Faculty of Business Studies and Economics
University of Bremen

Doctoral Thesis

THREE ESSAYS IN HOUSING ECONOMICS

A thesis submitted to the Doctoral Commission Dr. rer. pol. of the University of Bremen in fulfilment of the requirements for the degree of Dr. rer. pol.

ANGELINA HACKMANN

Date of submission: November 5, 2022
Date of colloquium: March 1, 2023

First examiner
Prof. Dr. Torben Klarl

Second examiner
Prof. Dr. Christian Cordes

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABS	Asset-Backed Securities
APP	Asset Purchase Programme
CET	Central European Time
DSTI	Debt-Service-To-Income
DTI	Debt-To-Income
EA	Euro Area
EA-MPD	Euro Area Monetary Policy Database
EC	European Commission
ECB	European Central Bank
ED	European DataWarehouse
ESRB	European Systemic Risk Board
FRG	Federal Republic of Germany
GDP	Gross Domestic Product
GDR	German Democratic Republic
GVA	Gross Value Added
LTV	Loan-To-Value
MPC	Marginal Propensity to Consume
MSA	Metropolitan Statistical Area
OECD	Organization for Economic Co-operation and Development
OIS	Overnight Index Swap
OLS	Ordinary Least Squares
RMBS	Residential Mortgage-Backed Securities
SMP	Securities Market Programme
SOEP	Socio-Economic Panel
SPVAR	Structural Panel Vector Autoregression
U.S.	United States of America
UK	United Kingdom
VAR	Vector Autoregression

CHAPTER

ONE

INTRODUCTION

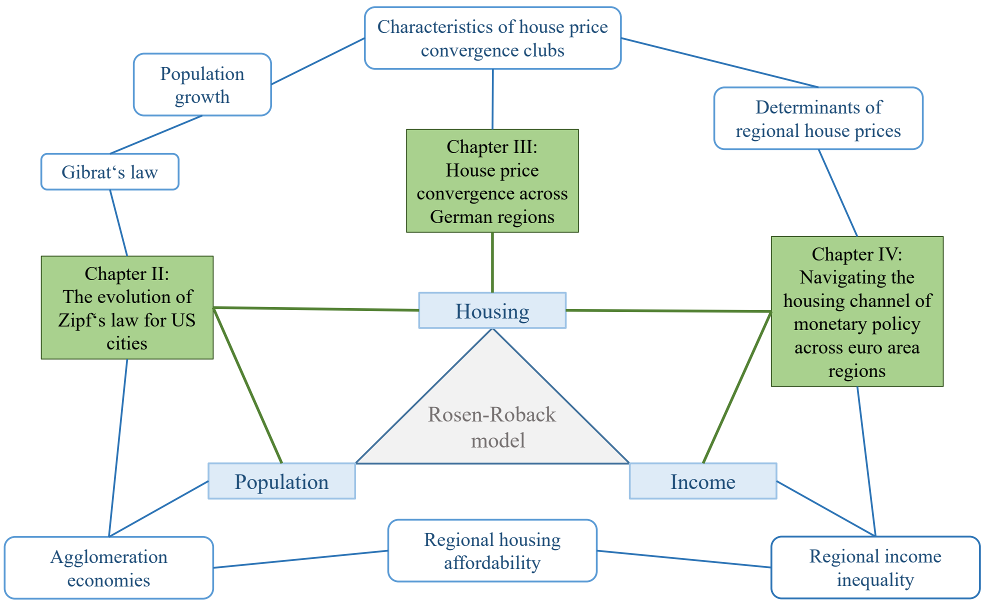
1.1 Introduction on regional housing markets

Housing markets have often been in the focus of policy makers, most notably during the Global Financial Crisis with its synchronized downturn in house prices in many countries (Claessens, Kose and Terrones, 2010). In recent years, the housing market—in light of rapidly increasing house prices and its pronounced local dimension—has come to the front of the media and policy debate again (The Financial Times, 2021, OECD, 2020, European Commission, 2021). In more detail, real house prices rose by 57% in Germany and by 62% in the United States (U.S.) between the years 2011 and 2021 (OECD, 2022). However, this development at the national level masks considerable heterogeneity across cities, states and regions with a pronounced difference between urban and rural areas: According to a nominal house price index by the Research Data Center FDZ Ruhr based on real estate offers, house prices increased on average by 52% in rural districts in contrast to an increase of 230% in the seven largest cities in Germany between the years 2008 and 2020 (RWI and ImmobilienScout24, 2022, Bauer et al., 2013).

This divergent development in regional housing markets is accompanied by and based on urbanization processes, which are reshaping city size distributions and have an impact on housing inequality and affordability (Thissen, Burger and van Oort, 2010, Ganong and Shoag, 2017). As reported in the latest projections by the United Nations, urbanization processes are expected to increase until 2050. While in the year 2000, 75% (79%) of the

German (U.S.) population lived in urban areas, the share is expected to grow to 84% (89%) in the year 2050. The growing concentration of population will increase the pressure on urban housing markets and stimulate a divergence process between urban and rural housing markets. This, in turn, impedes the inter-regional convergence process in human capital and per-capita income (Ganong and Shoag, 2017). Besides the regional perspective, housing market dynamics have an impact on the income distribution as they are closely related to changes in employment of lower income households (Mian and Sufi, 2014, Choi and Green, 2017, Cairó and Cajner, 2018) and have an uneven impact on disposable income, as well as consumption and saving patterns across income groups (Dustmann, Fitzenberger and Zimmermann, 2018).

Figure 1.1: Structure of the dissertation



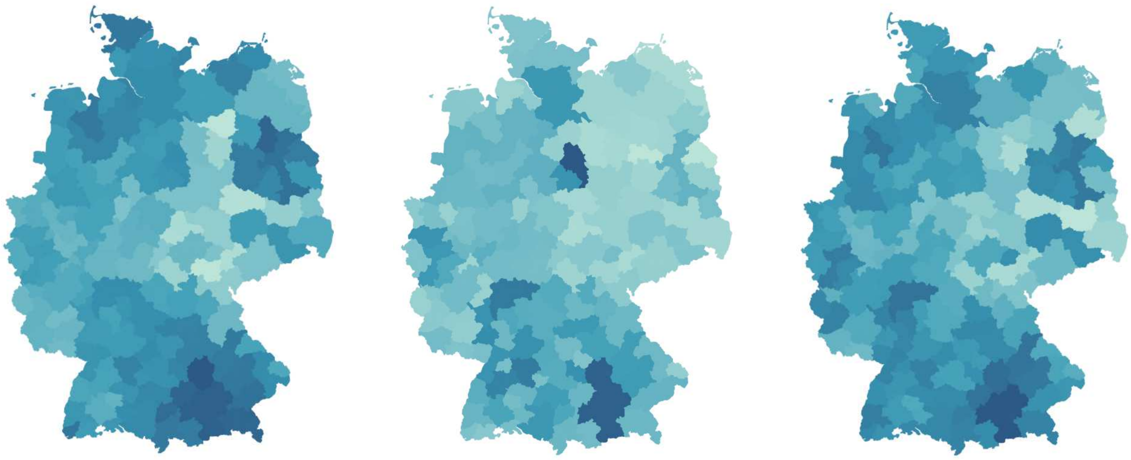
In light of potential consequences for inequality and housing affordability, the present thesis delivers a comprehensive contribution to several fields of studies related to regional housing markets. It comprises this introductory chapter and three scientific articles, which contribute to understanding the regional heterogeneity in housing markets, its origin as well as its implications. Chapter 2 deals with the evolutionary process of city size distributions, in particular the evolution of Zipf's law, and its implications for (sub-)urbanization processes. In chapter 3, the convergence process of regional housing markets and characteristics of house price convergence clubs are investigated. Chapter 4 assesses the role of regional housing markets in the transmission of monetary policy to economic activity and presents implications for regional inequality.

This introductory chapter will provide the foundation for embedding the subsequent chapters in the larger context of regional housing markets. Figure 1.1 displays the connection of the three scientific articles (chapters 2-4) together with related research topics, which will be discussed in this introductory chapter. To get a better understanding of why this dissertation focuses on the regional perspective, section 1.2 provides some stylized facts about regional housing markets—in the case of Germany. This is followed by the presentation of a theoretical framework, the Rosen-Roback model, which explains differences in local housing prices by local income levels and amenities (section 1.3). In addition, a modification by Moretti (2010) is presented to analyse migration patterns across cities. Relating thereto, section 1.4 deals in a first step with the question as to why are people still attracted to urban areas despite high house prices. It presents explanations for hierarchical city size distributions with a special focus on agglomeration economies. As a second step, the famous regularity of city size distributions, Zipf's law, is discussed. Section 1.5 focuses on regional house price dynamics. In particular, it gives an overview on various house price determinants and then it presents reasons for regional house price divergence. The latter is especially connected to urbanization processes and the development of city size distributions and has important implications for housing affordability and inequality. Section 1.6 presents the connection between monetary policy, the housing market and the real economy. It describes the overall mechanism how monetary policy transmits to economic activity and then it takes a closer look at the concrete housing channels at play. Section 1.7 describes the scientific articles (chapters 2-4) and section 1.8 gives an outlook on future challenges, which can be derived from the results of the articles.

1.2 Stylized facts about regional housing markets in the case of Germany

Most countries in the world are characterized by significant regional heterogeneity with regard to housing markets, economic outcomes and population developments. As an example, this section presents the spatial heterogeneity of German labour market regions. These 142 regions are defined from the 401 administrative districts in terms of existing commuting interrelations.

Figure 1.2: Regional heterogeneity



(a) House prices (b) GDP (c) Population

Sources: Research Data Center FDZ Ruhr (RWI Essen) and ARDECO database (European Commission).
Notes: The house price index for labour market regions is calculated as average of corresponding NUTS-3 regions, while population and GDP are sums of corresponding NUTS-3 regions. Panel (a): The house price index ranges between 75 and 204. Panel (b): GDP per capita ranges between 20,227 and 65,018. Panel (c): Population growth ranges between −13.4% and 10.3%.

A closer look at the regional dataset confirms a high degree of heterogeneity. As shown in panel (a) in Figure 1.2 the nominal house price indices in 2017 (2007=100) range between 75.0 in Saalfeld-Rudolstadt (Thuringia) and 203.7 in Ingolstadt (Bavaria).[1] A large proportion of regions with high house price indices can be found in the south of Germany with Ingolstadt and Munich (Bavaria) as the most expensive regions. Regarding the north of Germany, the labour market regions Oldenburg and Flensburg show the highest house price indices. Another area with high house price indices consists of Berlin and its surrounding labour market regions. Lower house price indices can mostly be found in Central Germany, the east of Germany (except regions surrounding Berlin), and in the south-west (Rhineland Palatinate).

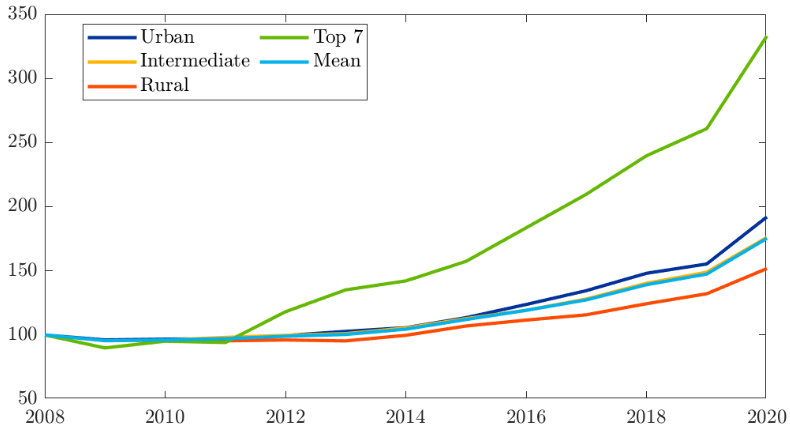
Panel(b) in Figure 1.2 provides a graphic illustration of income levels across the German labour market regions in 2017. It shows a large cross-regional heterogeneity with a minimum

[1]We employ the RWI-GEO-RED dataset providing a monthly house price index between 2007 and 2017 for German labour market regions by the FDZ Ruhr (research data center at the RWI - Leibniz-Institute for Economic Research). This house price index was calculated by means of a hedonic price regression and a granular dataset based on real estate offers published on the largest German listing website ImmobilienScout24. It contains information, inter alia, on the size of the house, its facilities, features, energy consumption and regional information to the $1km^2$ grid level. For a more detailed description on the granular dataset and the hedonic price regression, see Bauer et al. (2013).

GDP per capita of 20,227 EUR in Havelland (Brandenburg) and a maximum of 65,018 EUR in Wolfsburg (Lower Saxony). High income areas can mostly be found in the south of Germany, while lower income regions are mostly in Central Germany and the east of Germany.

Regarding population growth between the years 2007 and 2017, there is also a large regional dispersion (see panel (c) in Figure 1.2). While the labour market region Munich has grown the most between 2007 and 2017 (+10.3%), the region Elbe-Elster (Brandenburg) has decreased in population size by 13.36%. The regional dispersion is comparable to the one for house prices with the largest population increases in the south of Germany, especially around Munich, and the strongest decreases in population size in Central Germany and the east of Germany (except regions surrounding Berlin).

Figure 1.3: Regional house price development



Sources: Research Data Center FDZ Ruhr (RWI Essen) and Eurostat.
Notes: "Urban" represents the mean house price index of 86 regions, which are predominantly urban (excluding Top 7 cities), "Intermediate" the mean house price index of 191 regions,which are at an intermediate stage between urban and rural and "Rural" describes the mean house price index of 114 regions, which are predominantly rural.[2] "Top 7" shows the mean house price index of the seven largest cities (Hamburg, Berlin, Düsseldorf, Cologne, Frankfurt am Main, Stuttgart and Munich). "Mean" represents the mean of all 398 NUTS-3 regions in our panel.

The evolution of house prices in Germany in the 1990s and 2000s differs distinctly from the developments in other European countries (Jones, 2012, Westerheide, 2012). While prices in Germany were rather stable, with modest price movements and even drops, other countries, such as Ireland or Spain, experienced severe house price increases and booms. Gros

[2] The classification is based on the urban/rural typology of NUTS-3 regions by the European Commission (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Urban-rural_typology).

(2007) and Terrones (2004) relate this stagnation of German house prices to the reunification followed by a construction boom in the first half of the 1990s. Together with a slowdown in population growth (Ahearne et al., 2005) and the slow growth in real disposable income (Kholodilin, Menz and Siliverstovs, 2010), it resulted in a housing overhang. Since around 2010, however, real estate market dynamics in Germany have picked up and prices have grown at an accelerated but highly heterogeneous pace across regions in recent years (Deutsche Bundesbank, 2020, Dahl and Goralczyk, 2017). Figure 1.3 shows this divergence process of urban and rural house prices over the last decade. The 401 German administrative districts can be categorized as rural, intermediate and urban areas, whereby the 7 largest cites are excluded from the group of urban areas and considered as separate group. While the sharp increase in prices starting in 2012 was largely concentrated in Germany's Top 7 cities,[3] it only took off gradually in non-urban regions. As of roughly 2015, the large upward pressure on nominal residential prices has become more widespread across German regions. While house prices increased on average by 52% between the years 2008 and 2020 in rural districts, house prices in the seven largest cities increased by 230% during the same time period. This severe acceleration of house price growth has caused the European Systemic Risk Board (ESRB) to issue a warning to Germany given its systemic risk to financial stability. The board noticed a significant overvaluation of house prices in urban areas reflecting a shortage of housing supply relative to demand (European Systemic Risk Board, 2019, European Systemic Risk Board, 2022). These results are in line with Hertrich (2020), who finds that the German housing market was overvalued by 11.3% in 2018. This overvaluation is mostly driven by an increased interest rate risk and a relatively advanced stage of the housing investment cycle.[4] The so-called debt-servicing capacity, on the other hand, which reflects a change in household income and/or interest rate environment, has a dampening effect (Hertrich, 2020).

[3]The Top 7 cities are Berlin, Hamburg, Munich, Cologne, Frankfurt am Main, Stuttgart, Düsseldorf.

[4]Interest rates, which are low for an extended time period, can generate housing price misalignments (O'Meara, 2015, Musso, Neri and Stracca, 2011, Zhu, Betzinger and Sebastian, 2017, Hülsewig and Rottmann, 2021). Regarding housing investment, Leamer et al. (2007) shows that most recessions are preceded by excessive housing investments relative to GDP. As stated by Hertrich (2020), the 2002-2003 recession in Germany was also preceded by unusually high housing investment-to-GDP ratios by historical standards.

1.3 Rosen-Roback theoretical framework

Why is housing so expensive in some places? Why do people still cluster in cities? Why do workers accept lower real wages in some places? These questions can be addressed by means of a spatial equilibrium model, the Rosen-Roback model (Rosen, 1979, Roback, 1982). In this tradition, high prices for housing must reflect in high income and/or high amenities, which means that there is a trade-off between income, amenities and housing costs. As a consequence, first, the flow of wages plus amenities minus housing costs are roughly the same between locations and individuals are indifferent across space. Second, differences in wages reflect differences in productivity so that firms are indifferent between locations and hiring workers. Third, builders must be indifferent between building and not building meaning that house prices cannot exceed the construction costs by much. In line with Friedman (1975) stating "there is no such thing as a free lunch", these three no-arbitrage relationships are the foundation of the Rosen-Roback model.

