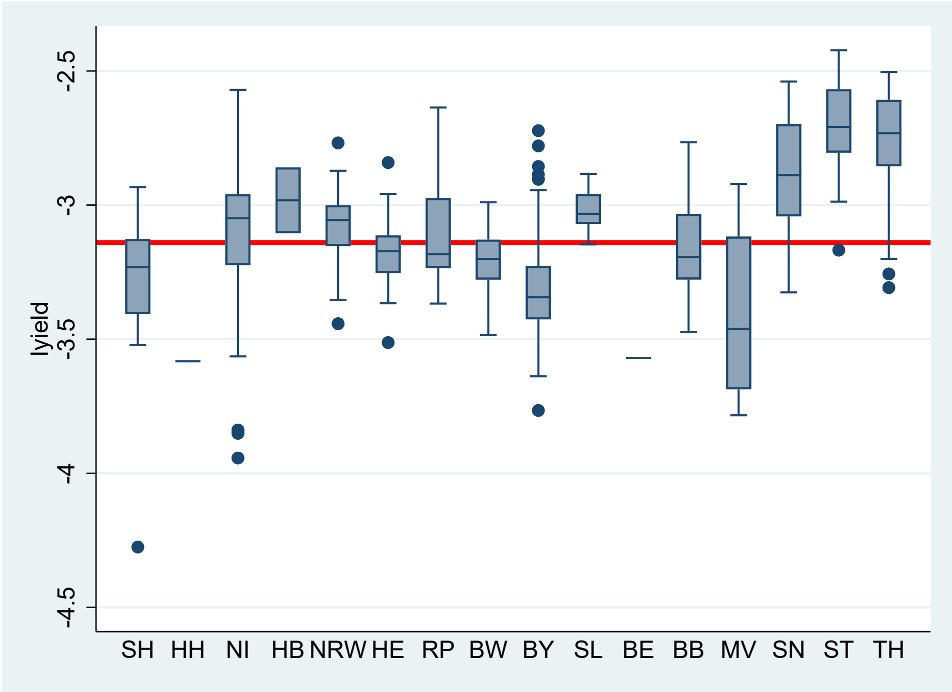


Figure 3: Logarithm of yields distinguished by federal state. Stata code: `mtjt05.do`, page 42. The red line represents the average log yield −3.14 across all federal states.

Next, Figure C.3, p. 39, depicts the distribution of yields comparing old and new states. It clearly shows that yields are somewhat normally distributed in old federal states. The yield value arising most often is $\exp(-3.2) = 4.08\%$. However, in new federal states this cannot be observed. At the same time, Table 3 reveals that for old federal states (and even some new states) there are counties with strikingly low yields. For instance, Nordfriesland has an average yield of about 1.4%. It is not surprising, because the island of Sylt is situated within that county, which is hugely attractive for affluent holidaymakers and people with retirement residences. Hence prices for apartments are higher, and yields lower.

Looking at the 20 counties with the lowest yields depicted in Table 3 it is striking that (with the exception of Berlin) all of them are situated in either Bavaria or northern and eastern German coastal areas (only Hamburg being more distantly located from the sea). Except for Garmisch-Partenkirchen, all counties examined in Bavaria

Table 3: Bottom and top 20 yields. Source: Postbank Studie 2020.

area	yield	rent	price	area	yield	rent	price
Nordfriesland	1.39	7.48	6452.41	Burgenlandkreis	8.87	5.27	713.03
Aurich	1.94	6.85	4239.30	Wittenberg	8.40	5.43	775.38
Leer	2.13	6.56	3703.35	Altenburger Land	8.18	5.26	771.95
Wittmund	2.15	6.14	3423.34	Kyffhäuserkreis	7.94	5.39	814.11
Rostock	2.27	6.55	3457.24	Vogtlandkreis	7.89	4.74	720.88
Miesbach	2.31	11.82	6127.15	Mansfeld-Südharz	7.77	5.38	831.42
Vorpommern-Rügen	2.35	6.57	3354.93	Saale-Orla-Kreis	7.73	5.70	884.43
Starnberg	2.63	13.32	6079.85	Jerichower Land	7.67	5.53	865.29
Garmisch-Partenkirchen	2.67	10.10	4543.01	Goslar	7.65	5.79	908.20
Vorpommern-Greifswald	2.67	6.86	3084.22	Sömmerda	7.57	6.03	956.29
München	2.69	18.10	8078.77	Zwickau	7.45	5.26	847.31
Hamburg	2.78	11.71	5054.23	Unstrut-Hainich-Kreis	7.45	5.45	878.38
Freising	2.79	11.49	4949.39	Hildburghausen	7.37	5.74	934.29
Berlin	2.82	10.89	4638.89	Wartburgkreis	7.37	5.61	913.90
Dachau	2.83	12.48	5296.53	Salzlandkreis	7.30	5.31	872.53
Cuxhaven	2.83	6.51	2758.45	Eichsfeld	7.28	5.91	973.96
Erding	2.88	10.72	4472.57	Görlitz	7.26	4.91	812.00
Ebersberg	2.89	12.40	5151.73	Birkenfeld	7.16	5.34	894.67
Friesland	2.92	6.28	2581.10	Dessau-Rosslau	7.05	5.77	982.47
Fürstenfeldbruck	2.93	12.96	5316.00	Sonneberg	6.85	5.69	996.38

are in close proximity to Munich and the city itself. These southern German counties exhibit enormously high prices relative to rents, whereas for the northern and eastern coastal areas the picture is different. For example, Vorpommern-Greifswald and Cuxhaven are structurally and economically weak (GDP per capita in Cuxhaven is only 23.007€ in 2019). Rents , on average, are low. However, the per square meter purchase prices in these weaker areas are relatively high. All in all, it appears that low yields are either driven by very high prices (as can be seen in Bavaria, Hamburg, Berlin and coastal holiday regions) or comparatively low rents in conjunction with relatively high prices. More broadly speaking, in terms of underlying reasons for low yields there appears to be a huge disparity between the north and south as well as the east and west. This is graphically undermined by Figure 2, p. 16, in which the four groups are defined using the three quartiles. It shows that the highest yields are obtained in eastern Germany whereas the highest concentration of low yields can be found in Bavaria and along the coastal areas. In southwestern and western Germany residential yields tend to be more moderate.

On the other end of the spectrum, 18 of the 20 counties with the highest yields observed are situated in eastern German states belonging to the former German Democratic Republic, specifically in the state of Thuringia (9 out of 18). These results are more intuitive, because purchase prices for apartments are relatively low compared with wealthier regions in western and especially southern Germany. Rents are also lower on average, however they are disproportionally higher relative to the underlying prices. The aforementioned agglomeration effects are clearly visible in all counties surrounding Berlin. In fact, 5 out of 9 surrounding counties exhibit yields lower than 4% and are below the eastern German average. The following table shows how the means in net rents, dwelling prices as well as yields are different between old states and new states. Since values for yields in all German counties were observed, they are included in Table 4 to delineate region specific differences. At the same time, however, net rents as well as dwelling prices are the key constituents of yields and must therefore be analyzed as well to observe how differences in yields come about. These differences are significant as can be seen by inspecting the disjoint confidence intervals.

Table 4: Means of net rents, dwelling prices and yields for old and new states. Stata code: `mtjt06.do`, page 43 (output edited).

	Old states ($n = 325$)		
variable	mean	standard error	95% confidence interval
netrent	7.9717	0.1046	[7.7660,8.1774]
dwellingprice2019	2458.8760	57.3387	[2346.073,2571.679]
yield	0.0419	0.0006	[0.04093,0.04279]
	New states ($n = 76$)		
variable	mean	standard error	95% confidence interval
netrent	6.1320	0.1158	[5.9012,6.3627]
dwellingprice2019	1524.6510	84.9342	[1355.453,1693.848]
yield	0.0553	0.0019	[0.05156,0.05894]

Differences in yields are also significant when comparing urban and rural areas. Figure 2, p. 16, demonstrates that metropolitan areas, on average, exhibit lower yields. In the case of some cities like Berlin and Hamburg, these lower yields can even be observed in peripheral areas. Examples of these are the county of Barnim bordering Berlin to the northeast as well as Pinneberg which is located to the northwest of Hamburg. It can be said that these observations align well with industry reports for the same time period (Catella, 2019; Deloitte, 2019).