1.3.1 Baseline Rosen-Roback model

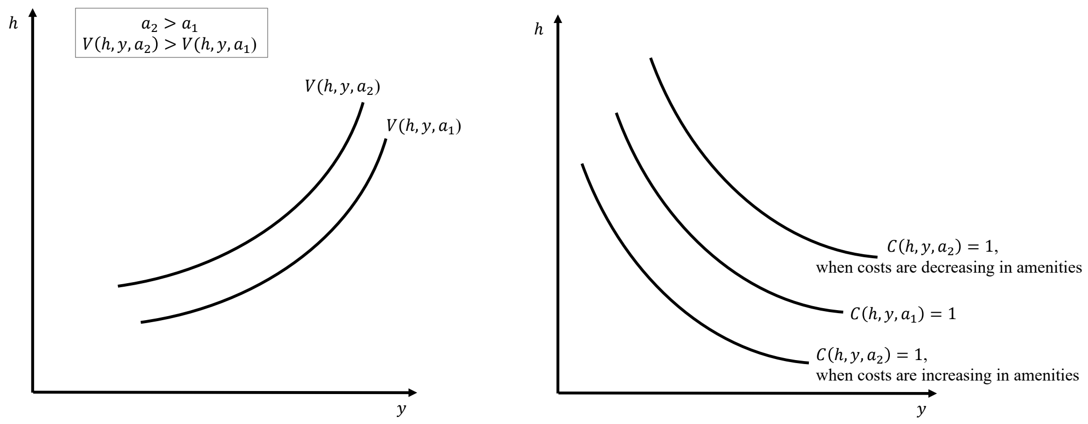
To understand spatial differences in housing markets and income levels, as a first step, we consider the baseline Rosen-Roback model. It is a general equilibrium model with capital and labour being completely mobile across cities. Housing costs and wages can differ between cities to account for differences in the cities' amenities and housing supply is fixed within cities, but assumed to be interchangeable between uses within a city. In a given city, the worker's utility level depends on income y, housing prices h and the city's amenities. It can be summarized by the indirect utility function

$$V(y, h, a) = k. \tag{1.1}$$

The worker's utility increases with increasing income y, decreasing prices for housing h and increasing amenities a. In order to compare cities with different amenities, we use indifference curves, which show, for a given level of amenity, the combinations of h and y, that yield the same utility (Figure 1.4 (a)).

On the production side, firms incur costs from labour (y) and from real estate (h). The firms' costs can also depend on the city's amenities: An amenity to workers can either also be

Figure 1.4: Indifference and Iso-profit curves with different amenity levels



(a) Indifference curves

(b) Iso-profit curves

Source: Author's illustration based on Roback (1980).
Notes: Panel (a): The upper curve is the worker's indifference curve for a higher level of amenity ($a_2 > a_1$). Panel (b): The upper (lower) curve shows the firm's iso-profit curve for a higher level of amenity ($a_2 > a_1$) when the firm's costs are decreasing (increasing) in amenities.

an amenity to firms, but it can also be a disamenity to firms and increase their costs. This will have different implications when comparing high and low amenity regions. The production function is assumed to have constant-returns-to-scale and the equilibrium condition for firms is that unit cost must equal product price, which is assumed to be unity. Otherwise, firms would relocate to more profitable cities.

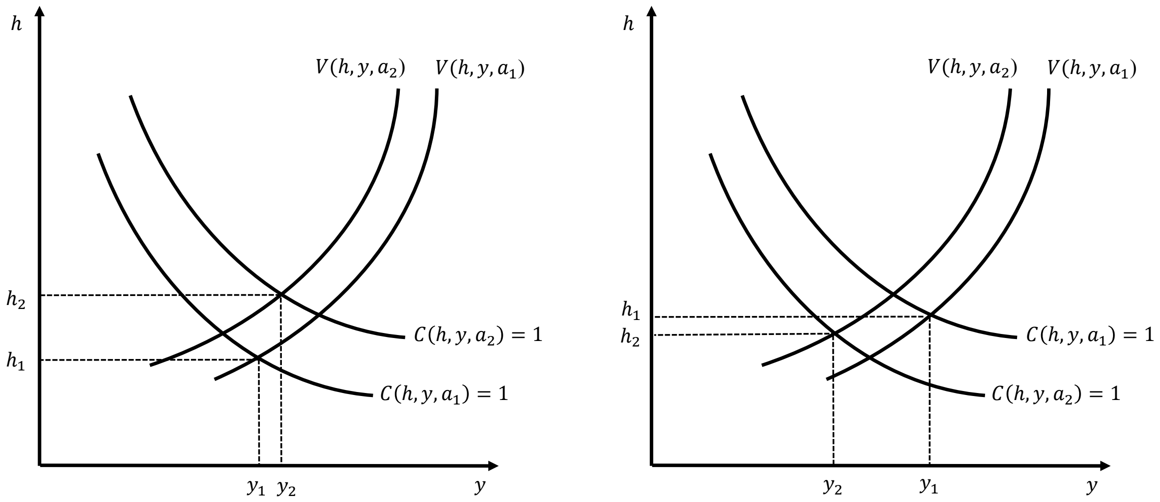
$$C(y, h, a) = 1. \tag{1.2}$$

This condition generates iso-profit curves, which show for a given level of amenity the combinations of h and y that yield zero profit (Figure 1.4 (b)).

As a next step, differences in the equilibrium for low- and high-amenity cities are considered for two cases, in which consumers value the amenity and (1) it decreases the firms' costs and (2) it increases the firms' costs.[5] In the first case (Figure 1.5 (a)), the high amenity region yields a higher level of utility for the consumers and thereby a higher indifference curve as well as a higher iso-profit curve, because the amenity decreases the firms' costs. Consequently, the high amenity region reflects in higher income ($y_2 > y_1$) and higher

[5]An example for case (1) would be the amenity clean air, which is appreciated by workers, but the firms' costs increase by having to use non-polluting technology. A productive amenity, case (2), could be the lack of severe snow storms, which would be appreciated by workers and decrease the firms' costs (Roback, 1982).

Figure 1.5: High- and low-amenity regions



(a) Amenity decreases firms' costs

(b) Amenity increases firms' costs

Source: Author's illustration based on Roback (1980).
Notes: Panel (a): An increase in amenities ($a_2 > a_1$) shifts the firm's iso-profit curve upwards as the firm's costs are decreasing in amenities. Panel (b): An increase in amenities ($a_2 > a_1$) shifts the firm's iso-profit curve downwards as the firm's costs are increasing in amenities.

house prices ($h_2 > h_1$). However, whether the income level increases, depends on the shift in the iso-profit curve, so by how much the amenity decreases firms' costs. For the second case (Figure 1.5 (b)), in which the amenity increases firms' costs, the iso-profit curve is shifted down for the high-amenity region. As a result, the high-amenity region has lower income levels ($y_2 < y_1$) and house prices ($h_2 < h_1$). Whether the house prices are actually lower in the high-amenity region, depends on how much the amenity increases firms' costs and thereby how much the iso-profit curve shifts. In general, this model shows that differences in amenities across cities can be adjusted for by house prices and income. Furthermore, the outcome depends on whether the firms' costs increase or decrease with the amenity and on how much they are affected.

1.3.2 The Rosen-Roback model and migration between cities

As a next step, the assumption that labour is perfectly mobile is modified, as people may have idiosyncratic preferences for a city, such as having family or being born there. Furthermore, the elasticity of housing supply is assumed to vary across locations due to geography and local land regulations. These modifications of the original Rosen-Roback framework by Moretti

(Moretti, 2010) can be used to analyse how an improvement in amenities in one city affects migration between cities.

Under the assumptions that (1) each city produces the same output, (2) each worker provides one unit of labor, and (3) firms and workers will locate where their utility and profit are maximized, the indirect utility of worker i in city C is given by:

$$V_{iC} = w_C - h_C + a_C + \epsilon_{iC}. \tag{1.3}$$

The real wage is the difference between the nominal wage w_C and housing costs h_C in city C, a_C represents local amenities and ϵ_{iC} worker i's idiosyncratic preferences for city C. Given two cities, A and B, worker i's relative preference for city A over B is

$$\epsilon_{iA} - \epsilon_{iB} \sim V[-s, s], \tag{1.4}$$

where s represents the degree of labour mobility. A large s means that worker i's idiosyncratic preference for city A is large and the willingness to move to city B in case of higher real wages or amenities in B is limited. If s is small, workers are more mobile and in the extreme case $s = 0$, there are no idiosyncratic preferences and worker mobility is perfect. Worker i chooses city A over B, if

$$\begin{aligned} V_{iA} &> V_{iB} \\ \epsilon_{iA} - \epsilon_{iB} &> (w_B - h_B) - (w_A - h_A) + (a_B - a_A). \end{aligned}$$

In equilibrium, the marginal worker needs to be indifferent between the cities A and B, which yields labour supply in city B as

$$w_B = w_A + (h_B - h_A) + (a_A - a_B) + s\frac{N_B - N_A}{N}. \tag{1.5}$$

N_A and N_B are the endogenously determined log number of workers in city A and B, respectively. The sum of N_A and N_B is equal to N and assumed to be fixed. The elasticity of labour supply in city B depends on the worker's idiosyncratic preferences s for city B. If s is small, the workers are relatively more mobile and the elasticity of labour supply is high, yielding a relatively flat labour supply curve. This is a modification of the original Rosen-Roback framework, which assumes all workers to be identical and indifferent across locations. The production function is given by a Cobb-Douglas function with constant returns to scale

$$\ln y_C = X_C + jN_C + (1-j)K_C, \tag{1.6}$$

where X_C is a productivity shifter in city C and K_C is the log of capital. Under the assumptions that firms are price takers and labour is paid its marginal product as well as perfect mobility of firms, labour demand is given by

$$w_C = X_C - (1-j)N_C + (1-j)K_C \ln j. \tag{1.7}$$

By re-arranging equation (1.5) and assuming that each worker consumes one unit of housing, we receive the demand for housing in city B

$$r_B = (w_B - w_A) + r_A + (a_B - a_A) - s\frac{N_B - N_A}{N}. \tag{1.8}$$

The supply of housing is given by

$$r_C = z + k_C N_C, \tag{1.9}$$

where k_C represents the elasticity of housing supply, which is exogeneously determined by a city's geography and land regulations. The more constrains there are to build new houses, the larger is the parameter k_C. The number of housing units in city C equals the number of workers in city C. The equilibrium on the labour market and on the housing market is given by equating equations (1.5) and (1.7), and equations (1.8) and (1.9), respectively.

What happens to a local economy in case of a labour supply shock, which is generated by an increase in the amenity level in a city? To answer this question, we consider two cities A and B over two time periods and the case that city B becomes more desirable for workers relative to city A in period 2. So, the amenity level increases in city B: $a_{B2} = a_{B1} + \Delta$, with $\Delta > 0$. Attracted by the higher amenity level, workers move from city A to B

$$N_{B2} - N_{B1} = \frac{N}{N(k_A + k_B) + 2s}\Delta \geq 0. \tag{1.10}$$

This indicates that more workers move from A to B when idiosyncratic preferences for location are less important (smaller s) and when there are fewer constraints on housing supply in city B (smaller k_B). The cost of housing in city B will increase as a consequence of in-migration and the magnitude depends on the increase in amenities Δ and the housing supply elasticity in city B relative to A[6]:

$$h_{B2} - h_{B1} = \frac{k_B N}{N(k_A + k_B) + 2s} \Delta \geq 0. \tag{1.12}$$

As the nominal wages in both cities do not change, real wages decrease in city B

$$(w_{B2} - w_{B1}) - (r_{B2} - r_{B1}) = -\frac{k_B N}{N(k_A + k_B) + 2s} \Delta \leq 0 \tag{1.13}$$

and increase in city A

$$(w_{A2} - w_{A1}) - (r_{A2} - r_{A1}) = \frac{k_A N}{N(k_A + k_B) + 2s} \Delta \geq 0. \tag{1.14}$$

As a consequence of this increase in amenities in city B, a certain amount of workers moves from city A to B, they are accepting lower real wages to live in the more desirable city and landowners in B experience an increase in their property value, while the opposite holds true for landowners in city A.

1.4 The size distribution of cities

Cities are very heterogeneous. They differ in their size, density, economic base, share of high- and low skilled workers as well as their productivity. As a first step, this section presents explanations for hierarchical city size distributions with a special focus on agglomeration economies and as a second step the famous regularity of city size distributions, Zipf's law, is discussed. Examining the evolution and development of urban systems is of great importance for regional housing markets as it allows propositions about future urban development, which affects housing demand and the housing market in general. Thissen, Burger and van Oort (2010), for example, empirically validate for the U.S. the positive effect of city size on urban

[6] The housing costs in city A will decrease:

$$h_{A2} - h_{A1} = \frac{k_A N}{N(k_A + k_B) + 2s} \Delta \geq 0. \tag{1.11}$$

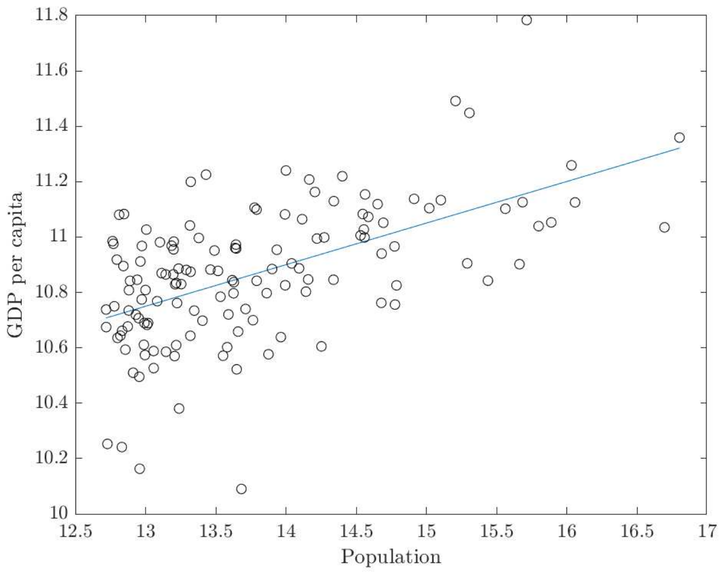
house prices.

1.4.1 Agglomeration economies

Having considered the theoretical interactions between income, amenities and housing costs as well as the impact of amenity shocks on migration between cities, we will now take a closer look at these amenities. What are the reasons that people are still attracted to urban areas despite high house prices and the overall high cost of living there? First of all, locational fundamentals can explain around one-fifth of the observed geographical concentration (Ellison and Glaeser, 1999). Locational fundamentals pin down city locations and attract or repulse population and economic activity. Examples for these fundamentals are natural resources, infrastructure, climate or consumption and production amenities (Behrens and Robert-Nicoud, 2015, Davis and Weinstein, 2002).

Besides local fundamentals, and presumably the most important reason, why large proportions of the population cluster in cities and urban areas, are agglomeration effects. In general, agglomeration economies entail that firms can produce higher output with the same amounts of input. Theoretically, there are two major strands explaining these agglomeration externalities: Specialization (Marshall, 1890) and diversification (Jacobs, 1969). According to Marshall (1890), local concentration of production of the same industry results in asset-sharing, meaning easier access to specific goods or services by specialized suppliers, as well as the creation of a local labour market pool, engendering an increase in production. Furthermore, the regional concentration of firms of the same industry can increase knowledge spillovers and thereby the stock of knowledge available to each individual firm. These specialization externalities may enhance the firms' ability to innovate (Marshall, 1890, Arrow, 1962, Romer, 1986 and Feldman and Audretsch, 1999). In contrast to this idea, Jacobs (1969) argues, that knowledge from one industry can be applied to problems in other industries, so that firms benefit from a high level of industrial diversity. The exchange of complementary knowledge across industries may facilitate and enhance innovation processes, which lead to increasing returns (Jacobs, 1969). Based on these two theories, larger cities can be linked to a local expansion of knowledge due to high levels of industrial specialization and diversification. Figure 1.6 illustrates the presence of agglomeration economies for the 135 largest U.S. metropolitan areas. There is a strong positive relationship between economic activity per capita and population. More precisely, the elasticity of economic activity per capita with respect to urban population is 0.15 (standard error 0.02). This is in line with Glaeser and Gottlieb (2009), who detect a slope coefficient of 0.13 for U.S. metropolitan areas.[7]

Figure 1.6: Regional population and economic activity



Source: OECD Database.
Notes: The x-axis plots log population and the y-axis log GDP per capita for the 135 largest U.S. metropolitan areas in 2020.

Regions also differ fundamentally in their human capital. Chen and Rosenthal (2008) find evidence for the U.S. that especially highly skilled workers are attracted to high quality business environments, whereas retirees are rather drawn to regions with low costs of living and high consumer amenities. Consequently, there are regions which are heavily populated by skilled workers. A concentration of skilled human capital leads to a lower probability of unemployment and higher wage rates, which in turn attract more people and result in higher population densities in a few regions (Krugman, 1991,Krugman, 2011). This is in line with Black and Henderson (1999), who identify a strong positive relationship between individual city population growth rates and local human capital growth rates. Furthermore, Glaeser et al. (2004) state that human capital predicts population and productivity growth at the city and metropolitan area level. For Germany, this is confirmed by Growe (2010), who finds

[7] Grover, Lall and Timmis (2021) figure out a meta-analysis of 1,200 estimates of agglomeration elasticities from 70 studies covering 33 countries over the period 1973 to 2020. They find agglomeration elasticities of 1 to 5 percent for developed countries. However, there is a large heterogeneity within these estimates due to differences in the outcome and agglomeration measure, consideration of urban costs, and the variation in pay-off across sectors and skill.

that workers with knowledge-based professions are highly attracted to urbanized areas and that most of these professions are over-represented in highly urbanized areas.

Urban agglomerations do not only enhance productivity of firms based in that area, but they also stimulate innovation processes and entrepreneurial activity, which lay the foundation for future (productivity) growth (Porter, 1998). Due to the access to information, technology, a pool of specialized workers and a specialized supplier base as well as the exposure to competitive pressure and constant comparison, firms can experiment at lower costs and innovation and entrepreneurial activity can be enhanced (Porter, 1998).