To better highlight these differences and closely aligned with Bundesamt für Bauwesen und Raumordnung (2021), in the following a threefold subcategorization of cities is applied to better account for urban and rural differences. For this analysis a selection is made and only cities (Kreisfreie Stadt) are taken into account whereas rural counties (Landkreis) are omitted. Figure 4 illustrates three distinct classes which are divided as follows: Small and medium sized cities (0 to 100.000 inhabitants), big cities (100.000 to 500.000 inhabitants) and major cities (anything above 500.000 inhabitants). The results reveal that yields are similar for the first two groups and average at around $\exp(-3.2) = 4.08\%$. However, they are substantially lower for big urban areas with a population beyond 500.000 inhabitants (of which there are 20 in total), averaging at approximately $\exp(-3.3) = 3.69\%$.

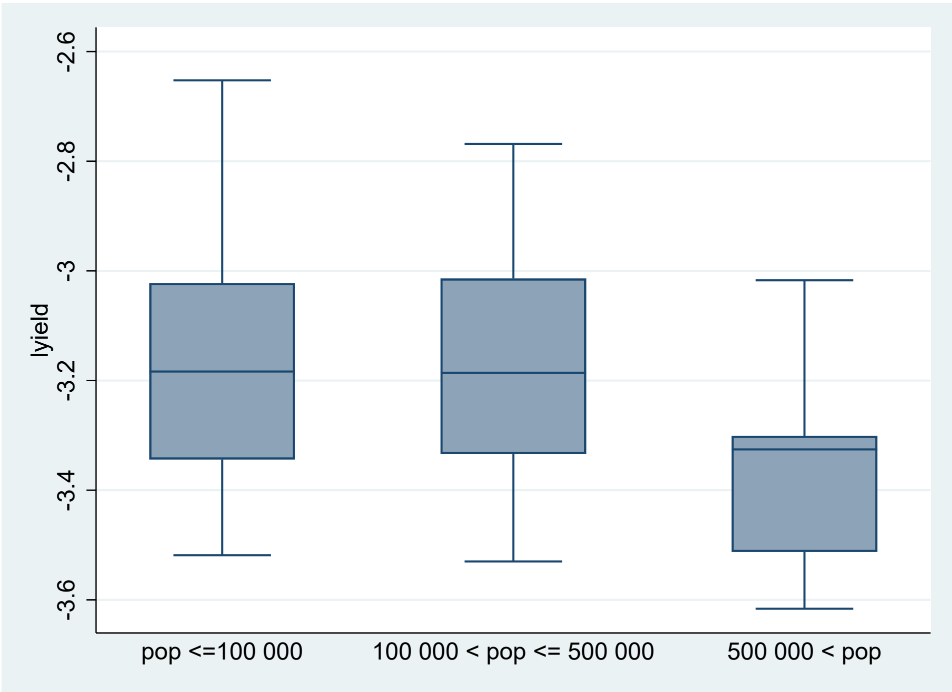


Figure 4: Yields sorted by size of the city. Stata code: `mtjt02.do`, page 41.

Looking at the boxplots and the corresponding descriptive statistics exhibited in Table 5 suggests that there are differences in the yields depending on the population sizes of the cities. To underpin this assumption a one-way ANOVA test for equal expectations in group results is performed in `STATA`, accompanied by the Scheffé option to uncover possible differences in group results, see Table B.2, p. 35 in the appendix.

The test is significant at the 5% level revealing that there are differences in the expectations. The source of this may be low yields in cities with more than 500.000 inhabitants.

Table 5: Descriptive statistics for apartment yields grouped by city size. Stata code: `mtjt02.do`, page 41 (output edited).

City's number of inhibitants (in 1000)	n	mean	std. dev.
up to 100	39	−3.1633	0.2133
more than 100 and up to 500	55	−3.1673	0.1824
more than 500	13	−3.3447	0.1993
Total	107	−3.1874	0.2030

Supplementarily, Figure 5 provides a distinction between urban and rural areas, scrutinizing cities (Kreisfreie Städte) and counties (Landkreise) separately. Remarkably, yields are similar between the two groups. However, rural areas exhibit more outliers in both directions.

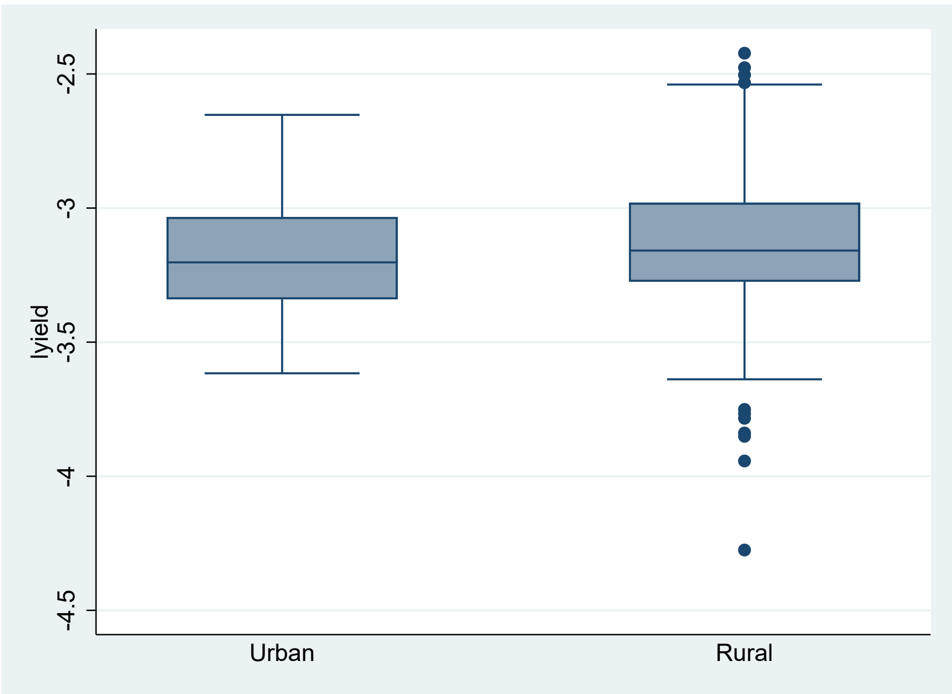
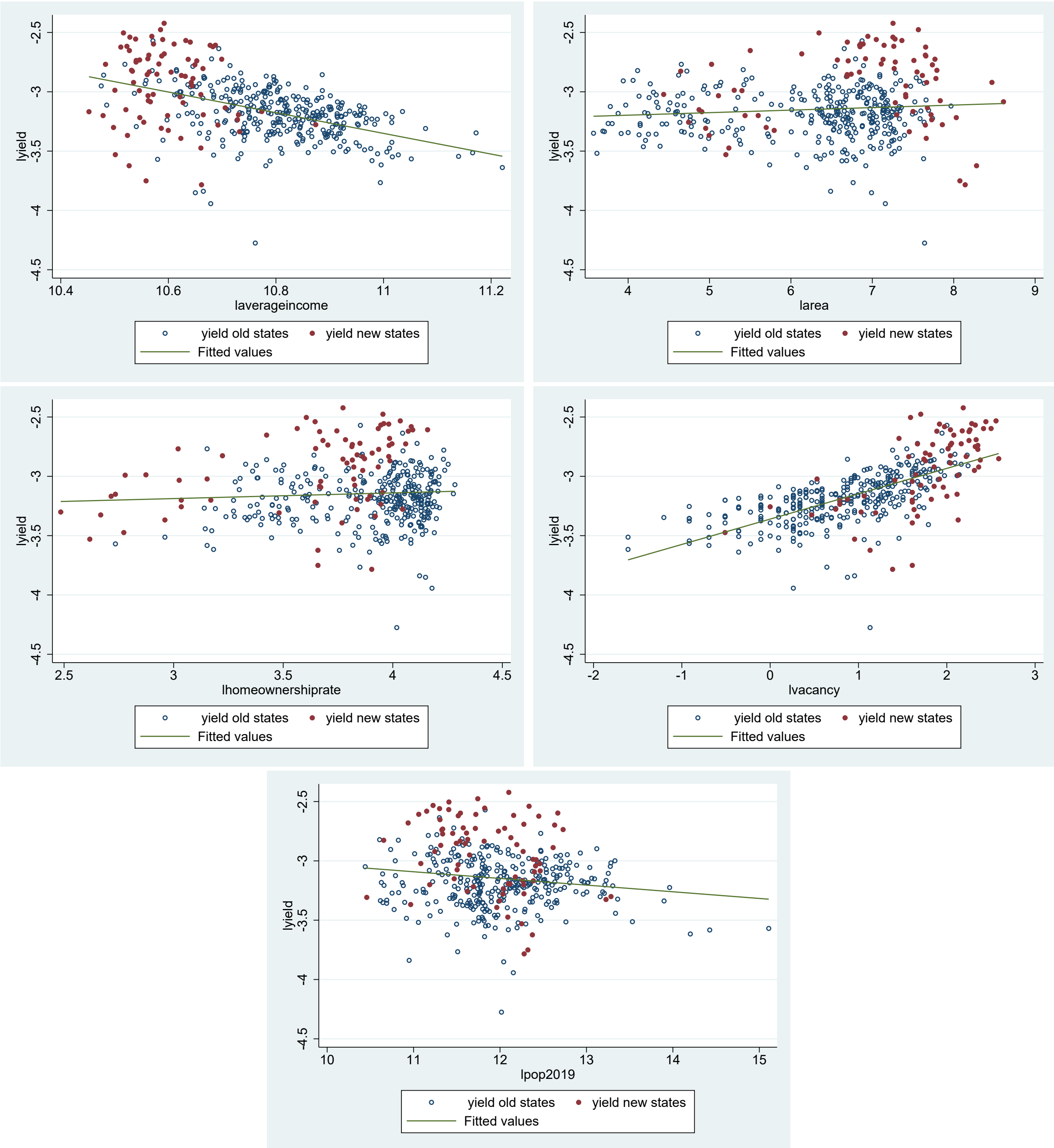