The most important drivers of innovation are likely knowledge spillovers. As stated by Feldman (1994), the exchange of information is beneficial to firms and it reduces the uncertainty and costs of innovation. Various studies measure knowledge spillovers by means of wage regressions. Glaeser and Mare (2001) find, for example, that workers learn more quickly in dense metropolitan areas, which may lead to an increase in human capital and thereby more experimentation and innovation. This results in an urban wage premium in levels and growth in larger cities (Heuermann, Halfdanarson and Suedekum, 2010, Baum-Snow and Pavan, 2013, De la Roca and Puga, 2012) and D'Costa and Overman (2014) show that the urban wage premium increases in city size. Rooted in agglomeration economies, higher worker productivity in dense labour markets results in higher wages in urban areas and thereby an urban wage premium (Puga, 2010, Moretti, 2010). Regarding German labour markets, Hirsch et al. (2022) find that the urban wage premium can not only be explained by higher productivity in denser labour markets, but also by a larger share of the marginal product of labor, which the workers receive. Regarding patent-based evidence, Koch and Simmler (2020) identify three local public knowledge spillover channels. First, public institutions present scientific knowledge, which spills over to applied researchers. These technological spillovers are enhanced by geographical proximity (Belenzon and Schankerman, 2013). Second, local firms may collaborate with public institutions to create new knowledge. And again, geographical proximity matters, as face-to-face interactions are still preferred (Rybnicek and Königsgruber, 2019). Third, firms may benefit from non-specific public knowledge in their region. Koch and Simmler (2020) find that this non-specific public knowledge channel accounts for around two-thirds of the overall local knowledge spillovers of public research and development in Germany. Various other studies find a significant local dimension in patent citations confirming the localized nature of knowledge flows for innovation (Jaffe, Trajtenberg and Henderson, 1993, Thompson, 2006, Carlino et al., 2010).

However, the equilibrium city size distribution is not only a result of agglomeration economies. It is rather a trade-off between agglomeration economies and urban crowding costs (Duranton and Puga (2014)). Urban costs can be costs for housing/land, traffic congestion, commuting costs or other urban disamenities. In general, while agglomeration economies make (research) labour more efficient and innovative in larger cities, crowding costs make it more costly (Duranton, 2007). While there is a large strand of literature examining agglomeration economies, only a few studies focus on urban crowding costs. As an example, Combes, Duranton and Gobillon (2012) find an elasticity of crowding costs with respect to city size of 0.03 using French data. The authors state that wage increases due to population growth in smaller cities will be higher in absolute terms than the rise in urban costs. For larger cities Combes, Duranton and Gobillon (2012) state the opposite: Population growth leads to higher wages, which will eventually be dominated by rising urban costs. Consequently, the net benefit from cities should be bell-shaped. Rosenthal and Strange (2004) argue that the elasticity of productivity to city sizes ranges between 3 and 8 percent (gross agglomeration economies) and Duranton (2007) expects this elasticity to decreases to a range of 0 and 5 percent when taking crowding costs into account (net agglomeration economies).

1.4.2 Zipf's law

After having considered determinants of agglomeration economies and their impact on city size distributions, we are now turning to a key fact about the size distribution of cities: Zipf's law.[8] It states that the rank of a city r is inversely related to its size $P(r)$ (Zipf, 1949)

$$r = \frac{P_1}{P(r)}, \tag{1.15}$$

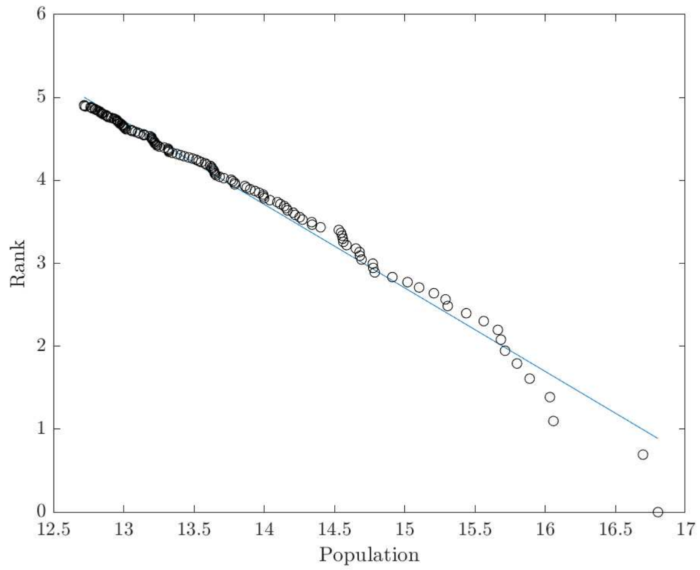
with P_1 showing the size of the largest city in the urban hierarchy.

To visualize this relationship, Figure 1.7 shows a scatterplot with the log size (i.e. population) of the 135 largest U.S. metropolitan areas on the x-axis and the log rank on the y-axis. The dots are shaped as a straight line and running a regression yields

$$\ln r = 17.82 - 1.008 \ln P(r) \tag{1.16}$$

[8] Mansury and Gulyas (2007) find that Zipf's law breaks down unless the extent of agglomeration economies exceeds the negative disagglomeration forces.

Figure 1.7: Zipf's law



Sources: OECD Database.
Notes: The x-axis plots log population and the y-axis plots the log rank of the 135 largest U.S. metropolitan areas in 2020.

with $R^2 = 0.98$. Hence, the size distribution of the largest U.S. metropolitan areas can be described by Zipf's law. This striking fit of Zipf's law for the U.S., but also other city size distributions, has generated many empirical studies and attempts at theoretical explanations. In the following, these theoretical frameworks will be reviewed succeeded by an empirical literature overview.

1.4.2.1 Theoretical underpinnings for Zipf's law

There are several theoretical underpinnings for Zipf's law. A large strand of the literature proposes various economic shocks in combination with Gibrat's law as foundation for Zipf's law. Alternatively, human capital accumulation and central place theories are used to explain the existence of Zipf's law.

The first group of models is based on the interaction between growth processes of cities and economic activity. Gabaix (1999) proposes a model with city growth rates, which are identically and independently distributed for all cities, regardless of their size. Given random amenity shocks and a reflective lower bound for city sizes, the model generates a city size distribution that follows Zipf's law at least in the upper tail. This process is based on Gibrat's law (Gibrat, 1931), which states that the proportionate growth process of cities, meaning growth rates of cities do not depend on the cities' sizes, results in a Pareto size distribution with Pareto exponent equal to one (Zipf's law). Córdoba (2008*a*) allows the variance of growth rates to depend on city sizes but not the mean and finds that this variation of Gibrat's law results in a Zipf exponent different from one.

Eeckhout (2004) develops a model in which a fixed number of equally old cities are hit by random productivity shocks. In this setting with free mobility, population is a function of productivity and the log normal distribution of city productivity results in a log normal city size distribution. Gibrat's law holds in this model and Zipf's law emerges when a lower bound for the city size is added. The model of Rossi-Hansberg and Wright (2007) also relies on exogenous productivity shocks, but in contrast to Eeckhout (2004), these shocks are industry-specific and there are no idiosyncratic productivity differences between cities. Once a productivity shock hits an industry, the production in cities, which are specialised in this industry, increases together with the city's population. These migration processes eliminate local increasing returns to scale and lead to constant returns to scale for the urban system in aggregate.

By means of innovation-driven shocks at the level of industries and cities, Duranton (2007) provides a simple mechanism to explain the growth and decline of cities. The model is based on the assumption that cities comprise multiple industries and that these industries are attracted to places where innovations are high. Duranton (2007) improves the model by including standard urban features, agglomeration economies and crowding costs, which make research labour more innovative and more costly, respectively, in larger cities. Unlike other random growth models, it does not only predict that the size distribution of cities follows an approximate Pareto distribution. It also replicates the fast churning of industries across cities (for West Germany, see also Findeisen and Südekum, 2008). Córdoba (2008*b*) proposes a standard urban model, which produces a Pareto distribution of city sizes, under the conditions that all cities must have the same expected growth rate and that the underlying source of randomness is also distributed Pareto.

In contrast to these random growth models, Hsu (2012) proposes a theory of city size distribution by means of a hierarchy approach based on central place theory. This model suggests that city size differences occur due to heterogeneity in the economics of scale across goods. Furthermore, it links central place theory and the power law by demonstrating that the power law for cities arises from regularly varying distributions of scale economies. Behrens and Robert-Nicoud (2015) can generate Zipf's law with an extension of the canonical model proposed by Henderson (1974). They show how locational fundamentals, agglomeration economies, the spatial sorting and selecting of heterogeneous agents affect the city size distribution.

1.4.2.2 Empirical evidence of Zipf's law

Empirically, Zipf's law can be confirmed either by estimating the log-rank log-size regression or by validating Gibrat's law of proportional growth. The great majority of studies uses cross-sectional data to check whether or not Zipf's law holds exactly for a system of cities.[9] For instance, Krugman (1996) and Gabaix (1999) use data for U.S. Metropolitan Statistical Areas (MSAs) and find that the one-parameter Zipf model holds exactly for a minimum threshold of 280,000 inhabitants. These findings are recently confirmed by Schmidheiny and Suedekum (2015) using novel data from an EC-OECD project. Zipf's law also occurs when applying other city definitions, like economic areas (Berry and Okulicz-Kozaryn, 2012), natural cities (Jiang and Jia, 2011) or geographic clusters (Rozenfeld et al., 2011). Some studies find opposing results for the U.S. city size distribution (Eeckhout, 2004) or detect that Zipf's law only holds for the upper tail of the distribution while the body and lower tail are lognormal (Levy, 2009, Malevergne, Pisarenko and Sornette, 2011 and Ioannides and Skouras, 2013). Using U.S. census data, Soo (2005) finds that the largest cities are more evenly and the largest urban agglomerations are more unevenly distributed than predicted by the exact Zipf's law (also see Gan, Li and Song, 2006 and Ioannides and Overman, 2003). Other studies show that Gibrat's law only weakly holds for the U.S. urban system, e.g., only for the upper-tail of the city size distribution, and consequently reject Zipf's law (González-Val, 2010, González-Val, 2012, González-Val, Lanaspa and Sanz-Gracia, 2014).

Regarding the German city size distribution, there are only few studies examining whether Zipf's law applies. Recently, Budde and Neumann (2019) define urban areas according to variable thresholds of population density across 1 square kilometer grids by a clustering algorithm. The authors find an increasing downward deviation of the Zipf parameter from value 1 with an increasing scale of urban territories assigned to agglomerations. However,

[9] A detailed literature review on the empirical findings on Zipf's law is given by Arshad, Hu and Ashraf (2018).

the parameter deviates upwards when restricting the considered territories to more densely populated core areas. Giesen and Südekum (2011) detect Zipf's law at both national and regional levels. For the regional level, the authors consider the federal states, random regions, as well as large cities that are geographically adjacent. In line with a proposition by Gabaix (1999), which states that if Gibrat's law holds for each region in a country, then Zipf's law will hold for each region and also for the national city size distribution, Giesen and Südekum (2011) find that Gibrat's law holds not only at the national level in Germany, but in each region regardless of which type.

Focusing on the long-term perspective of Zipf's law, again, the results depend on the employed city definition. For U.S. MSAs, Black and Henderson (2003), Dobkins and Ioannides (2000) and Dobkins and Ioannides (2001) find an increasing urban concentration, which is higher than predicted by the exact Zipf's law. Other authors focuse on states (Soo et al., 2012), counties (Beeson, DeJong and Troesken, 2001 and Desmet and Rappaport, 2017) or minor civil divisions (Michaels, Rauch and Redding, 2012) in the U.S.. González-Val (2010) compares U.S. incorporated places and finds that the city sizes are lognormally distributed and more unequally distributed than predicted by the exact Zipf's law. For the upper tail of the city size distribution, the author finds that the cities become more equally distributed over time. This is in line with the results in chapter 2. The article reveals that for most of the years between 1840 and 2016 the U.S. city size distributions are more equally distributed than expected by the exact Zipf's law and that they have become more equally distributed over time. In contrast to González-Val (2010), who explains the convergence of the city sizes with a loss of importance of the largest cities, for most of the time span, we find evidence for leading cities dominating the remaining largest cities.

1.5 Regional house price dynamics

A departure from Zipf' law and therewith transformation processes in city size distributions have important implications for the dynamics in regional housing markets. As a first step, this section presents several determinants of house prices. Second, taking the regional perspective on housing markets, theoretical grounds for regional house price divergence will be presented. Whether regional house prices are converging or diverging over time has an important impact on (regional) inequality and housing affordability, which will be in the focus of section 1.5.2.2.

1.5.1 Determinants of house prices

Based on the literature, we can identify six groups of house price determinants: (1) demand side, (2) supply side, (3) structural, (4) fiscal, (5) macroprudential, and (6) monetary factors.

The main determinants of house prices are demand side factors, such as income, demographics, labour market characteristics and real interest rates as well as supply side factors, such as the housing stock and construction costs. For the UK, Cameron, Muellbauer and Murphy (2006) apply a dynamic equilibrium-correction equation system and find that changes in house prices between 1997 and 2003 can largely be explained by income and population growth as well as house building. This is confirmed for London house prices between 1983 and 2016 by Sivitanides (2018), who detects that London house prices are strongly related to London population, housing completions and national GDP. Miles (2012) presents a model of the housing market, which identifies a positive relationship between house prices and population density, stating that if population density is on an upward trajectory, increases in population and incomes, further increase prices and diminish rises in new housing. Besides the role of income, Xu and Tang (2014) identify the importance of interest rates and construction costs for UK house price development. In line with these results, an early study regarding U.S. MSAs by Abraham and Hendershott (1996) shows that changes in equilibrium house prices can be explained by changes in income, interest rates and construction costs. Besides these factors, Jud and Winkler (2002) find that real house price appreciation in MSAs is strongly influenced by population growth, which is also confirmed by Holly, Pesaran and Yamagata (2010) on the state level. These supply side and demand side determinants of house prices are also identified in various cross-country analyses. Caldera and Johansson (2013), for example, estimate a stock-flow model of the housing market within an error correction framework and find for 21 OECD countries that house prices tend to increase with households' disposable income, population and decrease with the stock of housing and interest rates.[10] In addition to changes in economic fundamentals (income, interest rates, stock prices and inflation), by means of a multivariate unobserved component model, Algieri (2013) detect unobserved factors, such as structural changes in the markets and changing preferences, to play a key role in explaining house price changes.

Structural factors, which have an impact on house prices, can comprise local regulatory

[10] These results are in line with other cross-country samples, such as Égert and Mihaljek (2007) and Adams and Füss (2010).

and geographic constraints as well as rental regulation. By means of a theoretical framework, Hilber and Vermeulen (2016) model the impact of local supply constraints on local house prices. Using a panel dataset of local planning authorities in England between 1974 and 2008, the authors detect a substantial positive impact of local regulatory constraints on the response of local house prices to changes in local earnings. House prices would have been about 35% lower in 2008 if regulatory constraints had been absent (Hilber and Vermeulen, 2016). This relationship between regulatory constraints and house prices has also been widely studied for the U.S.. According to work by Albouy and Ehrlich (2018), fewer regulatory and geographic restrictions substantially decrease house prices relative to land and construction input costs (see also Saiz, 2010, Saks, 2008, Glaeser, Gyourko and Saks, 2005*a* and Glaeser, Gyourko and Saks, 2005*b*). In contrast, Glaeser and Ward (2009) do not find a significant impact of local land use regulation on house prices across municipalities in the Greater Boston area. With a focus on house price volatility, Paciorek (2013) shows that regulation can explain a large fraction of the difference in volatility of house prices between highly regulated and relatively unregulated areas. Furthermore, stringent regulatory and geographic restrictions decrease the responsiveness of investment to demand shocks, which in turn amplifies house price volatility. Another structural factor impacting house prices is rent regulation. Using a panel dataset of 25 countries between 1980 and 2017, Cavalleri, Cournède and Özsöğüt (2019) find that tighter rent controls are linked with lower supply elasticities, which positively impact the response of house prices to demand shocks. Presumably due to a different sample period, which covers very limited cross-country variation in rent regulation, Bétin and Ziemann (2019) find no significant impact of rent control on house prices.

House prices can also be affected by fiscal policy measures, such as transaction and property taxes, interest relief and the degree of lender recourse on mortgages. By means of a theoretical model, which is matched for U.S. data, Sommer and Sullivan (2018) show that an elimination of the regressive mortgage interest deduction leads to rising after-tax costs of occupying a square foot of housing, and consequently, to a decrease in house prices. For a large set of countries, Kuttner and Shim (2016) find that an increase in housing-related taxes, such as property taxes, transaction taxes, stamp taxes and less generous deductibility of mortgage interest slow real house price appreciation. These results are in line with Crowe et al. (2013), who observes that an increase in property tax rates in U.S. MSAs results in a decline in house price growth and thereby limits house price booms. In addition, Muellbauer

(2005) and Poghosyan and de Mooij (2016) state that higher property taxes can stabilize house prices and lower their volatility. The probability for house price booms can also be reduced by the degree of lender recourse on mortgages. Cerutti, Dagher and Dell'ariccia (2017) detect a lower probability of house price booms in countries where borrowers face the downside risk from full recourse.

Another group of factors determining house prices is concerned with macroprudential regulation. By means of a large cross-country study including 57 advanced and emerging countries, Akinci and Olmstead-Rumsey (2018) detect that macroprudential tightening is associated with lower house price growth. As measures of macroprudential regulation, the authors mention e.g. capital requirements, dynamic loan-loss provisioning as well as caps on Loan-To-Value (LTV) or Debt-To-Income (DTI) ratios. For a panel of 56 countries, Richter, Schularick and Shim (2019) confirm that changes in the LTV limit have a substantial effect on house prices. Even for a large panel of 119 countries, Cerutti, Claessens and Laeven (2017) provide empirical evidence that borrower-based policies, such as limits on LTV and DTI ratios, as well as financial institutions-based policies, such as limits on leverage and dynamic provisioning, are effective in reducing the growth rates of credit and house prices. In contrast, Alam et al. (2019) and Kuttner and Shim (2016) find only a weak impact of macroprudential measures on house prices. Focusing on 28 EU countries and considering 99 lending standard restrictions, Poghosyan (2020) shows that these lending restriction measures are generally effective in curbing house prices. The impact reaches its peak after three years. This is in line with Gross and Población (2017), who find that Debt-Service-To-Income (DSTI) and LTV ratio caps can curb house price growth for four European countries - Austria, Belgium, Germany and Portugal. Cerutti, Dagher and Dell'ariccia (2017) provide empirical evidence that house price booms are more likely in countries with higher LTV ratios.

Besides changes in interest rates, there are various other direct and indirect channels, through which monetary policy can influence housing markets. These channels will be discussed in greater detail in section 1.6.2.

1.5.2 Regional house price divergence

After having determined several factors influencing house prices, we are now turning to the regional perspective on housing markets, which is the focus of this dissertation, and answer the following questions: Why are regional house prices diverging? How can regional house price divergence impact wealth and income inequality?

1.5.2.1 Theoretical grounds for house price divergence

According to the spatial equilibrium concept introduced by Rosen (1979) and extended by Roback (1982), local housing prices are determined by local wages and amenities in a way that the marginal resident of each region will receive identical utility and households will not wish to move. Consequently, heterogeneous wages and amenities can result in heterogeneous housing prices across regions (see section 1.3). This model was extended by Van Nieuwerburgh and Weill (2010) who provide a spatial equilibrium model in order to analyse the joint dynamics of the entire cross-section of house prices, construction, and wages. While keeping the dispersion of ability and housing supply regulation constant, an increase in productivity, and therewith wage, dispersion results in households moving towards high-productivity metropolitan areas. As a consequence of this influx combined with limited housing supply, local house prices in these areas increase, while the prices in low-productivity areas decrease. Hence, overall house price dispersion is enhanced. The model by Van Nieuwerburgh and Weill (2010) illustrates that an increase in income inequality is an essential part in explaining the increased house price dispersion in the U.S.. Housing supply regulations seem to have a quantitatively small impact on the inequality of house prices.

Regarding the house price distribution within cities, from the well-known Alonso-Mills-Muth model (Alonso, 1964, Mills, 1967, Muth, 1969), we expect higher house prices in the center and lower prices towards the fringe of a city. This model assumes a city with a central business district and a fixed population with a given income level. The further away a worker lives from the center, the higher the commuting costs. Since not everyone can live in the center, the prices for land and housing are more expensive the more central the area is. Hence, a household can choose between living centrally in a small and expensive home or living further towards the city fringe with larger and less expensive housing. These assumptions result in a monocentric urban structure with high house prices in the central area of a city and lower prices towards the fringe and the city's periphery. This model can explain divergent house price developments even within a city.

Deviating from this monocentric theory, other studies have documented a more complex structure of these spatial variations in housing markets caused by subcenters and local amenities (Muth, 1969, Dubin and Sung, 1987). In these alternative city centers, there is more workforce available with low commuting costs and therefore lower wages. At some point these advantages outweigh the loss of agglomeration effects that coincides with a move out of the city center. This results in a city structure with several centers and thereby a deviation from one central area with higher house prices than its surrounding areas.

Ferreira and Gyourko (2012) highlight the heterogeneity of within city housing markets by showing that timing and magnitudes of house price booms vary across neighborhoods. They show that local income growth is an important factor at the start of the boom in many neighborhoods. These results are in line with Van Nieuwerburgh and Weill (2010), who state with their spatial, dynamic equilibrium model that an increase in income inequality explains an essential part of increasing house price dispersion.

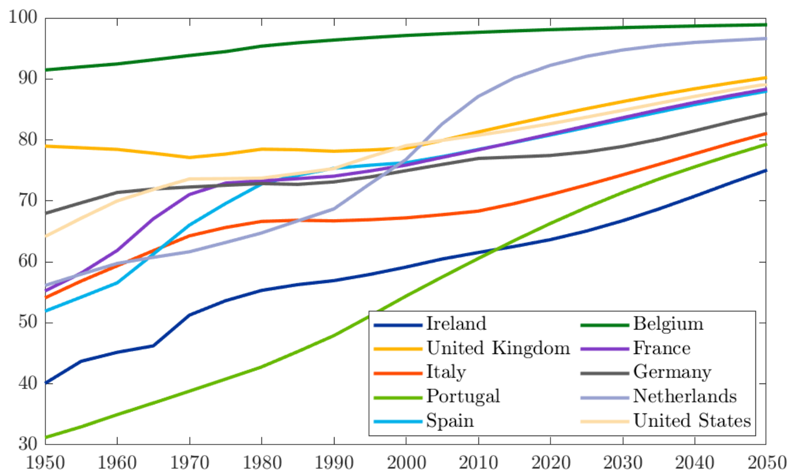
Meen (1999) explains the diffusion of house prices by means of the so-called ripple effect. That means that shocks to a local housing market can spread out to the surrounding markets, which leads to house prices moving together in the long run. The author finds that house prices rise at first in the south-east of the UK and then gradually spread out to the rest of the country. Jones and Leishman (2006) point out the importance of considering ripple effects and migration linkages between local, rather than regional, housing markets. This is in line with a gentrification process within regions described by Glaeser, Kolko and Saiz (2001) and Guerrieri, Hartley and Hurst (2013). They argue that a demand shock induces wealthier households to move to the city's fringe or periphery, which leads to an expansion of high-priced areas. As a consequence of this influx, the positive externality from living in that area, and thereby also its house prices, increase. This gentrification process in the city's fringe or periphery leads to a larger area of high house prices and therewith a more homogeneous house price distribution within the region. A former monocentric urban structure diminishes through this gentrification process.

The development away from a monocentric urban structure together with a larger area of high house prices was found for the largest German labour market regions, most pronounced for Munich and its surrounding regions in chapter 3 of this thesis. Furthermore, the article reveals overall house price divergence of German labour market regions, which has important implications for inequality and housing affordability, especially in urban areas. These connections and the associated concerns they raise are discussed in the next section.

1.5.2.2 Urbanization, house price divergence, and inequality

Urbanization processes are reshaping city size distributions and have an impact on regional house markets and housing affordability. According to the latest projections by the United Nations, urbanization processes are expected to increase until 2050 (see Figure 1.8). While in the year 2000, 75% (79%) of the German (U.S.) population lived in urban areas, the share is expected to grow to 84% (89%) in the year 2050. This expected increase is even higher for the Netherlands with an expected share of population residing in urban areas in 2050 of 97%. The growing concentration of population will increase the pressure on urban housing markets and if supply does not keep up with these dynamic changes in demand, a house price divergence process between urban and rural areas will be stimulated.

Figure 1.8: Share of urban population



Sources: United Nations, World Urbanization Prospects: The 2018 Revision.
Notes: Percentage of population at mid-year residing in urban areas.

These divergent processes of regional housing markets, give rise to concerns about income and wealth inequality as well as housing affordability, especially in urban areas. As shown by Ganong and Shoag (2017), the inequality of income per person among U.S. MSAs was 30% higher in 2016 than in 1980. An increase in regional divergence was also found for Europe (Rosés and Wolf, 2021). An increase in house prices, induced by tighter land regulation, slows population flows to high-income areas and thereby impedes the inter-regional convergence process in human capital and per capita income (Ganong and Shoag, 2017).

Furthermore, regions with a very dynamic house price development, are more exposed to the unequal impact of housing markets on the income and wealth distribution. In particular, housing markets can have an impact on employment: Due to a bust in the housing market and a related decrease in house prices, the housing net worth declines. Consequently, either through direct wealth effects or indirectly through tighter borrower constraints driven by the decline in collateral value, consumer demand decreases and, thereby, employment falls (Mian and Sufi, 2014). This impact on employment affects inequality as it is stronger for lower income households. A greater proportion of lower skilled workers may be employed in housing-related occupations or in occupations, such as the food and accommodation industry. These jobs are most sensitive to changes on housing markets and resulting changes in local demand (Choi and Green, 2017, Cairó and Cajner, 2018). In contrast, high-skilled workers are less likely to be let go due to their firm-specific skills and a more difficult replacement process (Becker, 2009, Abbritti and Consolo, 2022, Oesch, 2010). Hence, the income of low skilled workers, and thereby low income households, is likely to be more sensitive to changes in local labour market conditions arising from fluctuations in house prices.

Besides this labour market channel, real incomes of low income households are affected differently by house price fluctuations than incomes of high income households. With regard to the rising income inequality in Germany since the mid-1990s, Dustmann, Fitzenberger and Zimmermann (2018) study the effects of rising housing expenditures on households with different income levels. The authors state that an increase in housing expenditures leads, for low-income households, to a decrease in saving rates and a loss in real disposable income. For high-income households, these changes are reversed and the share spent on housing decreases. Thus, increases in house prices, lead to inequality in disposable income, as well as a divergence in consumption and saving patterns across income groups. Albouy, Ehrlich and Liu (2016) state that rising rents have a larger impact on poor households and increase real income inequality. These results are in line with Backhaus, Gebers and Schröder (2015) who analyse German rent-income ratios by means of the German Socio-Economic Panel (SOEP). The authors find that a rising square-meter price has a negative effect on low-income households. By means of a life-cycle model, Scoccianti (2010) finds that there is a redistribution of welfare from low-income to higher income households, when households, which are at the beginning of their life-cycle, experience an increase in house prices. When a shock is received later in the life-cycle, it increases welfare across all earning levels.

Focusing on wealth effects, Kuhn, Schularick and Steins (2020) examine households at the top and in the middle of the wealth distribution. The authors find that housing booms lead to substantial wealth gains for leveraged middle-class households, as their portfolios are dominated by housing, while rich households predominantly own business equity. Hence, in contrast to income inequality, housing booms tend to decrease wealth inequality, all else equal. Aladangady, Albouy and Zabek (2017) examine housing inequality in the U.S. over the last 85 years. While housing value inequality fell between 1930 and 1970 due to increasing homeownership rates, it has increased since then. Their evidence on divergent housing values indirectly supports the view that wealth inequality has increased albeit moderated through an increase in homeownership. This deduction is in line with Causa, Woloszko and Leite (2019), who study housing and wealth distributions across OECD countries. First, the authors state that wealth inequality is much higher and more dispersed across countries than income inequality. Second, they find a negative relationship between homeownership and wealth, such that countries with lower homeownership rates exhibit higher wealth inequality, and vice versa. Consistent with Kuhn, Schularick and Steins (2020) and from a static cross-country perspective, Causa, Woloszko and Leite (2019) identify housing as an equalising factor of the wealth distribution, as it is a larger source of wealth among middle class households than at the top.

1.6 Monetary policy, the housing market and the real economy

As stated in section 1.5.1, monetary policy can be a determinant of house prices and according to our empirical results in chapter 4, house price changes, induced by monetary policy shocks, can, in turn, have an impact on economic activity. This section presents existing literature on the role of the housing channel, explains the transmission mechanism of monetary policy through the housing channel, and in a third step considers the impact of monetary policy on inequality.

1.6.1 The role of the housing channel in the empirical literature

At the latest since the bursting of the housing bubble in the United States and the global financial crisis of 2008, the linkages between the housing market and economic activity have been widely debated. Nocera and Roma (2017) find that a housing demand shock, in terms of a 1% increase in house prices, impacts real private consumption and GDP albeit to a very different extent across euro area countries. In comparison to the U.S., Musso, Neri and Stracca (2011) state that the impact of a housing demand shock on consumption is stronger for the euro area. Goodhart and Hofmann (2008) apply a Bayesian SSVS-VAR model for euro area countries and also find that house price shocks have a significant impact on GDP.[11]

A second strand of the literature deals more specifically with the impact of monetary policy measures on the housing markets. For the euro area, Musso, Neri and Stracca (2011) find that a contractionary monetary policy shock has a negative impact on residential investment and house prices albeit smaller than for the U.S. (see also Rahal, 2016), which is in line with the findings by Huber and Punzi (2020). In addition, the authors state that the response of euro area housing variables to a monetary policy shock tends to be stronger after the financial crisis and, by means of a counterfactual analysis, they find that house prices would have been much lower without the central bank's quantitative easing and forward guidance during this period. Differentiating between an interest rate and a balance sheet shock, Rosenberg (2020) finds that an expansionary policy rate shock has a larger and more persistent impact on house prices than balance sheet shocks do.

Several studies analyse the impact of monetary policy measures across countries. Corsetti, Duarte and Mann (2020) show that a contractionary monetary policy shock decreases GDP and house prices and raises mortgage rates, while the size of the impact differs across countries. These results on house prices are in line with Zhu, Betzinger and Sebastian (2017) who identify country specific mortgage market conditions as key to determining how monetary policy shocks are transmitted to the respective housing market. Hülsewig and Rottmann (2021) corroborate this cross country heterogeneity in housing market responses for a recent period of unconventional monetary policy measures. By means of a Panel VAR, Nocera and Roma (2017) show that monetary policy has a strong and long lasting impact on house prices, confirming the existence of a credit channel in the euro area housing market as well as a high cross-country heterogeneity. This heterogeneity is confirmed by Koeniger, Lennartz and Ramelet (2021), who find differences in the pass-through of monetary policy shocks to mortgage rates, housing tenure transitions, rents and house prices for Germany, Italy, Switzerland as well as regions within Italy, which can be explained by the incidence of mortgagors

[11] Regarding the linkages between the housing market and economic activity, see also Iacoviello (2005), Leamer et al. (2007), Smets and Jarociński (2008) and Ghent and Owyang (2010).

and renters, institutional features, and the extent of public housing. The article in chapter 4 of this dissertation extends the analysis of housing market heterogeneity by considering sub-national housing markets of eight euro area countries. While most studies focus either on the impact of monetary policy on the housing market or on the links between housing and the business cycle, our paper connects these two strands and examines how monetary policy affects economic activity through the housing channel.

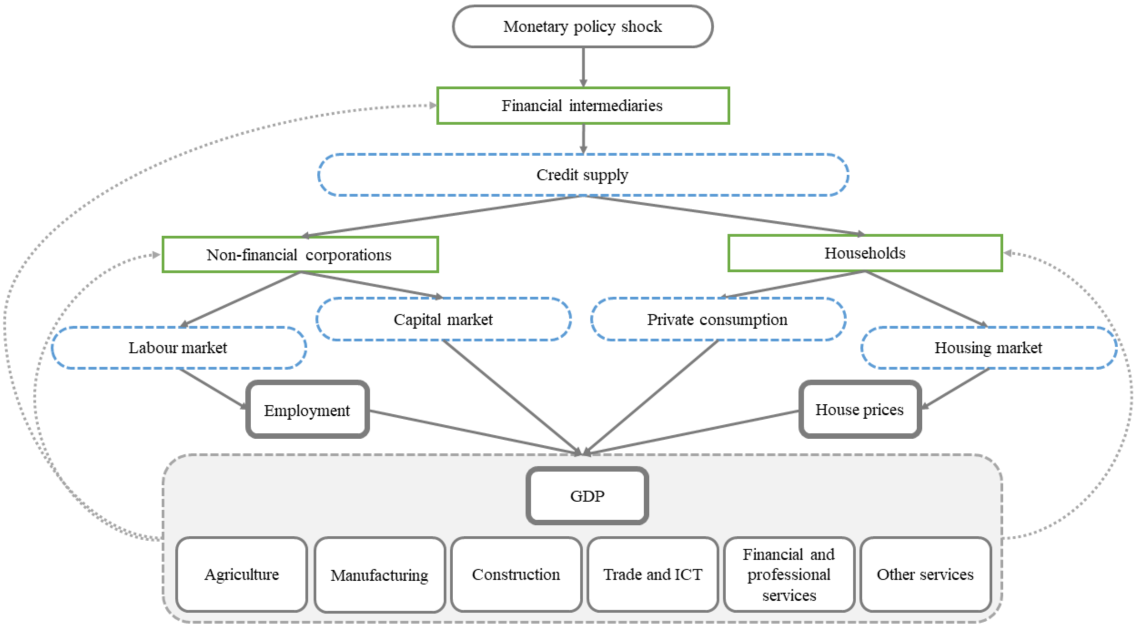
There are only a few studies examining this transmission mechanism. By means of a two-stage approach, Giuliodori (2005) assesses the monetary policy transmission through the housing sector. As a first step the total effect of interest rate shocks on private consumption is estimated and as a second step the author runs a counterfactual simulation exercise, in which the effect of interest rate shocks on consumption working through house price changes are shut off. The propagating role of house prices to consumption is strong for the UK, and, to a smaller extent, the Netherlands and Spain, while it is negligible for France and Italy. For the UK, Elbourne (2008) employs a two-stage approach estimating the response of house prices to an interest rate shock and, in a second step, estimating the response of consumption to a house price shock. When combining these two responses, the author finds that $12-15\%$ of the drop in consumption after a contractionary interest rate shock can be attributed to changes in house prices. Ozkan et al. (2017) find comparable results for the U.S. by means of a quantitative heterogeneous agents New Keynesian model. Furthermore, the authors state that expansionary monetary policy is more effective in a high-LTV economy.

1.6.2 The transmission mechanism of monetary policy

This section presents the theoretical framework to investigate the housing channel of monetary policy. As a first step, we consider the overall mechanism on how monetary policy transmits to the overall economy. In a second step, we take a closer look at the concrete housing channels at play.