Figure 5: Yields in Urban versus Rural areas. Stata code: `mtjt04.do`, page 42.

The scatter diagrams in Figure 6 illustrate the relationships between log yield and the other variables in the shortlist. They are subdivided into yields of new and old states as well as straight lines drawn according to the method of least squares. After that, the subsequent section is aimed at providing a better understanding at how these differences in yields come about whereby a focus is drawn on the underlying drivers of yields.

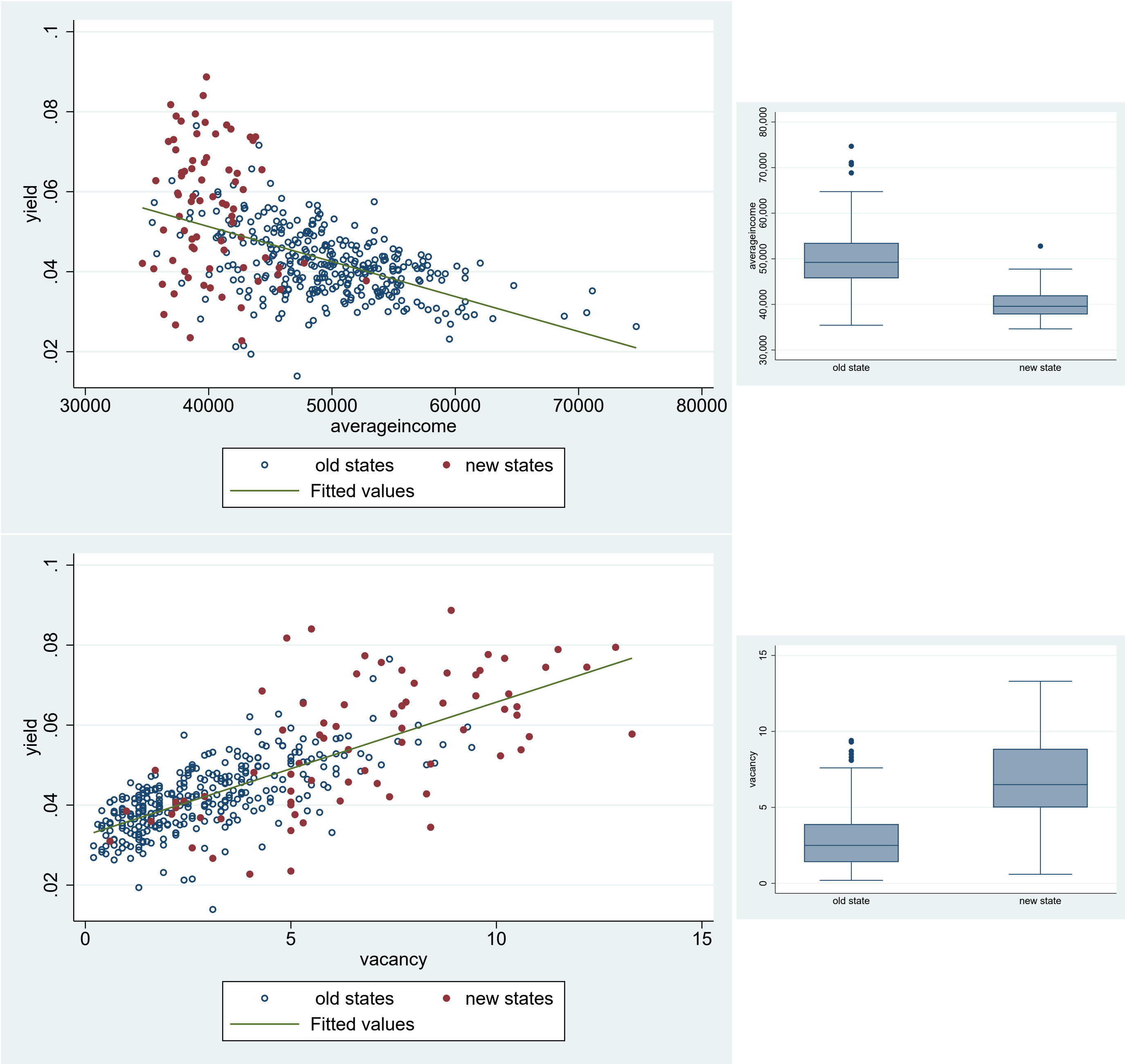
Figure 6: Pairwise scatter plots of log variables and log yields. Stata code: `mtjt12.do`, page 44.



3.2. Differences in drivers of yields

In this subsection the east-west divide is examined. Likewise, further sources are affecting yield values in Germany are addressed. In the analysis a parsimonious regression model including only a few selected variables is used. Scanning the literature (Egner and Grabietz, 2018; Belke and Keil, 2018; Arestis and González, 2014; Chichernea et al., 2008), the variables vacancy and average household income appear to be promising. To linearize the relations the data are logarithmized. The two scatter plots of `laverageincome` and `lyield` as well as `lvacancy` and `lyield` in Figure 7 are subdivided into points of new and old states. They reveal that there is not only an east-west divide in the values of yield and hence of `lyield`, but also in those of `lvacancy` and `laverageincome`. This is further undermined by the boxplots on the right of Figure 7 showing the boxplots for old and new states, respectively. This observation may be the cause of the problem one faces when including a dummy variable distinguishing between new and old states (`newold`), i.e., there may exist multicollinearity caused by an inherent association of the dummy variable with the other exogenous variables `vacancy` and `laverageincome`. An inherent association between `newold` and `laverageincome` seems plausible since households in Eastern federal states have a lower average income than households in the Western federal states. The first boxplot in Figure 7 underlines the difference in average income between East and West.

Figure 7: Pairwise scatter plots and boxplots of exogenous variables subdivided into west and east. Stata code: `mtjt14.do`, page 44.