Monetary policy propagates to the real economy through several direct and indirect channels. For illustrative purposes, we consider a closed economy with households, firms, financial intermediaries and a central bank (Figure 1.9). This framework is consistent with a broad class of general equilibrium models used to analyse the role of the housing market in the transmission of monetary policy, including models with collateral constraints (Iacoviello, 2005; Guerrieri and Iacoviello, 2017), non-rational expectations (Adam and Woodford, 2021) and household heterogeneity (Kaplan, Moll and Violante, 2018). Let us assume that the

Figure 1.9: Transmission mechanism of a monetary policy shock



Sources: Author's illustration.

central bank engenders an expansionary monetary policy shock, i.e. risk-free rates decline more than expected. This monetary policy accommodation improves supply conditions on the credit market, so that financial intermediaries expand their lending to the private sector. In turn, households and firms' current spending decisions are enhanced, thus stimulating aggregate demand across the consumption, housing, capital and labour markets. At the same time, monetary policy accommodation improves the economic expectations of private sector agents, which exerts upward pressures especially on financial and non-financial asset prices. As a consequence, the demand for labor, and thereby employment, increases, which generates positive income effects and supports economic activity.

Regarding the housing market, house price increases boost homeowners' wealth, lowering their savings while increasing private consumption. As house prices grow compared with construction costs, favourable Tobin's Q effects also make housing investment more attractive. To the extent that housing is posted as collateral, an increase in house prices relaxes borrowing constraints and allows homeowners to smooth consumption over the life cycle, further enhancing aggregate demand. Overall, monetary accommodation expands the resources available for the private sector, generating positive income and wealth effects for both households and firms and supporting activity.

After having given a broad overview of the transmission mechanism of a monetary policy shock to the overall economy, and given the focus of this dissertation on housing markets, we want to take a closer look at the concrete housing channels at play. Mishkin (2007) argues that monetary policy affects the housing market, and in turn the overall economy, through at least six direct and indirect channels.

The channel (1) is given by direct interest rate effects through the user cost of capital.[12] When the central bank decreases short-term interest rates, long-term interest rates will also decline, given their linkage to expected future short-term rates. This, in turn, decreases the user cost of capital and the demand for housing increases. A growing demand for housing leads to an increase in house prices as well as construction activity and thereby aggregate demand.

The channel (2) affects the user cost of capital through the expected real rate of appreciation of housing prices. As explained for channel (1), expansionary monetary policy leads to an increase in the demand for housing, and consequently, house prices. Therefore, expectations of a future monetary policy accommodation can increase the expected real rate of appreciation of housing prices. This, in turn, lowers the current user cost of capital, so that housing demand and construction activity increase.

The channel (3) refers to the supply side of housing. Expansionary monetary policy in form of lower interest rates, reduces the cost of producing new housing and thereby increases construction activity.

The housing market is indirectly affected by monetary policy through wealth effects on consumption from changes in house prices. In general, an increase in wealth leads to higher consumption, which is based on the view that the long-run marginal propensity to consume out of wealth is higher than the real interest rate (Modigliani and Brumberg, 1954, Ando and Modigliani, 1963). The channel (4) bases on a mechanism described earlier, that expansionary monetary policy enhances housing demand and leads to an increase in house prices. This results in an increase in total wealth, which stimulates household consumption

[12] The user cost of capital can be formalized as $uc = ph[((1-t)i - \pi^e) - (\pi_h^e - \pi^e) + \delta]$, where ph is the relative price of new housing capital, i is the mortgage rate, π^e is the expected inflation rate and δ is the depreciation rate of housing, so that $((1-t)i - \pi^e)$ is the expression for after-tax real interest rates, and $(\pi_h^e - \pi^e)$ denotes the expected real rate of appreciation of housing prices (Mishkin, 2007).

and thereby aggregate demand.

The channel (5) is based on balance sheet and credit-channel effects of monetary policy on consumer spending. An increase in house prices, which is induced by expansionary monetary policy, leads to more potential collateral for the homeowner. This has a positive impact on the amount and terms of credit for new residential mortgages or home equity loans available to this household. At the same time, if homeowners have additional collateral against which they can borrow, mortgage equity withdrawals can be used to withdraw cash. Both mechanisms, the easing of credit constraints as well as mortgage equity withdrawals can result in higher consumption of homeowners.

The last channel (6) regards balance sheet and credit-channel effects of monetary policy on housing demand. Lower nominal rates, even if real interest rates are unaffected, increase households' current cash-flows. As a consequence, credit-constrained households can afford larger mortgages and thereby larger amounts of housing so that the demand for housing increases. This has a positive impact on aggregate demand. This also holds true for credit-constrained homeowners with variable-rate mortgages. Lower short-term rates on variable-rate mortgages lead to a decline in interest rate payments, and consequently higher cash-flows. This results in an increase in the size of mortgage that credit-constrained homeowners will be able to afford, and an increase in housing demand.

1.6.3 Monetary policy and inequality

The focus on the comparison between the housing channel and the employment channel of monetary policy in chapter 4 is motivated by their impact on inequality, which is not unambiguous. On the one hand, a loosening of monetary policy may reduce inequality by supporting economic activity and employment, which has a disproportionately positive impact on lower income households or regions. On the other hand, it may disproportionately support households or regions with higher housing wealth and thereby increase inequality (see section 1.5.2.2). There is a controversial discussion in the growing literature on monetary policy and inequality, which mainly focuses on households and individuals. Some studies find that expansionary monetary policy can mitigate income inequality as lower income households disproportionately benefit from positive effects via the stimulus to economic activity and employment, which outweigh those via financial markets (for the U.S.: Coibion et al., 2017; for the euro area: Casiraghi et al., 2018, Altavilla et al., 2021). This stands in contrast to

Amberg et al. (2021), who state for Sweden that the income response is U-shaped and to Andersen et al. (2020), who find with data from Denmark that monetary easing raises income shares at the top of the income distribution while reducing them at the bottom, and thus, leading to higher income inequality. The impact of monetary policy on wealth inequality is discussed as controversially. Lenza and Slacalek (2021) state that monetary policy has only a negligible impact on wealth inequality, Casiraghi et al. (2018) find a U-shaped response and according to Andersen et al. (2020) monetary easing is more beneficial to the net wealth of higher income households and thereby increasing wealth inequality.

Besides these studies on household level data, little attention has been given to the pronounced geographical dimension of inequality and how it is affected by monetary policy. Hauptmeier, Holm-Hadulla and Nikalexi (2020) focus on the heterogeneity of the impact of monetary policy on euro area regions. They find that monetary easing shocks have a significantly more pronounced and persistent effect on output in poorer regions than on output in richer regions implying a mitigation of regional inequality. For the U.S., Beraja et al. (2019) examine the transmission of monetary policy via mortgage markets at the regional level. In contrast to former recessions, for the great financial crisis, they find that depressed regions reacted less to interest rate cuts, which increased regional consumption inequality.

In general, there is growing evidence, both in the theoretical (Kaplan, Moll and Violante, 2018) and empirical (Lenza and Slacalek, 2021; Hauptmeier, Holm-Hadulla and Nikalexi, 2020) literature, pointing to a larger role for labour income relative to housing wealth in transmitting monetary policy to the real economy.

1.7 An overview of the dissertation papers

In this section, an overview of the following chapters of this dissertation is provided. All of the papers are linked by their regional perspective. While the first article (chapter 2) contributes to the literature on city size distributions and their development over time, the second article (chapter 3) examines the convergence behavior of regional housing markets and its determinants. The third article (chapter 4) assesses the role of the housing market in the conventional and unconventional transmission of monetary policy across regions.

1.7.1 The evolution of Zipf's law for U.S. cities

Using a novel methodology and based on a dataset for the 100 largest U.S. cities between 1840 and 2016, the aim of this paper is to get a better understanding of the evolutionary process of the U.S. city size distribution. As a first step, modifications of the exact Zipf's law are presented, which – in contrast to the exact Zipf's law - allow for a more evenly or unevenly sized distribution as well as the absence of leading cities in the top level of the urban ranking. By estimating the most general form of these models, the three-parameter Zipf model, which can be traced back to Mandelbrot (1982), we can identify for each time period, whether the U.S. city size distribution can be described by Zipf's law or by a modified version of it. Thus, we can also infer information on the evolution of the city size distribution. To validate these results and to find evidence for primate cities in the size distribution, we make use of the finding made by Chen (2012) who shows that Zipf's law can be derived by the hierarchical scaling law based on a hierarchical urban structure. Intuitively, if the top level of a hierarchy is vacant, there is no evidence for primate cities.

With the exemption of the years 1850 to 1890, the U.S. city size distribution significantly (at a 5% level of significance) follows a two-parameter Zipf model in the years 1840-2016 even when considering larger samples. Hence, we can clearly reject the exact form of Zipf's law for U.S. city data. Furthermore, the scaling exponent fluctuates around the value one for the first 60 years (between 1840 and 1900) and decreases afterwards until it reaches a value of 0.75 in the year 2016. These results indicate that for most of the years the U.S. city size distributions are more equally distributed than expected by the exact Zipf's law and that they have become more equally distributed over time. By exploiting the dual relationship between the hierarchical scaling and Zipf's law, the estimated Zipf models can be validated and a more precise understanding of the structure of urban hierarchies can be achieved. For the 100 largest cities and for most of the time span 1840-2016 the city structure follows a hierarchical scaling law and we find evidence for leading cities dominating the remaining largest U.S. cities. The absence of leading cities for the years 1850, 1880 and 1890 can be confirmed by finding that the hierarchical scaling law fits an urban structure without top levels. Furthermore, the city size distribution diverges from the hierarchical scaling law starting with 1960 until the year 2016.

In contrast to González-Val (2010), who explains the convergence of the city sizes with a loss of importance of the largest cities, for most of the time span, we find evidence for leading cities dominating the remaining largest cities. Our results indicate that the growth of the smaller cities plays the main role in the convergence process. At the same time, Black and Henderson (2003) and Dobkins and Ioannides (2000) found that U.S. MSAs have become more unequally distributed during the twentieth century. Connecting these results to the convergence of city sizes, we found, confirms an increasing suburbanization in the growth process of the largest U.S. urban areas starting in the 1960s (Soo, 2005).[13]

1.7.2 House price convergence across German regions

The aim of this study is to examine the evidence for regional house price convergence in Germany over the years 2007 to 2017. As a first step, we employ a regression-based convergence test by Phillips and Sul (2007) in order to determine whether house prices across all 141 labour market regions converge to a common steady state. If overall convergence has to be rejected, we test whether subgroups of regions, which are determined on the basis of historical/geographical linkages as well as the seven largest cities, show a convergent house price behavior. If we cannot find house price convergence within these pre-determined subgroups, a club convergence algorithm by Phillips and Sul (2007), which endogenously forms subgroups of regions with converging house prices, will be applied. As these clubs do not necessarily coincide with commonly known classifications of German regions, as a last step, we study general characteristics of these subgroups as well as possible factors driving convergence club membership.

The log t convergence test and the clustering algorithm by Phillips and Sul (2007) have been applied to various research fields, such as the convergence of consumer prices (Phillips and Sul, 2007), GDP (Phillips and Sul, 2009, Bartkowska and Riedl, 2012, Lyncker and Thoennessen, 2017), carbon dioxide emissions (Burnett, 2016) as well as housing markets in the U.S. (Kim and Rous (2012)), the UK (Montagnoli and Nagayasu, 2015), Spain (Blanco, Martín and Vazquez, 2016), Poland (Mateusz, 2019, Matysiak and Olszewski, 2019), Australia

[13] According to Boustan and Shertzer (2013), a large portion of suburbanization in the U.S. over the twentieth century can be explained by factors associated with the natural evolution process of urbanization, like rising incomes, which led to a larger demand for housing and land, as well as transportation improvements, especially the growing network of interstate highways. Furthermore, the authors state that factors associated with the flight-from-blight theory of suburbanization, like school quality, taxes, crime-rates and socioeconomic factors of the population, reinforced the spatial dispersion. Also see Mieszkowski and Mills (1993), Bayoh, Irwin and Haab (2006) and Kim (2000).

(Awaworyi Churchill, Inekwe and Ivanovski, 2018), South Africa (Apergis, Simo-Kengne and Gupta, 2015) and European countries (Tsai, 2018). Still, to the best of our knowledge, no existing study has examined either intra- or inter-regional house price convergence for Germany. With regard to the strong and broadly based increase of house prices in recent years (Deutsche Bundesbank, 2020), we intend to fill this gap.

Rejecting the hypothesis of overall convergence, the results of the log t test propose a dispersion of house prices across the 141 German labour market regions. Given the overall divergent house price development, we perform a clustering analysis, as a first step on subgroups of regions, which are determined on the basis of historical/geographical linkages or similarities in their size, namely East and West Germany, the German states and the Top 7 cities. The hypothesis of overall house price convergence within each of these subgroups has to be rejected. In contrast to these pre-determined subgroups of regions, as a next step, the convergence clustering algorithm by Phillips and Sul (2007) with alterations by Schnurbus, Haupt and Meier (2017) is applied, which endogenously forms 7 subgroups of regions with converging house prices. By construction of the algorithm, the members of the first club, Ingolstadt and Munich, show on average the highest house price growth between 2007 and 2017 as well as the highest house price index in 2017. These variables are decreasing in size with an increasing club number. Regarding the geographical distribution, lower house price clubs mostly comprise regions in the middle, the east (except regions around Berlin) and in the south west of Germany (Rhineland Palatinate and Saarland). Convergence clubs with a higher average house price growth mostly consist of regions in southern Germany, particularly in Bavaria and Baden-Wurttemberg. Furthermore, the seven largest cities and most of their surrounding regions belong to high house price clubs indicating a high house price cluster formation which is most pronounced for Munich and its surrounding regions. As described by Glaeser, Kolko and Saiz (2001) and Guerrieri, Hartley and Hurst (2013), this cluster formation around large cities, which goes in line with a diminishing monocentric urban structure, can be explained through a gentrification process in the city's fringe or periphery.

As these convergence clubs are determined endogenously, the cluster formation does not depend on arbitrarily selected variables or thresholds, but as a consequence, its results are also somewhat atheoretical. Thus, we apply an ordered logit model in order to determine the key drivers of this club formation. The results indicate that population growth and density as well as the supply side variable, namely the housing stock, are key drivers in determining house price club membership. A one unit increase in population growth significantly increases the probability of belonging to a higher house price club, whereas an increase in population density and housing stock decreases this probability.

1.7.3 Navigating the housing channel across euro area regions

Profound economic and institutional differences across regions have long challenged the effectiveness of monetary policy in the euro area. The unequal geography of the transmission of monetary policy has also stoked concerns about its possible side effects on regional inequality, especially owing to the unconventional measures conducted by the European Central Bank (ECB) over the last decade. In this context, the housing market—in light of its role in the propagation of shocks, its distributional implications and its local dimension—has often come to the front of the media and policy debate on the intended and unintended effects of monetary policy across euro area regions.

Our paper contributes to the literature on this debate by assessing empirically the role of the housing market in the conventional and unconventional transmission of monetary policy across regions in the first two decades of the euro area. We first construct a large dataset with a panel of 106 regions in eight euro area countries (Belgium, Germany, Spain, France, Ireland, Italy, the Netherlands, and Portugal) covering the period 1999-2018. We compile novel indicators for regional house prices and loan-to-value (LTV) ratios on the basis of loan-level data. We also collect regional indicators for aggregate and sectoral activity, labour market developments and housing market features.

We then consider monetary policy through its conventional and unconventional transmission mechanisms by constructing a measure of monetary policy shocks. To isolate the impact of "genuine" monetary policy shocks, we adopt a high-frequency identification and impose sign and zero restrictions on high-frequency changes in risk-free interest rates and stock prices around the ECB's monetary policy announcements. We assume that the conventional transmission mechanism of monetary policy has mainly operated through short-term rates, whereas long-term rates were primarily related to the unconventional transmission mechanism of monetary policy enacted in the aftermath of the Global Financial Crisis.

Making use of our regional dataset and our measure of conventional and unconventional monetary policy shocks, we design a methodology to assess the role of the housing market in the transmission of monetary policy to the real economy. Using a structural panel vector autoregression (SPVAR) model with regional GDP, employment and house prices as endogenous variables, and euro area monetary policy shocks as exogenous variable, we first assess the average impact of a monetary accommodation on GDP, employment and house prices across regions. Accounting for the endogenous reaction of GDP to employment and house prices, we further quantify the role of the employment and the housing channels in conveying monetary stimulus. We finally provide an anatomy of the long-term drivers of the diverse impact of monetary policy across euro area regions.

Our results point to an effective, yet widely heterogeneous transmission of monetary policy across the euro area, with monetary policy stimulating economic activity mainly through labour income, compared with housing wealth. Nevertheless, the housing channel becomes more relevant in the unconventional transmission of monetary policy. Moreover, we find that monetary policy has a larger impact on the economy of regions with lower labour income and a higher homeownership rate. This suggests that poorer regions stand to benefit the most from expansionary monetary policy, which may decrease cross-regional inequality.

1.8 Outlook

The three articles of this dissertation, dealing with city size distributions (chapter 2), house price convergence (chapter 3) and the housing channel of monetary policy (chapter 4), deliver three important implications for the future development of regional housing markets.

The suburbanization process and diminishing urban structure we identified for the U.S. (chapter 2) and for Germany (chapter 3) can be related to a gentrification process in the cities' fringe or periphery (Glaeser, Kolko and Saiz, 2001, Guerrieri, Hartley and Hurst, 2013). It can be expected that this development will be enhanced by the COVID-19 crisis. Accompanying regulations, such as closed schools, offices, stores, and cultural facilities, added to the drawbacks of living in small and expensive homes in a central area while letting the amenities of central homes diminish. Even for the post-pandemic era, more flexible working arrangements, and consequently the acceptance of longer commuting times, may encourage more households to move to the cities' peripheries and enhance the ongoing gentrification processes for the largest cities and their surrounding regions. As a result, high house price areas around the largest cities may become larger and the price pressure in the cities' centers may decrease. This development requires investments in infrastructure, broadband, public transport, and services for the cities' peripheries, which will need to be in the focus of policymakers.

Furthermore, chapter 3 reveals that current population growth is highly correlated with future house price growth in German labour market regions. According to population forecasts by Eurostat, regions in southern Germany are mostly expected to grow until the year 2030, whereas regions in eastern Germany, with the exception of Berlin and its surrounding areas, are projected to decrease regarding their population size. These future population developments, in turn, are correlated with current house price growth. Connecting these results indicates that already high house price growth regions are expected grow further in population size and thereby develop even higher future house price increases, while the opposite holds true for regions with lower house price growth. As a consequence, the divergence of regional house prices would become more pronounced in Germany. Diverging house prices, in turn, slow down the inter-regional convergence process in human capital and per-capita income (Ganong and Shoag, 2017). Furthermore, Dustmann, Fitzenberger and Zimmermann (2018) state that an increase in housing expenditures leads, for low-income households, to a decrease in saving rates and a loss in real disposable income. For high-income households, these changes are reversed. This increase in income inequality negatively affects housing affordability (Albouy, Ehrlich and Liu, 2016, Backhaus, Gebers and Schröder, 2015, Scoccianti, 2010).

After years of a low interest rate environment, recently, there was a turning point in the stance of monetary policy in many countries. As shown in chapter 4, monetary policy accommodation over the last years had a positive impact on house prices in the euro area, albeit heterogeneous across regions. Hence, the current increase in interest rates may also have a heterogeneous and, given evidence on the impact of contractionary monetary policy, on average a negative impact on house prices (e.g. Corsetti, Duarte and Mann, 2020). Furthermore, in chapter 4, accommodative monetary policy is found to have a larger impact in areas with lower labour income and higher homeownership rates indicating that poorer regions stand to benefit the most from expansionary monetary policy. While monetary policy accommodation is found to mitigate regional inequality through its stimulus to the economy, the current increase in interest rates may have unintended consequences for regional inequality in the euro area, particularly in the case of resurgent fragmentation risks.

References

Abbritti, Mirko, and Agostino Consolo. 2022. "Labour market skills, endogenous productivity and business cycles." European Central Bank Working Papers 2651.

Abraham, Jesse M., and Patric H. Hendershott. 1996. "Bubbles in Metropolitan Housing Markets." *Journal of Housing Research*, 7(2): 191–207.

Adam, Klaus, and Michael Woodford. 2021. "Robustly optimal monetary policy in a new Keynesian model with housing." *Journal of Economic Theory*, 198: 105352.

Adams, Zeno, and Roland Füss. 2010. "Macroeconomic determinants of international housing markets." *Journal of Housing Economics*, 19(1): 38–50.

Ahearne, Alan G, John Ammer, Brian M Doyle, Linda S Kole, and Robert F Martin. 2005. "House prices and monetary policy: a cross-country study." Citeseer.

Akinci, Ozge, and Jane Olmstead-Rumsey. 2018. "How effective are macroprudential policies? An empirical investigation." *Journal of Financial Intermediation*, 33: 33–57.

Aladangady, Aditya, David Albouy, and Mike Zabek. 2017. "Housing inequality." NBER Working Paper 21916.

Alam, Zohair, Adrian Alter, Jesse Eiseman, Gaston Gelos, Heedon Kang, Machiko Narita, Erlend Nier, and Naixi Wang. 2019. "Digging Deeper–Evidence on the Effects of Macroprudential Policies from a New Database." *IMF Working Papers*, 2019(066).

Albouy, David, and Gabriel Ehrlich. 2018. "Housing productivity and the social cost of land-use restrictions." *Journal of Urban Economics*, 107(C): 101–120.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu. 2016. "Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis." National Bureau of Economic Research, Inc NBER Working Papers 22816.

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

Alonso, William. 1964. *Location and land use: Toward a general theory of land rent.* Cambridge/Mass.: Harvard Univ. Press.

Altavilla, Carlo, Wolfgang Lemke, Tobias Linzert, Jens Tapking, and Julian von Landesberger. 2021. "Assessing the efficacy, efficiency and potential side effects of the ECB's monetary policy instruments since 2014." European Central Bank Occasional Paper Series 278.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. 2021. "Five Facts about the Distributional Income Effects of Monetary Policy." CESifo Working Paper Series 9062.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró. 2020. "Monetary policy and inequality." Department of Economics and Business, Universitat Pompeu Fabra Economics Working Papers 1761.

Ando, Albert, and Franco Modigliani. 1963. "The" life cycle" hypothesis of saving: Aggregate implications and tests." *The American economic review*, 53(1): 55–84.

Apergis, Nicholas, Beatrice D. Simo-Kengne, and Rangan Gupta. 2015. "Convergence In Provincial-Level South African House Prices: Evidence From The Club Convergence And Clustering Procedure." *Review of Urban & Regional Development Studies*, 27(1): 2–17.

Arrow, Kenneth J. 1962. "Economic Welfare and the Allocation of Resources for Invention." *The Rate and Direction of Inventive Activity: Economic and Social Factors.*

Arshad, Sidra, Shougeng Hu, and Badar Nadeem Ashraf. 2018. "Zipf's law and city size distribution: A survey of the literature and future research agenda." *Physica A: Statistical mechanics and its applications*, 492: 75–92.

Awaworyi Churchill, Sefa, John Inekwe, and Kris Ivanovski. 2018. "House price convergence: Evidence from Australian cities." *Economics Letters*, 170(C): 88–90.

Backhaus, Teresa, Kathrin Gebers, and Carsten Schröder. 2015. "Evolution and Determinants of Rent Burdens in Germany." DIW Berlin, The German Socio-Economic Panel SOEP papers on Multidisciplinary Panel Data Research 806.

Bartkowska, Monika, and Aleksandra Riedl. 2012. "Regional convergence clubs in Europe: Identification and conditioning factors." *Economic Modelling*, 29(1): 22–31.

Bauer, Thomas, Sven Feuerschütte, Michael Kiefer, Philipp an de Meulen, Martin Micheli, Torsten Schmidt, and Lars-Holger Wilke. 2013. "Ein hedonischer Immobilienpreisindex auf Basis von Internetdaten: 2007–2011." *AStA Wirtschafts- und Sozialstatistisches Archiv*, 7(1): 5–30.

Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and city size." *Review of Economics and Statistics*, 95(5): 1535–1548.

Bayoh, Isaac, Elena G Irwin, and Timothy Haab. 2006. "Determinants of residential location choice: How important are local public goods in attracting homeowners to central city locations?" *Journal of Regional Science*, 46(1): 97–120.

Becker, Gary S. 2009. *Human capital: A theoretical and empirical analysis, with special reference to education.* University of Chicago press.

Beeson, Patricia E, David N DeJong, and Werner Troesken. 2001. "Population growth in US counties, 1840–1990." *Regional Science and Urban Economics*, 31(6): 669–699.

Behrens, Kristian, and Frédéric Robert-Nicoud. 2015. "Agglomeration theory with heterogeneous agents." *Handbook of regional and urban economics*, 5: 171–245.

Belenzon, Sharon, and Mark Schankerman. 2013. "Spreading the word: Geography, policy, and knowledge spillovers." *Review of Economics and Statistics*, 95(3): 884–903.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *The Quarterly Journal of Economics*, 134(1): 109–183.

Berry, Brian, and Adam Okulicz-Kozaryn. 2012. "The city size distribution debate: Resolution for US urban regions and megalopolitan areas." *Cities*, 29: 17–23.

Black, Duncan, and Vernon Henderson. 1999. "A theory of urban growth." *Journal of political economy*, 107(2): 252–284.

Black, Duncan, and Vernon Henderson. 2003. "Urban evolution in the USA." *Journal of economic geography*, 3(4): 343–372.

Blanco, Francisco, Victor Martín, and Guillermo Vazquez. 2016. "Regional house price convergence in Spain during the housing boom." *Urban Studies*, 53(4): 775–798.

Boustan, Leah, and Allison Shertzer. 2013. "Population trends as a counterweight to central city decline, 1950–2000." *Demography*, 50(1): 125–147.

Budde, Rüdiger, and Uwe Neumann. 2019. "The size ranking of cities in Germany: caught by a MAUP?" *GeoJournal*, 84(6): 1447–1464.

Burnett, J. Wesley. 2016. "Club convergence and clustering of U.S. energy-related CO2 emissions." *Resource and Energy Economics*, 46(C): 62–84.

Bétin, Manuel, and Volker Ziemann. 2019. "How responsive are housing markets in the OECD? Regional level estimates." OECD Publishing OECD Economics Department Working Papers 1590.

Cairó, Isabel, and Tomaz Cajner. 2018. "Human capital and unemployment dynamics: Why more educated workers enjoy greater employment stability." *The Economic Journal*, 128(609): 652–682.

Caldera, Aida, and Asa Johansson. 2013. "The price responsiveness of housing supply in OECD countries." *Journal of Housing Economics*, 22(3): 231–249.

Cameron, Gavin, John Muellbauer, and Anthony Murphy. 2006. "Was There A British House Price Bubble? Evidence from a Regional Panel." University of Oxford Economics Series Working Papers 276.

Carlino, Gerald A., Jake Carr, Robert M. Hunt, and Tony E. Smith. 2010. "The agglomeration of R&D labs." Federal Reserve Bank of Philadelphia Working Papers 10-33.

Casiraghi, Marco, Eugenio Gaiotti, Lisa Rodano, and Alessandro Secchi. 2018. "A "reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households." *Journal of International Money and Finance*, 85(C): 215–235.

Causa, Orsetta, Nicolas Woloszko, and David Leite. 2019. "Housing, wealth accumulation and wealth distribution: Evidence and stylized facts." OECD Publishing OECD Economics Department Working Papers 1588.

Cavalleri, Maria Chiara, Boris Cournède, and Ezgi Özsöğüt. 2019. "How responsive are housing markets in the OECD? National level estimates." OECD Publishing OECD Economics Department Working Papers 1589.

Cerutti, Eugenio, Jihad Dagher, and Giovanni Dell'ariccia. 2017. "Housing finance and real-estate booms: A cross-country perspective." *Journal of Housing Economics*, 38(C): 1–13.

Cerutti, Eugenio, Stijn Claessens, and Luc Laeven. 2017. "The use and effectiveness of macroprudential policies: New evidence." *Journal of Financial Stability*, 28: 203–224.

Chen, Yanguang. 2012. "The mathematical relationship between Zipf's law and the hierarchical scaling law." *Physica A: Statistical Mechanics and its Applications*, 391(11): 3285–3299.

Chen, Yong, and Stuart S Rosenthal. 2008. "Local amenities and life-cycle migration: Do people move for jobs or fun?" *Journal of Urban Economics*, 64(3): 519–537.

Choi, Hyun J., and R. K. Green. 2017. "House prices shock and changes in inequality." AEA Conference Paper.

Claessens, Stijn, M Ayhan Kose, and Marco E Terrones. 2010. "The global financial crisis: How similar? How different? How costly?" *Journal of Asian Economics*, 21(3): 247–264.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary policy and inequality." *Journal of Monetary Economics*, 88(C): 70–89.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2012. "The Costs of Agglomeration: Land Prices in French Cities." C.E.P.R. Discussion Papers CEPR Discussion Papers 9240.

Córdoba, Juan Carlos. 2008*a*. "A generalized Gibrat's law." *International Economic Review*, 49(4): 1463–1468.

Córdoba, Juan-Carlos. 2008*b*. "On the distribution of city sizes." *Journal of urban Economics*, 63(1): 177–197.

Corsetti, Giancarlo, Joao B. Duarte, and Samuel Mann. 2020. "One Money, Many Markets: Monetary Transmission and Housing Financing in the Euro Area." International Monetary Fund IMF Working Papers 2020/108.

Crowe, Christopher, Giovanni Dell'Ariccia, Deniz Igan, and Pau Rabanal. 2013. "How to deal with real estate booms: Lessons from country experiences." *Journal of Financial Stability*, 9(3): 300–319.

Dahl, Juliana, and Malgorzata Goralczyk. 2017. "Recent Supply and Demand Developments in the German Housing Market." Directorate General Economic and Financial Affairs (DG ECFIN), European Commission European Economy - Economic Briefs 025.

Davis, Donald R, and David E Weinstein. 2002. "Bones, bombs, and break points: the geography of economic activity." *American economic review*, 92(5): 1269–1289.

D'Costa, Sabine, and Henry G Overman. 2014. "The urban wage growth premium: Sorting or learning?" *Regional Science and Urban Economics*, 48: 168–179.

De la Roca, Jorge, and Diego Puga. 2012. "Learning by working in big cities." C.E.P.R. Discussion Papers CEPR Discussion Papers 9243.

Desmet, Klaus, and Jordan Rappaport. 2017. "The settlement of the United States, 1800–2000: The long transition towards Gibrat's law." *Journal of Urban Economics*, 98: 50–68.

Deutsche Bundesbank. 2020. "Monthly Report - October 2020." Deutsche Bundesbank Institutional Reports 72/10.

Dobkins, L.H., and Y. M. Ioannides. 2000. "Dynamic evolution of the US city size distribution." In *The economics of cities.* , ed. J. M. Huriot and J. F. Thisse, 217–260. Cambridge University Press.

Dobkins, Linda Harris, and Yannis M Ioannides. 2001. "Spatial interactions among US cities: 1900–1990." *Regional science and urban Economics*, 31(6): 701–731.

Dubin, Robin A, and Chein-Hsing Sung. 1987. "Spatial variation in the price of housing: rent gradients in non-monocentric cities." *Urban Studies*, 24(3): 193–204.

Duranton, Gilles. 2007. "Urban evolutions: The fast, the slow, and the still." *American Economic Review*, 97(1): 197–221.

Duranton, Gilles, and Diego Puga. 2014. "The growth of cities." *Handbook of economic growth*, 2: 781–853.

Dustmann, Christian, Bernd Fitzenberger, and Markus Zimmermann. 2018. "Housing Expenditures and Income Inequality." Institute of Labor Economics IZA Discussion Papers 11953.

Eeckhout, Jan. 2004. "Gibrat's law for (all) cities." *American Economic Review*, 94(5): 1429–1451.

Égert, Balázs, and Dubravko Mihaljek. 2007. "Determinants of house prices in central and eastern Europe." *Comparative economic studies*, 49(3): 367–388.

Elbourne, Adam. 2008. "The UK housing market and the monetary policy transmission mechanism: An SVAR approach." *Journal of Housing Economics*, 17(1): 65–87.

Ellison, Glenn, and Edward L Glaeser. 1999. "The geographic concentration of industry: does natural advantage explain agglomeration?" *American Economic Review*, 89(2): 311–316.

European Commission. 2021. "Euro Area Housing Markets: Trends, Challenges and Policy Responses." European Commission Discussion Paper 147.

European Systemic Risk Board. 2019. "Warning of the European Systemic Risk Board." ESRB Report.

European Systemic Risk Board. 2022. "Vulnerabilities in the residential real estate sectors of the EEA countries." ESRB Report.

Feldman, Maryann P. 1994. *The geography of innovation.* Vol. 2, Springer Science & Business Media.

Feldman, Maryann P, and David B Audretsch. 1999. "Innovation in cities:: Science-based diversity, specialization and localized competition." *European economic review*, 43(2): 409–429.

Ferreira, Fernando, and Joseph Gyourko. 2012. "Heterogeneity in neighborhood-level price growth in the united states, 1993-2009." *American Economic Review*, 102(3): 134–40.

Findeisen, Sebastian, and Jens Südekum. 2008. "Industry churning and the evolution of cities: Evidence for Germany." *Journal of Urban Economics*, 64(2): 326–339.

Friedman, Milton. 1975. *There's No Such Thing as a Free Lunch.* Open Court.

Gabaix, Xavier. 1999. "Zipf's law for cities: an explanation." *The Quarterly journal of economics*, 114(3): 739–767.

Gan, Li, Dong Li, and Shunfeng Song. 2006. "Is the Zipf law spurious in explaining city-size distributions?" *Economics Letters*, 92(2): 256–262.

Ganong, Peter, and Daniel Shoag. 2017. "Why has regional income convergence in the U.S. declined?" *Journal of Urban Economics*, 102(C): 76–90.

Ghent, Andra C., and Michael T. Owyang. 2010. "Is housing the business cycle? Evidence from US cities." *Journal of Urban Economics*, 67(3): 336–351.

Gibrat, Robert. 1931. "Les inégalités économiques." *Libraire du Recueil Siray.*

Giesen, Kristian, and Jens Südekum. 2011. "Zipf's law for cities in the regions and the country." *Journal of economic geography*, 11(4): 667–686.

Giuliodori, Massimo. 2005. "The Role Of House Prices In The Monetary Transmission Mechanism Across European Countries." *Scottish Journal of Political Economy*, 52(4): 519–543.

Glaeser, Edward L, Albert Saiz, Gary Burtless, and William C Strange. 2004. "The rise of the skilled city." *Brookings-Wharton papers on urban affairs*, 47–105.

Glaeser, Edward L, and Bryce A Ward. 2009. "The causes and consequences of land use regulation: Evidence from Greater Boston." *Journal of urban Economics*, 65(3): 265–278.

Glaeser, Edward L, and David C Mare. 2001. "Cities and skills." *Journal of labor economics*, 19(2): 316–342.

Glaeser, Edward L, and Joshua D Gottlieb. 2009. "The wealth of cities: Agglomeration economies and spatial equilibrium in the United States." *Journal of economic literature*, 47(4): 983–1028.

Glaeser, Edward L, Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of economic geography*, 1(1): 27–50.

Glaeser, Edward L, Joseph Gyourko, and Raven E Saks. 2005*a*. "Why have housing prices gone up?" *American Economic Review*, 95(2): 329–333.

Glaeser, Edward L, Joseph Gyourko, and Raven Saks. 2005*b*. "Why is Manhattan so expensive? Regulation and the rise in housing prices." *The Journal of Law and Economics*, 48(2): 331–369.