In order to explicitly express the dependence of yields on `newold` the following model is used:

$$\texttt{lyield}_i = \beta_0 + \beta_1 \texttt{lvacancy}_i + \beta_2 \texttt{laverageincome}_i + \beta_3 \texttt{newold}_i + \varepsilon_i\,, \quad i = 1, \ldots, 401\,. \quad (3)$$

Subsequently, Table 6 exhibits estimation results of five specifications based on Equation (3). The first four involve the dummy variable `newold`. It turns out that `newold` becomes insignificant whenever `lvacancy` is added onto the model (estimations (3) and (4)). As mentioned before the reason for that may be the association of `newold` with `lvacancy` and `laverageincome`. When comparing the different estimations, a regional effect can be observed. Estimations (1) to (3) indicate a statistically significant difference in yields between Eastern and Western German counties. Despite these major regional differences, including both `lvacancy` and `laverageincome` turns `newold` insignificant. This may indicate that the regional effect is priced in quite efficiently. Further statistical analyses of the results can be found in the Appendix A.

Table 6: Estimation results involving `newold`, `lvacancy` and `laverageincome`. Stata code: `mtjt11.do`, page 43.

	(1) lyield	(2) lyield	(3) lyield	(4) lyield	(5) yield
newold	**0.2523***	**0.1073**	**0.06506*	**0.02602**	
	(0.000)	**(0.002)**	**(0.017)**	**(0.387)**	
laveragein~e		**-0.6948***		**-0.2586**	**-0.01657***
		(0.000)		**(0.004)**	**(0.000)**
lvacancy			**0.1987***	**0.1829***	**0.008369***
			(0.000)	**(0.000)**	**(0.000)**
_cons	**-3.1948***	**4.3129***	**-3.3579***	**-0.5502**	**0.2144***
	(0.000)	**(0.000)**	**(0.000)**	**(0.571)**	**(0.000)**
N	**401**	**401**	**401**	**401**	**401**
adj. R-sq	**0.150**	**0.243**	**0.452**	**0.462**	**0.490**

4. Discussion

The central research question of this study is to determine whether yield differences between Eastern and Western Germany persist and to what extent relevant drivers would attribute thereto. The first part of the analysis uncovers yield differences on a county level and shows that yields are, on average, higher in the eastern part of Germany compared to the rest. Although influential drivers of yields are uncovered in the second part, the findings of the analysis do not reveal significant yield differences between eastern and western federal states.

In addition, the results obtained in this study are mostly in line with previous research findings (Egner and Grabietz, 2018; Redding and Sturm, 2008). Examining the underlying drivers of differences uses regression techniques similar to those in Chichernea et al. (2008) and yields results that are similar to those obtained in Belke and Keil (2018), especially with regards to the prevalent influence of disposable household income. The overall yield structure observed is very low, which Belke and Keil (2018) relate to the loosened monetary policy of the European Central Bank. Because of that temporal effect, further studies should perform a similar analysis not on the cross section, but over time. This could yield compelling results with regards to how yields structures have evolved in different parts of the country. The recommended time span are the years 2009–2019, to capture the time of high investment activity after the Great Financial Crisis and before the Covid-19 pandemic. At the same time, a longitudinal analysis would be able to better assess differences in drivers of yields and may provide a more holistic overview as to which drivers are preeminent in different years. The main challenge in providing longitudinal analyses relates to data collection, because relevant data are sparsely available from public resources.

Furthermore, it must be noted that the results need to be interpreted with caution. From an investors point of view, merely observing high yields cannot be used as the only investment decision criterion. One needs to account for the underlying property value and its development over time as well (Himmelberg et al., 2005; Kajuth et al., 2013; Kholodilin et al., 2018). It may be (as seen in parts of Eastern Germany) that high yields are driven by low property values. These low property values may further diminish over time. In that context, high yields concur with capital losses which yields do not account for.

Likewise, it is challenging to collect suitable data for Germany, since data are very fragmented, and few are publicly accessible. Adding further variables is recommended, for example one that controls for supply constraints such as the effect of zoning regulation (Chichernea et al., 2008). As noted in Section 3.2, one or more relevant influential variables may be missing to better explain the yield differences between Eastern and Western Germany using equation (3). The results may change depending on the additional variables chosen and the significance of `newold` may as well. Therefore, this must be noted as a caveat of the thesis, as the research for further relevant influential variables goes beyond the scope of this thesis.

In addition, further studies could unveil differences between states in Eastern and

Western Germany and broadly comparing yields between federal states. For instance, it is striking that the highest yields were observed specifically in the Eastern German state of Thuringia, for which the author does not have an intuitive answer, and which is beyond the scope of this research. The seminal annual publication *UBS Global Real Estate Bubble Index* hints at potential caveats with regards to yields compressions unfolding in some German metropolitan cities. Comparing major German urban cities to global counterparts appears to be insightful as well. Finally, building up on this analysis could unveil promising findings for investors wishing to better understand and drawing valuable conclusions about the prospects of yields before making long term investments into apartments.

5. Conclusion

The results of this study are twofold. In a first step, differences in yields across space are unveiled. These are mostly as assumed, i.e. yields in eastern German states (except for Berlin) being considerably higher than their western German counterparts. In addition, when comparing rural and urban areas, yields are notably lower in the latter, specifically in metropolitan areas. Likewise, yields are found to be particularly low in the economically most prosperous regions, and vice versa. This is especially true for southern Bavaria and the German coastline, where yields are remarkably low. The author finds that low yields are either driven by high prices (as can be observed in Bavaria) or come about by comparatively low rents in conjunction with relatively high prices. On the other hand, high yields are typically explained by lower purchase prices and lower demand. This means that these are, on average, lower in Eastern Germany compared to the rest.

With regards to estimating the drivers of yields, the analysis reveals that average disposable household income and vacancy rates may be insightful constituents. When it comes to assessing yield structures across space, average disposable household income and vacancy rates are found to significantly influence yields throughout the model estimations.

This is in line with the findings of the first part of the analysis, i.e. yields being highly associated with economic prosperity. Hence, lower disposable household income and an overarching economic disadvantage of Eastern Germany leads to higher yields and underlines the assumption that an East-West divide in yields remains ubiquitous for the year 2019 when assessing yield differences across space. However, the analysis does not show a significant difference between Eastern and Western states when using a dummy variable and including both vacancy rates and average income. While multicollinearity may not be the problem here, missing variables could be a potential explanation. Although representing a caveat of this thesis, this insight could be used to build upon further research on the topic of the East-West divide in yields.

6. References

B. Algieri. House price determinants: Fundamentals and underlying factors. *Comparative Economic Studies*, 55(2):315–341, 2013.

P. Arestis and A. R. González. Modelling the housing market in OECD countries. *International Review of Applied Economics*, 28(2):131–153, 2014.

A. Baum. *Real Estate Investment: A Strategic Approach*. EBL-Schweitzer. CRC Press, 2015. ISBN 9781317647874.

A. Belke and J. Keil. Fundamental determinants of real estate prices: A panel study of german regions. *International Advances in Economic Research*, 24(1):25–45, 2018.

Bundesamt für Bauwesen und Raumordnung. Laufende Stadtbeobachtung – Raumabgrenzungen. `https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/gemeinden/StadtGemeindetyp/StadtGemeindetyp.html`, (accessed January 3, 2022), 2021.

Catella. Investment locations germany 2019: Residential – rents and yields. Technical report, Catella Research, 2019. URL `https://www.catella.com/globalassets/global/mix-germany-corporate-finance/catella-residential-market-map-2019_final_en.pdf`.

D. Chichernea, N. Miller, J. Fisher, M. Sklarz, and B. White. A cross-sectional analysis of cap rates by msa. *Journal of Real Estate Research*, 30(3):249–292, 2008.

Deloitte. Property index overview of European residential markets. Technical report, Deloitte, 2019. URL `https://www2.deloitte.com/content/dam/Deloitte/de/Documents/real-estate/property-index-2019-2.pdf`.

C. Dustmann, B. Fitzenberger, and M. Zimmermann. Housing expenditures and income inequality. *ZEW-Centre for European Economic Research Discussion Paper*, (18-048), 2018.

B. Egner and K. J. Grabietz. In search of determinants for quoted housing rents: Empirical evidence from major german cities. *Urban Research & Practice*, 11(4):460–477, 2018.

European Central Bank. Financial stability review. `https://www.ecb.europa.eu/pub/pdf/fsr/ecb.fsr202105~757f727fe4.en.pdf`, (accessed January 4, 2022), 2021.

C. Goodhart and B. Hofmann. House prices, money, credit, and the macroeconomy. *Oxford review of economic policy*, 24(1):180–205, 2008.

B. Hargreaves. Exploring the yields on residential investment property. *Pacific Rim Real Estate Society (PRRES)*, 2005.

C. Heij, C. Heij, P. de Boer, P. H. Franses, T. Kloek, H. K. van Dijk, et al. *Econometric methods with applications in business and economics*. Oxford University Press, 2004.

R. J. Hill and I. A. Syed. Hedonic price–rent ratios, user cost, and departures from equilibrium in the housing market. *Regional Science and Urban Economics*, 56:60–72, 2016.

C. Himmelberg, C. Mayer, and T. Sinai. Assessing high house prices: Bubbles, fundamentals and misperceptions. *Journal of Economic Perspectives*, 19(4):67–92, 2005.

O. Holtemöller and R. Schulz. Investor rationality and house price bubbles: Berlin and the german reunification. *German Economic Review*, 11(4):465–486, 2010.

M. D. Igan and M. P. Loungani. *Global housing cycles*. International Monetary Fund, 2012.

F. Kajuth, T. Knetsch, and N. Pinkwart. Assessing house prices in Germany: Evidence from an estimated stock-flow model using regional data. *Bundesbank Discussion Paper*, 2013.

K. A. Kholodilin, C. Michelsen, and D. Ulbricht. Speculative price bubbles in urban housing markets. *Empirical Economics*, 55(4):1957–1983, 2018.

J. Möbert. Outlook for the german residential property market 2021 and beyond. https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000517463/Outlook_for_the_German_residential_property_market.PDF?undefined&realload=q66DPqEE9WNN6MCmgdtYEq74Yqg4Qbf5IRo2snRfx2U7BsVOfmu06l6ntTmwS0f3, (accessed January 5, 2022), 2021.

OECD. Regions and Cities at a Glance 2018 – germany. https://www.oecd.org/cfe/GERMANY-Regions-and-Cities-2018.pdf, (accessed January 3, 2022), 2019.

S. Redding and D. M. Sturm. Cep discussion paper no 688 may 2005. 2005.

S. J. Redding and D. M. Sturm. The costs of remoteness: Evidence from german division and reunification. *American Economic Review*, 98(5):1766–97, 2008.

M. Voigtländer. Why is the german homeownership rate so low? *Housing Studies*, 24 (3):355–372, 2009.

W. C. Wheaton. Vacancy, search, and prices in a housing market matching model. *Journal of political Economy*, 98(6):1270–1292, 1990.

W. C. Wheaton and G. Nechayev. Does location matter? *The Journal of Portfolio Management*, 31(5):100–108, 2005.

World Bank Group. Cities in Europe and Central Asia: Germany. https://documents1.worldbank.org/curated/en/924911511933682347/pdf/121726-BRI-P154478-PUBLIC-Germany-Snapshot-Print.pdf, (accessed January 3, 2022), 2017.

Y. Zhang, X. Hua, and L. Zhao. Exploring determinants of housing prices: A case study of chinese experience in 1999–2010. *Economic modelling*, 29(6):2349–2361, 2012.

Appendix

A. Further statistical analyses

This section provides further technical analyses to validate the results obtained in Section 3.2.

Table A.7 shows the results of the OLS estimation of equation (3), p. 25. The F-statistic is significant implying that the model has a high explanatory value. Likewise, the adjusted R^2 is 0.4624 pointing to a modest goodness of fit. The influences of `lvacancy` and of `laverageincome` on `lyield` are both significant at the 1% level and all estimated coefficients have the expected signs. A 1% increase in averageincome leads to an estimated decrease of the yield of approximately 0.26% and a 1% increase in vacancy leads to an estimated increase of the yield of approximately 0.18%.

Table A.7: Estimation output for model (3). Stata code: `mtjt11.do`, page 43.

Source	SS	df	MS			
				Number of obs	=	401
				F(3, 397)	=	115.68
Model	12.0525245	3	4.01750818	Prob > F	=	0.0000
Residual	13.7874327	397	.03472905	R-squared	=	0.4664
				Adj R-squared	=	0.4624
Total	25.8399572	400	.064599893	Root MSE	=	.18636

lyield	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lvacancy	.182903	.014313	12.78	0.000	.1547642	.2110418
laverageincome	-.2586256	.0893141	-2.90	0.004	-.4342134	-.0830378
newold	.0260189	.0300137	0.87	0.387	-.0329868	.0850246
_cons	-.5501993	.9697323	-0.57	0.571	-2.456652	1.356253

The estimation of (3) yields the decomposition

$$\begin{aligned} \texttt{lyield}_i &= -0.5502 - 0.2586\,\texttt{laverageincome}_i + 0.1829\,\texttt{lvacancy}_i \\ &\quad + 0.0260\,\texttt{newold}_i + e_i \\ &= \widehat{\texttt{lyield}}_i + e_i\,, \quad i = 1, \ldots, 401\,. \end{aligned} \tag{4}$$

This serves as the input of the residual versus fit plot displayed in Figure C.4. It is a scatter plot of the linear prediction regressed with the residuals. It turns out that there is a clear-cut divide visible separating East and West.

With regards to multicollinearity, the variance inflation factors displayed in Table B.3, p. 35, reveal that multicollinearity does not cause problems in the estimation of equation (3). With respect to the assumption of homoscedastic error variances, the Breusch-Pagan Test hints at the fact that the assumption of homoscedastic error variances is in doubt (Table B.4, p. 36). A hint of a possible cause may be deducted from Figure C.4 exhibiting different variabilities in the residuals of regions in the East and the West. To safeguard the results, in the estimation of equation (3), heteroscedastic consistent standard errors have been computed (Table B.5, p. 36). It turns out that the observations reached before do not change.

Besides, conducting the Ramsey RESET test (Table B.6, p. 36) yields a significant result which leads to the conclusion that one or more relevant influential variables may be missing. However, assessing those is beyond the scope of this thesis. Lastly, assessing the normality assumption: Figure C.5, p. 40, shows the histogram of the corresponding residuals resulting from the estimation. Compared to the added density of a normal distribution there is a notable deviation from symmetry visible. Thus, the assumption of normally distributed errors is in doubt. To confirm this impression, a Shapiro-Wilk test of normal distribution is conducted confirming this observation (Table B.7, p. 36).

B. Tables

Table B.1: Two-sample t-test for equal means of yields in old and new states. Welch's version is used to account for possible unequal variances. Stata code: `mtjt05.do`, page 42.

```
Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
old stat	325	.041858	.000472	.0085094	.0409294	.0427866
new stat	76	.0552512	.0018506	.016133	.0515646	.0589377
combined	401	.0443963	.0005804	.0116222	.0432553	.0455373
diff		-.0133932	.0019098		-.0171903	-.0095961

```
    diff = mean(old stat) - mean(new stat)                    t =  -7.0128
Ho: diff = 0                        Welch's degrees of freedom =  85.2577

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Table B.2: One-way ANOVA test for equal yields in different cities. A Scheffé test for pairwise multiple comparisons is added. Stata code: `mtjt02.do`, page 41.

```
Summary for variables: yield
     by categories of: pop2019classes
```

pop2019classes	N	mean	sd
pop <=100 000	39	.0432661	.0097376
100 000 < pop <=	55	.0428175	.0079446
500 000 < pop	13	.0359316	.0073302
Total	107	.0421444	.0088126

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	.000575774	2	.000287887	3.91	0.0230
Within groups	.007656304	104	.000073618		
Total	.008232077	106	.000077661		

```
Bartlett's test for equal variances:  chi2(2) =   2.4238  Prob>chi2 = 0.298
```

Comparison of yield by pop2019cla~s
(Scheffe)

Row Mean- Col Mean	pop <=10	100 000
100 000	-.000449 0.969	
500 000	-.007334 0.032	-.006886 0.038

Table B.3: Variance inflation factors resulting from estimating (3), p. 25. Stata code: `mtjt11.do`, page 43.