González-Val, Rafael. 2010. "The evolution of US city size distribution from a long-term perspective (1900–2000)." *Journal of Regional Science*, 50(5): 952–972.

González-Val, Rafael. 2012. "A nonparametric estimation of the local Zipf exponent for all US cities." *Environment and Planning B: Planning and Design*, 39(6): 1119–1130.

González-Val, Rafael, Luis Lanaspa, and Fernando Sanz-Gracia. 2014. "New evidence on Gibrat's law for cities." *Urban Studies*, 51(1): 93–115.

Goodhart, Charles, and Boris Hofmann. 2008. "House Prices, Money, Credit and the Macroeconomy." European Central Bank Working Paper Series 888.

Gros, Daniel. 2007. "Bubbles in real estate? A Longer-Term Comparative Analysis of Housing Prices in Europe and the US." CEPS Working Document No. 276, October 2007. Update of No. 239/February 2006.

Gross, Marco, and Javier Población. 2017. "Assessing the efficacy of borrower-based macroprudential policy using an integrated micro-macro model for European households." *Economic Modelling*, 61: 510–528.

Grover, Arti Goswami, Somik V Lall, and Jonathan David Timmis. 2021. "Agglomeration Economies in Developing Countries: A Meta-Analysis." The World Bank.

Growe, Anna. 2010. "Human capital in the German urban system–patterns of concentration and specialisation." *European Journal of Spatial Development*, 8(3): 1–23.

Guerrieri, Luca, and Matteo Iacoviello. 2017. "Collateral constraints and macroeconomic asymmetries." *Journal of Monetary Economics*, 90(C): 28–49.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous gentrification and housing price dynamics." *Journal of Public Economics*, 100: 45–60.

Hauptmeier, Sebastian, Fédéric Holm-Hadulla, and Katerina Nikalexi. 2020. "Monetary policy and regional inequality." European Central Bank Working Paper Series 2385.

Henderson, J Vernon. 1974. "Optimum city size: The external diseconomy question." *Journal of political Economy*, 82(2, Part 1): 373–388.

Hertrich, Markus. 2020. "A novel housing price misalignment indicator for Germany." *German Economic Review*, 20(4): 759–794.

Heuermann, Daniel, Benedikt Halfdanarson, and Jens Suedekum. 2010. "Human capital externalities and the urban wage premium: Two literatures and their interrelations." *Urban Studies*, 47(4): 749–767.

Hilber, Christian AL, and Wouter Vermeulen. 2016. "The impact of supply constraints on house prices in England." *The Economic Journal*, 126(591): 358–405.

Hirsch, Boris, Elke J Jahn, Alan Manning, and Michael Oberfichtner. 2022. "The urban wage premium in imperfect labor markets." *Journal of Human Resources*, 57(S): S111–S136.

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. 2010. "A spatio-temporal model of house prices in the USA." *Journal of Econometrics*, 158(1): 160–173.

Hsu, Wen-Tai. 2012. "Central place theory and city size distribution." *The Economic Journal*, 122(563): 903–932.

Huber, Florian, and Maria Teresa Punzi. 2020. "International Housing Markets, Unconventional Monetary Policy, And The Zero Lower Bound." *Macroeconomic Dynamics*, 24(4): 774–806.

Hülsewig, Oliver, and Horst Rottmann. 2021. "Euro area house prices and unconventional monetary policy surprises." *Economics Letters*, 205(C).

Iacoviello, Matteo. 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review*, 95(3): 739–764.

Ioannides, Yannis, and Spyros Skouras. 2013. "US city size distribution: Robustly Pareto, but only in the tail." *Journal of Urban Economics*, 73(1): 18–29.

Ioannides, Yannis M, and Henry G Overman. 2003. "Zipf's law for cities: an empirical examination." *Regional science and urban economics*, 33(2): 127–137.

Jacobs, Jane. 1969. *The Economy of cities.* Vintage Books, New York.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *the Quarterly journal of Economics*, 108(3): 577–598.

Jiang, Bin, and Tao Jia. 2011. "Zipf's law for all the natural cities in the United States: a geospatial perspective." *International Journal of Geographical Information Science*, 25(8): 1269–1281.

Jones, Colin. 2012. "Introduction: the housing economy and the credit crunch." *Challenges of the housing economy: An international perspective*, 1–24.

Jones, Colin, and Chris Leishman. 2006. "Spatial dynamics of the housing market: An interurban perspective." *Urban Studies*, 43(7): 1041–1059.

Jud, G. Donald, and Dan T. Winkler. 2002. "The Dynamics of Metropolitan Housing Prices." *Journal of Real Estate Research*, 23(1/2): 29–46.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Kholodilin, Konstantin A, Jan-Oliver Menz, and Boriss Siliverstovs. 2010. "What drives housing prices down? Evidence from an international panel." *Jahrbücher für Nationalökonomie und Statistik*, 230(1): 59–76.

Kim, Sukkoo. 2000. "Urban Development in the United States, 1690-1990." *Southern Economic Journal*, 66(4): 855–880.

Kim, Young Se, and Jeffrey J. Rous. 2012. "House price convergence: Evidence from US state and metropolitan area panels." *Journal of Housing Economics*, 21(2): 169–186.

Koch, Leonie, and Martin Simmler. 2020. "How important are local knowledge spillovers of public R&D and what drives them?" *Research Policy*, 49(7).

Koeniger, Winfried, Benedikt Lennartz, and Marc-Antoine Ramelet. 2021. "On the transmission of monetary policy to the housing market." Swiss National Bank Working Papers 2021-06.

Krugman, Paul. 1991. *Geography and trade.* MIT press.

Krugman, Paul. 1996. "Confronting the mystery of urban hierarchy." *Journal of the Japanese and International economies*, 10(4): 399–418.

Krugman, Paul. 2011. "The new economic geography, now middle-aged." *Regional studies*, 45(1): 1–7.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins. 2020. "Income and wealth inequality in America, 1949–2016." *Journal of Political Economy*, 128(9): 3469–3519.

Kuttner, Kenneth N, and Ilhyock Shim. 2016. "Can non-interest rate policies stabilize housing markets? Evidence from a panel of 57 economies." *Journal of Financial Stability*, 26: 31–44.

Leamer, Edward E, et al. 2007. "Housing is the business cycle." 149–233, Federal Reserve Bank of Kansas City.

Lenza, Michele, and Jirka Slacalek. 2021. "How Does Monetary Policy Affect Income and Wealth Inequality? Evidence from Quantitative Easing in the Euro Area." C.E.P.R. Discussion Papers 16079.

Levy, Moshe. 2009. "Gibrat's law for (all) cities: Comment." *American Economic Review*, 99(4): 1672–75.

Lyncker, Konrad, and Rasmus Thoennessen. 2017. "Regional club convergence in the EU: evidence from a panel data analysis." *Empirical Economics*, 52(2): 525–553.

Malevergne, Yannick, Vladilen Pisarenko, and Didier Sornette. 2011. "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities." *Physical Review E*, 83(3): 036111.

Mandelbrot, Benoit B. 1982. *The fractal nature of geometry.* W. H. Freeman and Company, New York.

Mansury, Yuri, and Laszlo Gulyas. 2007. "The emergence of Zipf's Law in a system of cities: An agent-based simulation approach." *Journal of Economic Dynamics and Control*, 31(7): 2438–2460.

Marshall, Alfred. 1890. *Principles of Economics.* London, Macmillan and Co., Ltd.

Mateusz, Tomal. 2019. "House Price Convergence on the Primary and Secondary Markets: Evidence from Polish Provincial Capitals." *Real Estate Management and Valuation*, 27(4): 62–73.

Matysiak, George A., and Krzysztof Olszewski. 2019. "A panel analysis of Polish regional cities: residential price convergence in the primary market." Narodowy Bank Polski NBP Working Papers 316.

Meen, Geoffrey. 1999. "Regional house prices and the ripple effect: a new interpretation." *Housing studies*, 14(6): 733–753.

Mian, Atif, and Amir Sufi. 2014. "What explains the 2007–2009 drop in employment?" *Econometrica*, 82(6): 2197–2223.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding. 2012. "Urbanization and structural transformation." *The Quarterly Journal of Economics*, 127(2): 535–586.

Mieszkowski, Peter, and Edwin S Mills. 1993. "The causes of metropolitan suburbanization." *Journal of Economic perspectives*, 7(3): 135–147.

Miles, David. 2012. "Population Density, House Prices and Mortgage Design." *Scottish Journal of Political Economy*, 59(5): 444–466.

Mills, Edwin S. 1967. "An aggregative model of resource allocation in a metropolitan area." *The American Economic Review*, 57(2): 197–210.

Mishkin, Frederic S. 2007. "Housing and the Monetary Transmission Mechanism." National Bureau of Economic Research Working Papers 13518.

Modigliani, Franco, and Richard Brumberg. 1954. "Utility analysis and the consumption function: An interpretation of cross-section data." In *Post-Keynesian Economics.* , ed. K. Kurihara, 388–436. Rutgers University Press, New Brunswick.

Montagnoli, Alberto, and Jun Nagayasu. 2015. "UK house price convergence clubs and spillovers." *Journal of Housing Economics*, 30(C): 50–58.

Moretti, Enrico. 2010. "Local labor markets." In *Handbook of labor economics.* Vol. 4A of *Handbook of labor economics*, , ed. Orley Ashenfelter and David Card, Chapter 14, 1237–1313. Elsevier.

Muellbauer, John. 2005. "Property taxation and the economy after the Barker Review." *The Economic Journal*, 115(502): C99–C117.

Musso, Alberto, Stefano Neri, and Livio Stracca. 2011. "Housing, consumption and monetary policy: How different are the US and the euro area?" *Journal of Banking & Finance*, 35(11): 3019–3041.

Muth, Richard F. 1969. *Cities and housing. The spatial pattern of urban residential land use.* The University of Chicago Press, Chicago and London.

Nocera, Andrea, and Moreno Roma. 2017. "House prices and monetary policy in the euro area: evidence from structural VARs." European Central Bank Working Paper Series 2073.

OECD. 2020. "Housing Amid Covid-19: Policy Responses and Challenges." OECD Institutional Reports.

OECD. 2022. "Housing prices (indicator). Dataset. doi: 10.1787/63008438-en."

Oesch, Daniel. 2010. "What explains high unemployment among low-skilled workers? Evidence from 21 OECD countries." *European Journal of Industrial Relations*, 16(1): 39–55.

O'Meara, Graeme. 2015. "Housing bubbles and monetary policy: a reassessment." *The Economic and Social Review*, 46(4, Winter): 521–565.

Ozkan, Serdar, Kurt Mitman, Fatih Karahan, and Aaron Hedlund. 2017. "Monetary Policy, Heterogeneity, and the Housing Channel." Society for Economic Dynamics 2017 Meeting Papers 1610.

Paciorek, Andrew. 2013. "Supply constraints and housing market dynamics." *Journal of Urban Economics*, 77: 11–26.

Phillips, Peter C. B., and Donggyu Sul. 2007. "Transition Modeling and Econometric Convergence Tests." *Econometrica*, 75(6): 1771–1855.

Phillips, Peter C. B., and Donggyu Sul. 2009. "Economic transition and growth." *Journal of Applied Econometrics*, 24(7): 1153–1185.

Poghosyan, Tigran. 2020. "How effective is macroprudential policy? Evidence from lending restriction measures in EU countries." *Journal of Housing Economics*, 49(C).

Poghosyan, Tigran, and Ruud de Mooij. 2016. "Can Property Taxes Reduce House Price Volatility? Evidence from US Regions." *IMF Working Papers*, 2016(216).

Porter, Michael E. 1998. *Clusters and the new economics of competition.* Vol. 76, Harvard Business Review Boston.

Puga, Diego. 2010. "The magnitude and causes of agglomeration economies." *Journal of regional science*, 50(1): 203–219.

Rahal, Charles. 2016. "Housing markets and unconventional monetary policy." *Journal of Housing Economics*, 32(C): 67–80.

Richter, Björn, Moritz Schularick, and Ilhyock Shim. 2019. "The costs of macroprudential policy." *Journal of International Economics*, 118: 263–282.

Roback, Jennifer. 1982. "Wages, rents, and the quality of life." *Journal of political Economy*, 90(6): 1257–1278.

Roback, Jennifer Anne. 1980. "The value of local urban amenities: theory and measurement." PhD diss. University of Rochester.

Romer, Paul M. 1986. "Increasing returns and long-run growth." *Journal of political economy*, 94(5): 1002–1037.

Rosenberg, Signe. 2020. "Conventional and unconventional monetary policies: effects on the Finnish housing market." *Baltic Journal of Economics*, 20(2): 170–186.

Rosen, Sherwin. 1979. "Wage-based indexes of urban quality of life." *Current issues in urban economics*, 74–104.

Rosenthal, Stuart S, and William C Strange. 2004. "Evidence on the nature and sources of agglomeration economies." In *Handbook of regional and urban economics*. Vol. 4, 2119–2171. Elsevier.

Rosés, Joan R, and Nikolaus Wolf. 2021. "Regional growth and inequality in the long-run: Europe, 1900–2015." *Oxford Review of Economic Policy*, 37(1): 17–48.

Rossi-Hansberg, Esteban, and Mark Wright. 2007. "Urban Structure and Growth." *Review of Economic Studies*, 74(2): 597–624.

Rozenfeld, Hernán D, Diego Rybski, Xavier Gabaix, and Hernán A Makse. 2011. "The area and population of cities: New insights from a different perspective on cities." *American Economic Review*, 101(5): 2205–25.

RWI, and ImmobilienScout24. 2022. "Regional Real Estate Price Index for Germany - 2008-06/2022. RWI-GEO-REDX. Version: 1. RWI – Leibniz Institute for Economic Research. Dataset. doi.org/10.7807/immo:redx:v9."

Rybnicek, Robert, and Roland Königsgruber. 2019. "What makes industry–university collaboration succeed? A systematic review of the literature." *Journal of business economics*, 89(2): 221–250.

Saiz, Albert. 2010. "The geographic determinants of housing supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296.

Saks, Raven E. 2008. "Job creation and housing construction: Constraints on metropolitan area employment growth." *Journal of Urban Economics*, 64(1): 178–195.

Schmidheiny, Kurt, and Jens Suedekum. 2015. "The pan-European population distribution across consistently defined functional urban areas." *Economics Letters*, 133: 10–13.

Schnurbus, Joachim, Harry Haupt, and Verena Meier. 2017. "Economic Transition and Growth: A Replication." *Journal of Applied Econometrics*, 32(5): 1039–1042.

Scoccianti, Filippo. 2010. "Causes and Welfare Consequences of Real Estate Price Appreciation." In *Housing Markets in Europe*. 241–259. Springer.

Sivitanides, Petros Stavrou. 2018. "Macroeconomic drivers of London house prices." *Journal of Property Investment & Finance*, 36(6): 539–551.

Smets, Frank, and Marek Jarociński. 2008. "House Prices and the stance of Monetary Policy." European Central Bank Working Paper Series 891.

Sommer, Kamila, and Paul Sullivan. 2018. "Implications of US tax policy for house prices, rents, and homeownership." *American Economic Review*, 108(2): 241–74.

Soo, Kwok Tong. 2005. "Zipf's Law for cities: a cross-country investigation." *Regional science and urban Economics*, 35(3): 239–263.

Soo, Kwok Tong, et al. 2012. "The size and growth of state populations in the United States." *Economics Bulletin*, 32(2): 1238–1249.

Terrones, Marco. 2004. "The Global House Price Boom." International Monetary Fund World Economic Outlook.

The Financial Times. 2021. "'It has never been like this': US house price spiral worries policymakers." 4 August 2021.

Thissen, Mark, Martijn J Burger, and Frank G van Oort. 2010. "House Prices, Bubbles and City Size."

Thompson, Peter. 2006. "Patent citations and the geography of knowledge spillovers: Evidence from inventor-and examiner-added citations." *The Review of Economics and Statistics*, 88(2): 383–388.

Tsai, I-Chun. 2018. "House price convergence in euro zone and non-euro zone countries." *Economic Systems*, 42(2): 269–281.

Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill. 2010. "Why has house price dispersion gone up?" *The Review of Economic Studies*, 77(4): 1567–1606.

Westerheide, Peter. 2012. "Market stability, housing finance and homeownership in Germany." *Challenges of the Housing Economy: An International Perspective*, 153–169.

Xu, Lu, and Bo Tang. 2014. "On the determinants of UK house prices." *International Journal of Economics and Research*, 5(2): 57–64.

Zhu, Bing, Michael Betzinger, and Steffen Sebastian. 2017. "Housing market stability, mortgage market structure, and monetary policy: Evidence from the euro area." *Journal of Housing Economics*, 37(C): 1–21.

Zipf, George Kingsley. 1949. *Human Behaviour and the Principle of Least Effort.* Addison-Wesley, Cambridge MA.

CHAPTER

TWO

THE EVOLUTION OF ZIPF 'S LAW FOR U.S. CITIES

Authors: Angelina Hackmann[1], Torben Klarl[2][3]

Abstract

Exploiting the hierarchical structure of cities and based on a dataset for U.S. cities between 1840 and 2016, the aim of this paper is to analyze the evolution of the U.S. city size distribution. For that purpose we estimate a general three-parameter Zipf model, which can be traced back to Mandelbrot (1982), and validate our results by means of the hierarchical scaling law. Especially in the second half of the twentieth century, we find a pronounced departure from the exact Zipf's law. The city size distribution has become more equally distributed over time. Besides, the applied estimation method reveals evidence for leading cities dominating the remaining largest cities. Thus, the growing equality of the city sizes can be explained rather by growing smaller cities than by a loss of importance of the largest ones.