Variable	VIF	1/VIF
laveragein~e	1.79	0.559655
newold	1.60	0.625894
lvacancy	1.49	0.669576
Mean VIF	1.63	

Table B.4: Breusch-Pagan test for constant error variances resulting from estimating (3), p. 25. Stata code: `mtjt11.do`, page 43.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lyield

         chi2(1)      =    20.18
         Prob > chi2  =   0.0000
```

Table B.5: Re-estimating (3), p. 25, using robust standard errors. Stata code: `mtjt11.do`, page 43.

```
Linear regression                               Number of obs     =        401
                                                F(3, 397)         =     150.14
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4664
                                                Root MSE          =     .18636
```

lyield	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
lvacancy	.182903	.0120595	15.17	0.000	.1591945	.2066115
laverageincome	-.2586256	.0816131	-3.17	0.002	-.4190735	-.0981776
newold	.0260189	.0355885	0.73	0.465	-.0439466	.0959844
_cons	-.5501993	.8893184	-0.62	0.536	-2.298561	1.198163

Table B.6: Ramsey's RESET test for omitted variables after estimating (3), p. 25, by OLS. Stata code: `mtjt11.do`, page 43.

```
Ramsey RESET test using powers of the fitted values of lyield
       Ho:  model has no omitted variables
                 F(3, 394) =     12.21
                  Prob > F =      0.0000
```

Table B.7: Shapiro-Wilk test for normality of disturbances. Stata code: `mtjt11.do`, page 43.

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
resid	401	0.90774	25.455	7.703	0.00000

C. Figures

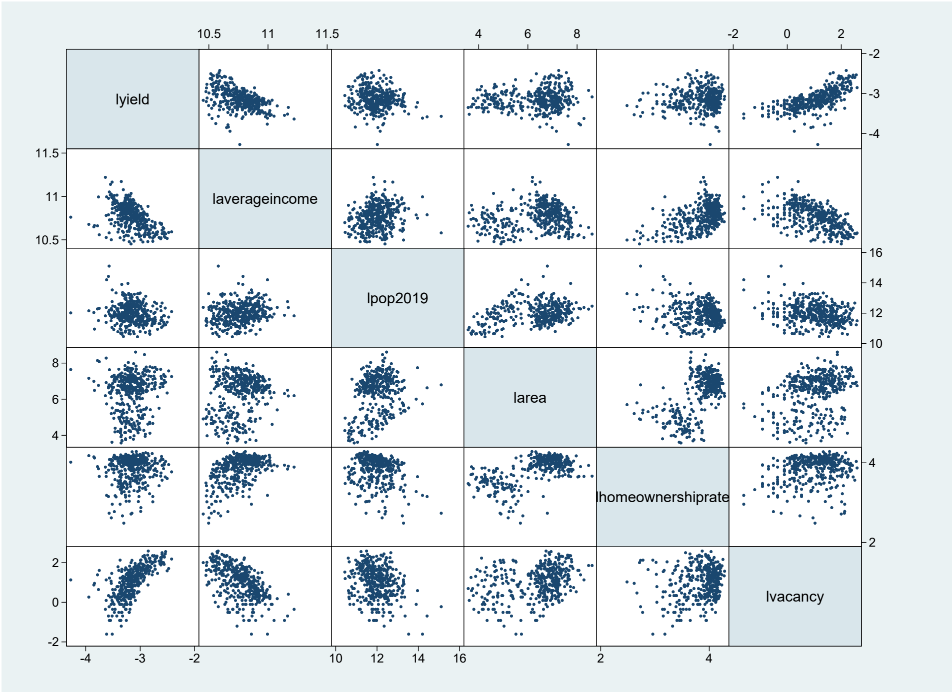


Figure C.1: Pairwise scatter plots. Stata code: `mtjt09.do`, page 43.

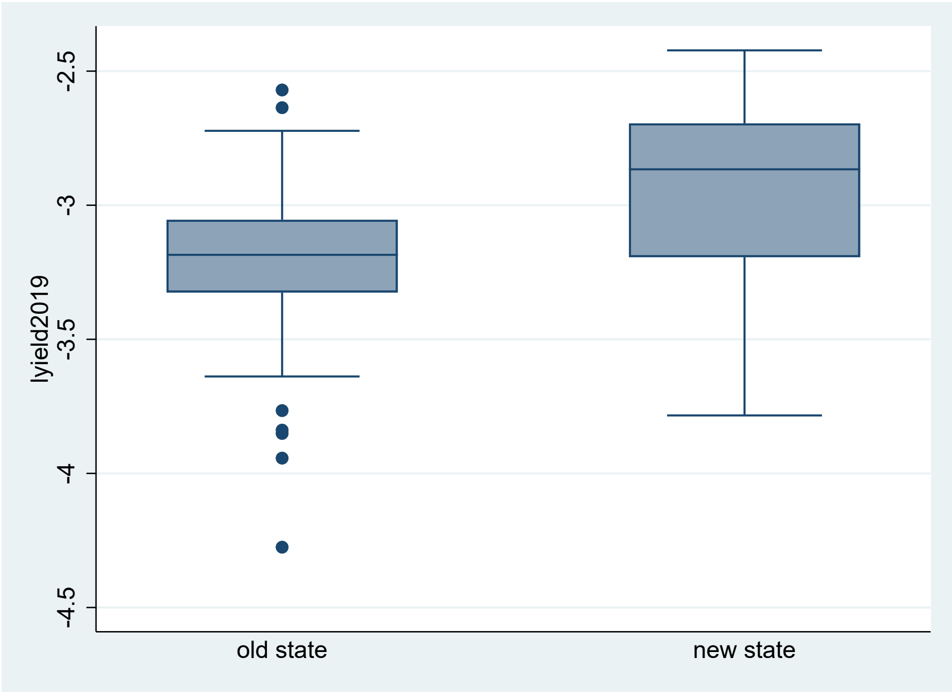


Figure C.2: Differences in logarithm of yields in old versus new federal states. Stata code: `mtjt05.do`, page 42.

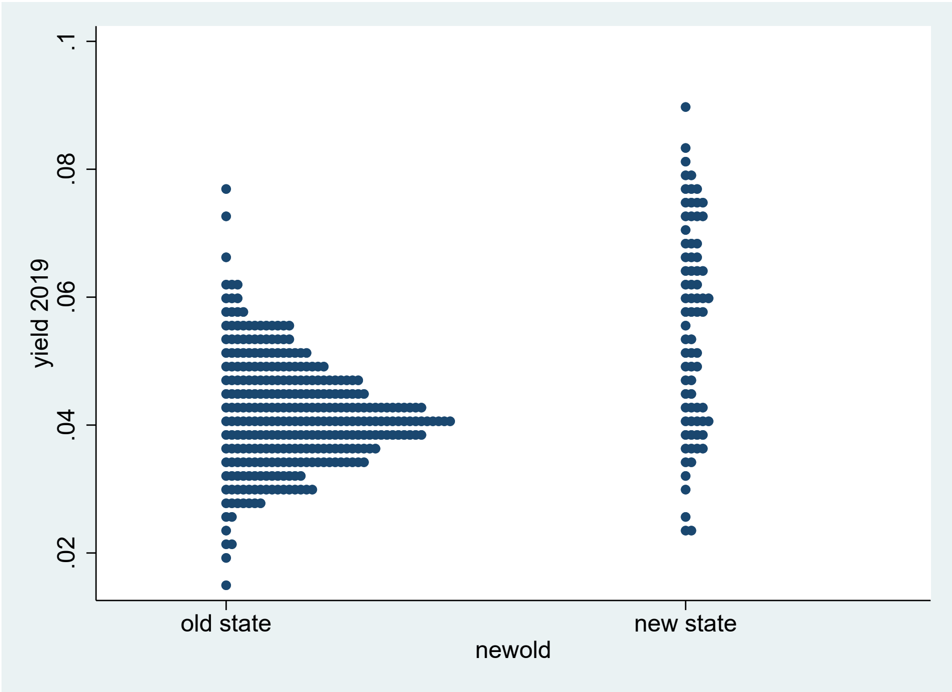
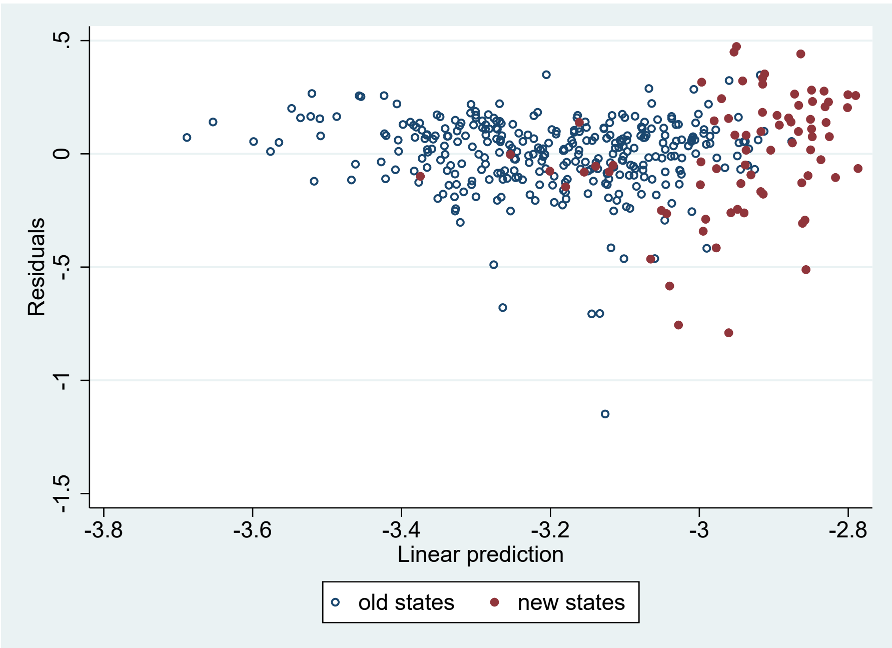


Figure C.3: Distribution of yields in old versus new states. Stata code: `mtjt05.do`, page 42.

Figure C.4: Residual versus fit plot distinguishing between Eastern and Western federal states. Stata code: `mtjt11.do`, page 43.



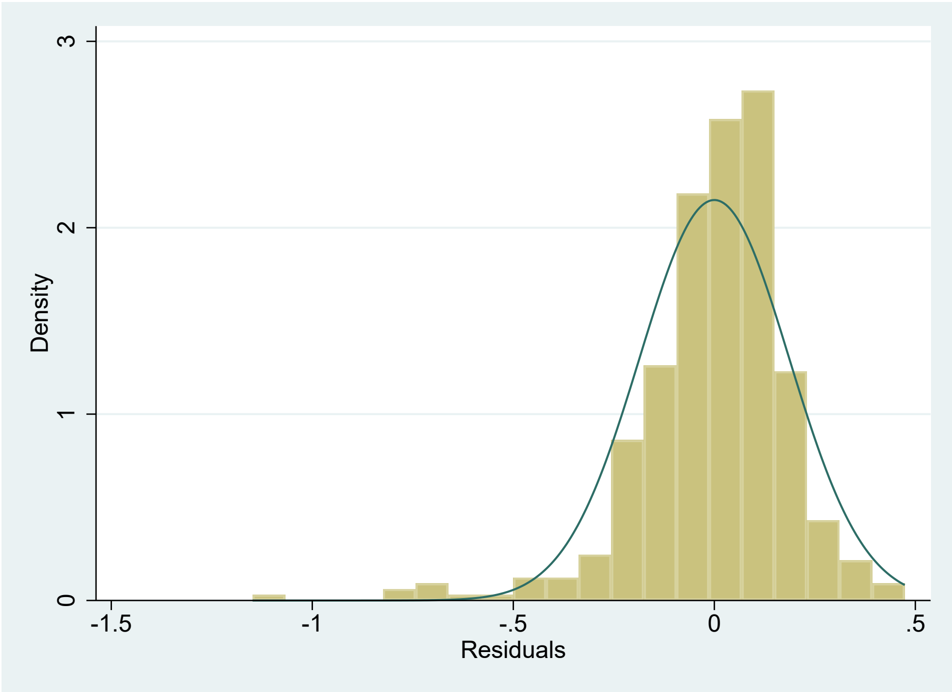


Figure C.5: Histogram of OLS residuals resulting from estimating (3) with the density of a fitted normal distribution. Stata code: `mtjt11.do`, page 43.

D. Stata Codes

mtjt01.do

```
mtjt01.do
/*
run C:\mtjt\Stata\mtjt.do\mtjt01.do, nostop
*/

program drop multtsline

/* Trick found in */
/* http://pierrefrancois.wifeo.com/documents/Intro-Stata---LSE-III.pdf*/

* Plotting multiple time series
* https://www.statalist.org/forums/forum/general-stata-discussion/
*           general/1355560-plotting-multiple-time-series
*! 1.0.0 NJC 5sept2016
program multtsline
    version 11
    syntax varlist(numeric) [if] [in]                                ///
    [, byopts(str asis) mylabels(str asis) *]

    quietly {
        marksample touse
        count if ‘touse’
        if r(N) == 0 exit 2000

        tsset
        local panelvar ‘r(panelvar)’
        local timevar ‘r(timevar)’

        local varlist : list varlist - timevar
```

```
        if "‘varlist’" == "" exit 102

        preserve
        keep if ‘touse’
        drop ‘touse’

        gettoken yvar varlist : varlist
        local J = 0
        while "‘yvar’" != "" {
            local ++J
            local lbl‘J’ : var label ‘yvar’
            if ‘"‘lbl‘J’’"’ == "" local lbl‘J’ "‘yvar’"
            local call ‘call’ ‘yvar’ ‘timevar’ ‘panelvar’
            gettoken yvar varlist : varlist
        }

        tempname y
        stack ‘call’, into(‘y’ ‘timevar’ ‘panelvar’) clear

        if ‘"‘mylabels’"’ != "" {
            tokenize ‘mylabels’
            forval j = 1/‘J’ {
                label def _stack ‘j’ ‘"‘‘j’’"’, add
            }
        }
        else forval j = 1/‘J’ {
            label def _stack ‘j’ ‘"‘lbl‘j’’"’, add
        }
        label val _stack _stack
    }

    sort ‘panelvar’ ‘timevar’

    line ‘y’ ‘timevar’, by(_stack, col(1) yrescale note("") ‘byopts’)    ///
    ytitle("") xtitle("")   yla(, ang(h)) c(L)                           ///
    subtitle(, pos(9) bcolor(none) nobexpand place(e)) ‘options’
end
use C:\mtjt\Stata\mtjt.do\dta\timeseries.dta, clear
generate year = 1987 + _n - 1
generate pricetorent = resprop/rents
tsset year
label var inflation "inflation rate (in %)"
label var dax "DAX score (by the end of year)"
label var resprop "residential property price index (base year 2015)"
label var rents "rent index (base year 2015)"
label var gdp "per capita gross domestic product (euro)"
label var baserate "mean annual base interest rate (ECB)"
label var pricetorent "price-to-rent"
multtsline baserate dax gdp resprop rents pricetorent if 2000 <= year & year <= 2019, xtitle("year")
graph export C:\mtjt\Stata\mtjt.do\graphics\161121c.pdf, as(pdf) name("Graph") replace
```

mtjt02.do

mtjt02.do

```
/*
do C:\mtjt\Stata\mtjt.do\mtjt02.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

generate pop2019classes=100  if pop2019 <= 100000
replace pop2019classes=500 if 100000 < pop2019 & pop2019 <= 500000
replace pop2019classes=10000 if 500000 < pop2019
label define popclass 100 "pop <=100 000" 500 "100 000 < pop <= 500 000" 10000 "500 000 < pop"
label value pop2019classes popclass
encode Regionsebene, gen(RegionsebeneAsFactor)

graph box lyield if(RegionsebeneAsFactor==1), over(pop2019classes)
graph export C:\mtjt\Stata\mtjt.do\graphics\201121a.pdf, as(pdf) name("Graph") replace

tabstat yield if(RegionsebeneAsFactor==1), by(pop2019classes) statistics(n mean sd)
```

```
oneway yield pop2019classes if(RegionsebeneAsFactor==1), scheffe
```

mtjt03.do

mtjt03.do

```
/*
do C:\mtjt\Stata\mtjt.do\spmap\mtjt03
*/

/*
For the following to work appropriately please install shp2dta and spmap
* ssc install shp2dta
* ssc install spmap
*/

/*
shp2dta using gadm36_DEU_2, database("germany-attr.dta") ///
coordinates("germany-coord.dta") genid(stid) gencentroids(cc) replace
*/

cd  C:\mtjt\Stata\mtjt.do\spmap
use C:\mtjt\Stata\mtjt.do\spmap\germany-attr.dta,clear
sort CC_2
rename CC_2 regionalcode
destring regionalcode,replace
merge 1:1 regionalcode using C:\mtjt\Stata\mtjt.do\spmap\yield.dta

/* Grouping yields according to quartiles, no legend */
spmap  yield  using "germany-coord.dta", legend(off)              ///
id(stid) fcolor(Rainbow) ocolor(white ..) osize(thin ..)          ///
clmethod(custom)    clbreaks(0  .036819  .0420813 .0498317  0.10)
graph export "C:\mtjt\Stata\mtjt.do\graphics\201221a.png", as(png) name("Graph") replace
```

mtjt04.do

mtjt04.do

```
/*
run C:\mtjt\Stata\mtjt.do\mtjt04.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

encode Regionsebene, gen(RegionsebeneAsFactor)
label define ruralurban 1 "Urban" 2 "Rural"
label value RegionsebeneAsFactor ruralurban
graph box lyield, over(RegionsebeneAsFactor)
graph export "C:\mtjt\Stata\mtjt.do\graphics\201121b.pdf", as(pdf) name("Graph") replace
```

mtjt05.do

mtjt05.do

```
/*
run C:\mtjt\Stata\mtjt.do\mtjt05.do
*/

graph box lyield , over(fedstate, sort(fedstatecode)) ///
yline(-3.14, lstyle(foreground) lwidth(0.7) lcolor(red))
graph export "C:\mtjt\Stata\mtjt.do\graphics\101121c.pdf", as(pdf) name("Graph") replace

graph box lyield , over(newold,sort(fedstatecode))  ytitle(lyield2019)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021o.pdf", as(pdf) name("Graph") replace

ttest yield, by(newold) welch
```

mtjt06.do

mtjt06.do

```
/*
run C:\mtjt\Stata\mtjt.do\mtjt06.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

mean netrent dwellingprice2019 yield if (newold==0)
mean netrent dwellingprice2019 yield if (newold==1)
```

mtjt09.do

mtjt09.do

```
/*
run C:\mtjt\Stata\mtjt.do\mtjt09.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

graph matrix lyield laverageincome lpop2019 larea lhomeownershiprate lvacancy,msize(vsmall)
graph export "C:\mtjt\Stata\mtjt.do\graphics\041221b.pdf", as(pdf) name("Graph") replace
```

mtjt11.do

mtjt11.do

```
/*
do C:\mtjt\Stata\mtjt.do\mtjt11.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

eststo clear
reg lyield newold
eststo mod1
reg lyield newold laverageincome
eststo mod2
reg lyield newold lvacancy
eststo mod3
reg lyield newold lvacancy laverageincome
eststo mod4
reg yield lvacancy laverageincome
eststo mod5
esttab mod1 mod2 mod3 mod4 mod5, ar2 p b(a4)

regress lyield lvacancy laverageincome newold
predict yfit, xb
predict resid, resid

graph twoway (scatter resid yfit if newold==0, msymbol(Oh) msize(small) mlabsize(large))  ///
(scatter  resid yfit if newold==1, msymbol(O) msize(small)),                              ///
legend(label(1 old states) label(2 new states))
graph export C:\mtjt\Stata\mtjt.do\graphics\020322m.pdf, as(pdf) name("Graph") replace

vif
hettest
regress lyield lvacancy laverageincome newold, robust
regress lyield lvacancy laverageincome newold
ovtest
hist resid,normal
graph export "C:\mtjt\Stata\mtjt.do\graphics\020322s.pdf", as(pdf) name("Graph") replace
swilk resid
drop yfit resid
```

mtjt12.do

mtjt12.do

```
/*
run C:\mtjt\Stata\mtjt.do\mtjt12.do
*/

run C:\mtjt\Stata\mtjt.do\ProvideData.do

graph twoway (scatter lyield laverageincome if newold==0, msymbol(Oh) msize(small) mlabsize(large)) ///
(scatter lyield laverageincome if newold==1, msymbol(O) msize(small))                               ///
(lfit lyield laverageincome), legend(label(1 yield old states) label(2 yield new states))          ///
ytitle(lyield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021b.pdf", as(pdf) name("Graph") replace

graph twoway (scatter lyield larea if newold==0, msymbol(Oh) msize(small) mlabsize(large))          ///
(scatter lyield larea if newold==1, msymbol(O) msize(small))                                        ///
(lfit lyield larea), legend(label(1 yield old states) label(2 yield new states))                   ///
ytitle(lyield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021e.pdf", as(pdf) name("Graph") replace

graph twoway (scatter lyield lhomeownershiprate if newold==0, msymbol(Oh) msize(small)              ///
mlabsize(large)) (scatter lyield lhomeownershiprate if newold==1, msymbol(O) msize(small))          ///
(lfit lyield lhomeownershiprate), legend(label(1 yield old states) label(2 yield new states))      ///
ytitle(lyield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021k.pdf", as(pdf) name("Graph") replace

graph twoway (scatter lyield lvacancy if newold==0, msymbol(Oh) msize(small) mlabsize(large))       ///
(scatter lyield lvacancy if newold==1, msymbol(O) msize(small))                                     ///
(lfit lyield lvacancy), legend(label(1 yield old states) label(2 yield new states))                ///
ytitle(lyield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021l.pdf", as(pdf) name("Graph") replace

graph twoway (scatter lyield lpop2019 if newold==0, msymbol(Oh) msize(small) mlabsize(large))       ///
(scatter lyield lpop2019 if newold==1, msymbol(O) msize(small))                                     ///
(lfit lyield lpop2019), legend(label(1 yield old states) label(2 yield new states))                ///
ytitle(lyield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\161021s.pdf", as(pdf) name("Graph") replace
```

mtjt14.do

mtjt14.do

```
/*
do C:\mtjt\Stata\mtjt.do\mtjt14.do
*/

/* run C:\mtjt\Stata\mtjt.do\ProvideData.do */

graph twoway (scatter yield averageincome if newold==0, msymbol(Oh) msize(small) mlabsize(large)) ///
(scatter yield averageincome if newold==1, msymbol(O) msize(small))                               ///
(lfit yield averageincome), legend(label(1 old states) label(2 new states))                      ///
ytitle(yield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\250222b.pdf", as(pdf) name("Graph") replace
graph box averageincome,over(newold,sort(fedstatecode)) ytitle(averageincome)
graph export "C:\mtjt\Stata\mtjt.do\graphics\250222f.pdf", as(pdf) name("Graph") replace

graph twoway (scatter yield vacancy if newold==0, msymbol(Oh) msize(small) mlabsize(large))       ///
(scatter yield vacancy if newold==1, msymbol(O) msize(small))                                     ///
(lfit yield vacancy), legend(label(1 old states) label(2 new states))                            ///
ytitle(yield)
graph export "C:\mtjt\Stata\mtjt.do\graphics\250222c.pdf", as(pdf) name("Graph") replace
graph box vacancy, over(newold,sort(fedstatecode)) ytitle(vacancy)
graph export "C:\mtjt\Stata\mtjt.do\graphics\250222e.pdf", as(pdf) name("Graph") replace
```