JEL Classification: R11, R12, R15

Keywords: City size distributions, Zipf's law, hierachical scaling law, urban systems

[1]University of Bremen. E-mail: angelina.hackmann@uni-bremen.de.
[2]University of Bremen. E-mail: tklarl@uni-bremen.de.
[3]Research Fellow at SPEA, Indiana University, Bloomington. E-mail: tklarl@indiana.edu.

2.1 Introduction

Gabaix (1999) proposed that "[...] city size processes must have the time to converge to Zipf's law". Accordingly, city size processes can be described as an evolutionary process where different states of urbanization require different forms of city size distributions. Stated in other words: Even if the city size process converges to the well-known exact Zipf's law[4], this law must not necessarily fit every stage of urbanization.

Using a novel methodology and based on a dataset for U.S. cities between 1840 and 2016, the aim of this paper is to answer four important questions: First, which form of the presented family of Zipf models can be used to describe the U.S. city size distribution? Second, do we observe that the U.S. city size distribution exhibits a smooth transition towards the exact Zipf's law from the beginning or are there periods showing a pronounced departure from Zipf's law? Third, if we observe periods of departure, will the city sizes be more equally or unequally distributed than predicted by the exact Zipf's law? Fourth, employing information from the hierarchical structure of cities, do we always find evidence for primate cities for a specific period of time?

In order to answer the raised first three questions, we estimate a more general three-parameter Zipf model, which can be traced back to Mandelbrot (1982). To validate these results and to answer the fourth question, we make use of the finding made by Chen (2012*a*) who shows that Zipf's law can be derived by the hierarchical scaling law based on a hierarchical urban structure. Intuitively, if the top level of a hierarchy is vacant, we can conclude that there is no evidence for primate cities.

The paper makes the following points: First, for the great majority of the examined years between 1840 and 2016, the U.S. city size distribution can be described by a two-parameter Zipf model with a decreasing scaling exponent. From this result we can conjecture that the U.S. city size distribution has become more equally distributed over time thereby diverging from the exact Zipf's law. Moreover, we find evidence for leading cities dominating the remaining largest cities, which indicates that the growing equality of the city sizes is due to the growth of smaller cities instead of a loss of importance of the largest cities. Relating our results with the findings made by Black and Henderson (2003) or Dobkins and Ioannides

[4]Based on the family of Zipf models, the terms "exact Zipf's law" and "one-parameter Zipf model" are used interchangeably in this paper. If the scaling exponent differs from one, we'll receive a two- or three-parameter Zipf model. See section 2 for more details.

(2000), we further conclude that especially in the last decades of the twentieth century, the growth of the largest U.S. areas has mainly taken the form of suburbanization.

The next section outlines the one-, two- and three-parameter Zipf model. In section 2.3 we show the correspondence between Zipf's law and the hierarchical scaling law. Section 2.4 presents the estimation and validation procedure, followed by section 2.5 which presents the data. Section 2.6 discusses the results. Section 2.7 contrasts our findings with the relevant literature and concludes.

2.2 A family of Zipf models

In this section we shortly introduce the three-, two- and one-parameter Zipf models. Suppose that $P(r)$ denotes the size of a city with rank $r = \{1, 2, 3, ...\}$, where the largest city has rank 1. Further, let k denote a scale-translational parameter and define q as the scaling exponent.

2.2.1 The three-parameter Zipf model

The three-parameter model can be traced back to Mandelbrot (1982):

$$P(r) = \frac{\Theta}{(r + \xi)^q}, \tag{2.1}$$

where ξ represents an adjustment parameter, and Θ shows a proportionality coefficient. Recently, Chen and Zhou (2008) have shown that Mandelbrot's (1982) law given with equation (2.1) can be rewritten by capturing the hierarchical urban structure, which we refer to as the three-parameter Zipf model:

$$P(r) = \frac{P_{1-k}}{(r + k)^q}, \tag{2.2}$$

with P_{1-k} showing the size of the $(1-k)$-th possible largest city. The scale-translational parameter k indicates a gap between the largest city in the real world (in the data set) and the possible largest city (estimated by the model). If $k = 1$, for example, in the optimal urban ranking predicted by the model, the proportionality coefficient represents the leading city $P_{1-k} = P_0$. It is larger than the largest city in the data P_1. In contrast to the exact Zipf's law, this model describes a situation of a more evenly $(q < 1)$ or a more unevenly $(q > 1)$ sized distribution without leading cities in the top level of the urban ranking $(k > 0)$.

2.2.2 The two-parameter Zipf model

The two-parameter Zipf model directly arrives from equation (2.2) by letting $k = 0$:

$$P(r) = \frac{P_1}{r^q}, \tag{2.3}$$

with P_1 showing the size of the largest city in the urban hierarchy.[5]

2.2.3 The one-parameter Zipf model

If we further assume that the scaling exponent is $q = 1$, from equation (2.3) we directly deduce the exact form of Zipf's law which states that the rank of a city r is inversely related to its size $P(r)$ (Zipf, 1949)[6]:

$$P(r) = \frac{P_1}{r}. \tag{2.4}$$

2.3 The correspondence between Zipf's law and the hierarchical scaling law

As shown by Chen (2012*a*), Zipf's law can be transformed into the hierarchical scaling law, which can be applied to reveal the scaling relations of the hierarchical structure of the city sizes. We will use the mathematical relationship between the two models in order to validate the estimated one-, two- or three-parameter Zipf model. In what follows, we briefly show the correspondence between the hierarchical scaling law and Zipf's law.

In a first step, we construct a hierarchy of cities. Suppose there are M levels of cities with $m = \{1, 2, ..., M\}$. Further, let f_m be the number of cities in the m-th level, whereas f_1 refers to the number of cities in the top level. P_m is the average size of the cities in the m-th level. Following Chen (2012*a*), the hierarchy of cities can be described by two discrete exponential

[5] If the leading cities are missing ($k > 0$) and the scaling exponent is $q = 1$, we will receive a special form of the two-parameter Zipf model: $P(r) = P_{1-k}/(r + k)$.

[6] The Pareto distribution between the cities' ranks r as the dependent variable and the cities' sizes $P(r)$ as the independent variable was initially observed by Auerbach (1913), who found a Pareto exponent close to one. In theory, this Pareto distribution is equivalent to Zipf's law proposed by Zipf (1949), where the cities' ranks form the independent variable and the cities' sizes the dependent variable. We will refer to this distribution as the one-parameter Zipf model or the exact Zipf's law, if the estimated scaling exponent q is equal to one, and as the two-parameter Zipf model, if $q \neq 1$. A large strand of the literature uses the terms "Zipf's law", "Pareto distribution" and "Power law distribution" effectively synonymously (Newman, 2005).

functions, namely the city number law

$$f_m = f_1 \delta^{m-1}, \tag{2.5}$$

and the city size law

$$P_m = P_1 \lambda^{1-m}, \tag{2.6}$$

with parameters $\delta = \frac{f_{m+1}}{f_m} > 1$ and $\lambda = \frac{P_m}{P_{m+1}} > 1$ referring to the number ratio and the size ratio, respectively.

Using these two equations, Chen (2016) shows that the three-parameter Zipf model can be reinterpreted as a fractal model of the rank-size distribution of cities:

$$f_m = \eta P_m^{-D}, \tag{2.7}$$

with $\eta \equiv f_1 P_1^D$ as a proportionality coefficient. The scaling exponent D is directly associated with the fractal dimension of urban hierarchies:

$$D = -\lim_{m \to \infty} \frac{\ln(\frac{N_{m+1}}{N_m})}{\ln(\frac{P_{m+1}}{P_m})} = -\frac{\ln(\frac{f_{m+1}}{f_m})}{\ln(\frac{P_{m+1}}{P_m})} = \frac{\ln(\delta)}{\ln(\lambda)} \equiv \frac{1}{q}, \tag{2.8}$$

with N_m as the cumulative number of city levels. Taken together, equations (2.7) and (2.8) show a direct correspondence between the hierarchical scaling law and the three-parameter Zipf model presented in equation (2.2).[7]

2.4 Estimation and validation procedure

The detailed approach, how to study the evolution of the U.S. city size distribution relies on Chen (2016). It is as follows:

2.4.1 First step: Determine scaling range

As a first step, the scaling range is determined, which is a straight line on the plot with the logarithmized size of the city on the y-axis and the logarithmized rank of the city on the x-axis. Cities beyond this scaling range represent underdeveloped cities and they are not

[7] See also Chen (2012*b*), p. 3295.

considered in the analysis. Applying an OLS estimation yields a residual value for each city and standardized residuals can be calculated. As proposed by Chen (2015), if a standardized residual value is smaller than -2 or larger than 2, then the associated data point will be treated as an outlier based on the significance level $\alpha = 0.05$ and it will be left out of the estimation.

2.4.2 Second step: Estimate Zipf model

Having defined the scaling range, we apply an OLS estimation to the logarithmized version of the three-parameter model

$$ln(P(r)) = ln(P_{1-k}) - qln(r + k) \tag{2.9}$$

and increase the scale-translational parameter k until the value of goodness of fit R^2 reaches its maximum. According to the k, we receive a Zipf model describing the city size distribution. If $k = 0$ is optimal and the estimation yields as a scaling exponent the value $q = 1$, then we will receive the one-parameter Zipf model (2.4). In case that $k = 0$ and $q \neq 1$ or else $k > 0$ and $q = 1$, we obtain the two-parameter Zipf model (2.3) or a special form of the two-parameter model, respectively. For $k > 0$ and $q \neq 1$, the result is the three-parameter Zipf model (2.2).

2.4.3 Third step: Validate Zipf model

The ascertained model can be transformed into the hierarchical scaling law (2.7), which is based on the hierarchy constructed by the city number law or city size law. In order to validate our Zipf models, we apply the city number law (2.5). Given a number ratio of e.g. $\delta = 2$, then the number of cities in the different levels will be a geometric sequence such as $1, 2, 4, ..., 2^{m-1}$. The average city size at each level can be easily calculated, leading to a number based urban hierarchy. We can make a least square calculation to examine whether the hierarchical scaling law can be well fitted to this hierarchical dataset and thereby whether our estimated Zipf model can be validated.

In order to confirm the above, we utilize the empirical relationship between exponents estimated by means of an OLS regression approach and by means of a Reduced Major Axis regression approach. The fractal dimension D is theoretically related to the Zipf scaling exponent q by $q = 1/D$ (see equation (2.8)). Empirically, q is calculated by $q^{OLS} = R^2/D^{OLS}$. The OLS-estimation can be applied to the Reduced Major Axis regression approach, from which we have

$$D^{RMA} = \sqrt{\frac{D^{OLS}}{q^{OLS}}} = \sqrt{\frac{R^2/q^{OLS}}{q^{OLS}}} = \frac{R}{q^{OLS}}. \tag{2.10}$$

If the difference $|D^{OLS} - D^{RMA}| = |(R^2/q^{OLS}) - (R/q^{OLS})| \to 0$ $(R \to 1)$, we can be sure that the city structure follows the hierarchical scaling law (2.7) and the ascertained Zipf model can be validated.

2.5 Dataset

In order to study the evolution of the U.S. city size distribution, a dataset from the U.S. Bureau of the Census is applied. It contains the population data of the 100 largest urban places in the U.S., which we refer to as "cities" in this paper.[8] As a sample selection criterion, we follow Rosen and Resnick's number threshold approach and examine a fixed number of cities every ten years from 1840 to 2016 (Rosen and Resnick, 1980).[9]

Besides the size distribution of the 100 largest cities, for the last four dates, we additionally consider larger data samples to check whether the results change when including more cities. The dataset is, therefore, supplemented by the sizes of the 601 largest cities for the years 1990 and 2000 as well as the sizes of the 300 largest cities for 2010 and 2016.

[8] Before 1950, urban places were defined as incorporated places with at least 2500 inhabitants. Since 1950, the Census Bureau has differentiated between large cities, which are considered in our study, and urbanized areas in order to account for suburban areas in the vicinity of large cities. The data can be accessed online from `https://www.census.gov/population/www/documentation/twps0027/twps0027.html`, `http://demographia.com/db-uscity98.htm` and `https://www.census.gov/data/tables/2016/demo/popest/total-cities-and-towns.html`.

[9] An overview of the number of inhabitants for selected years can be found in Table 2.1.

Table 2.1: Number of inhabitants of 100 largest U.S. cities

Rank	1860		1910	
1	New York	813,669	New York	4,766,883
2	Philadelphia	565,529	Chicago city	2,185,283
3	Brooklyn	266,661	Philadelphia	1,549,008
4	Baltimore	212,418	St. Louis	687,029
5	Boston	177,840	Boston	670,585
6	New Orleans	168,675	Cleveland	560,663
7	Cincinnati	161,044	Baltimore	558,485
8	St. Louis	160,773	Pittsburgh	533,905
9	Chicago	112,172	Detroit	465,766
10	Buffalo	81,129	Buffalo	423,715
20	Milwaukee	45,246	Kansas	248,381
30	Syracuse	28,119	Toledo	168,497
40	New Bedford	22,300	Paterson	125,600
50	Petersburg	18,266	Albany	100,253
60	Poughkeepsie	14,726	Springfield	88,926
70	Harrisburg	13,405	St. Joseph	77,403
80	Elizabeth	11,567	Evansville	69,647
90	New London	10,115	Charleston	58,833
100	Wilmington	9,552	South Bend	53,684
Rank	**1960**		**2010**	
1	New York	7,781,984	New York	8,175,133
2	Chicago	3,550,404	Los Angeles	3,792,621
3	Los Angeles	2,479,015	Chicago	2,695,598
4	Philadelphia	2,002,512	Houston	2,099,451
5	Detroit	1,670,144	Philadelphia	1,526,006
6	Baltimore	939,024	Phoenix	1,445,632
7	Houston	938,219	San Antonio	1,327,407
8	Cleveland	876,050	San Diego	1,307,402
9	Washington	763,956	Dallas	1,197,816
10	St. Louis	750,026	San Jose	945,942
20	Buffalo	532,759	El Paso	649,121
30	Newark	405,220	Baltimore	620,961
40	St. Paul	313,411	Colorado Springs	416,427
50	Tulsa	261,685	Wichita	382,368
60	Albuquerque	201,189	Lexington-Fayette	295,803
70	Gary	178,320	Newark	277,140
80	Bridgeport	156,748	Laredo	236,091
90	Montgomery	134,393	North Las Vegas	216,961
100	Greensboro	119,574	San Bernardino	209,924

2.6 Results

2.6.1 The evolution of the U.S. city size distribution

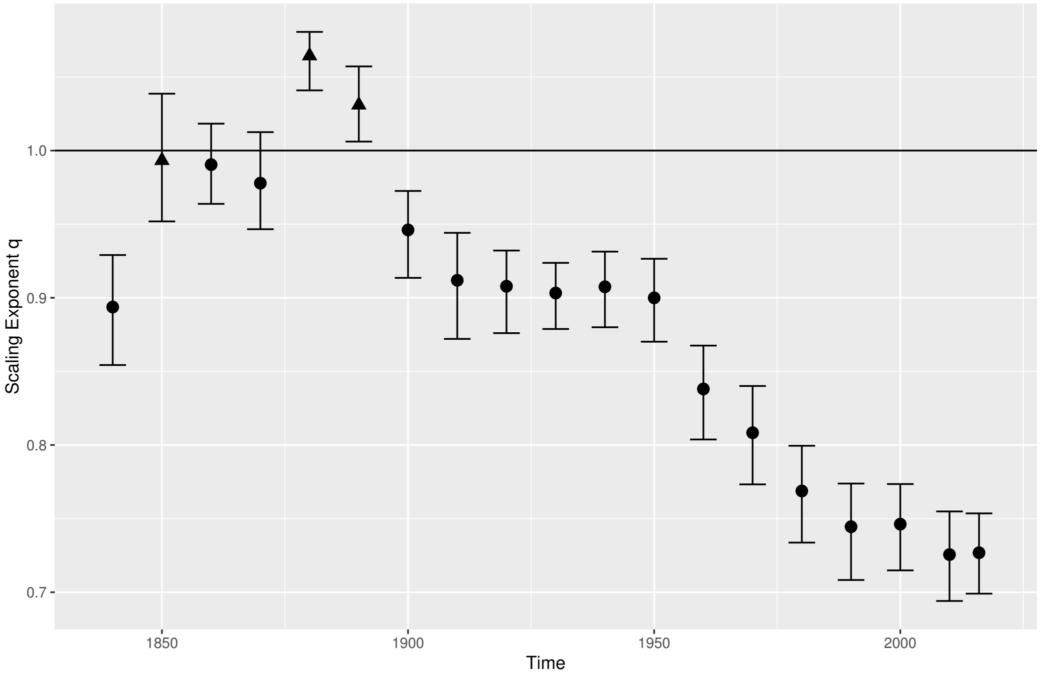
We follow Chen's procedure to estimate the scaling exponent q for different parameters k and stop when the value of goodness of fit (R^2) reaches its highest value (Chen, 2016).[10] Figure 2.1 shows the evolution of the scaling exponent q over time. For most of the years, this maximum is attained for $k = 0$ (see Table 2.2), thus rejecting a three-parameter Zipf model.

Further, the scaling exponent q decreases over the time horizon (see Figure 2.1). For

[10] As an example, the estimation results for the years 1880 and 2016 are depicted in Figure 2.2.

the first 60 years (from 1840 to 1900) q fluctuates around the value one. From 1910 to 1950, we observe that q remains constant taking a value slightly above $q = 0.9$. Starting with the year 1950, the calculated scaling exponent distinctly decreases to $q = 0.75$ in 1990, followed by a further reduction with a value of $q = 0.72$ until 2016. Estimating a scaling exponent q, which is significantly lower than one, indicates a city size distribution which is more equally distributed than expected by the exact Zipf's law.

Figure 2.1: Evolution of the scaling exponent q for the 100 largest U.S. cities between 1840 and 2016.



Notes: The circles (triangles) show the calculated scaling exponents for $k = 0$ ($k = 1$), when the R^2 reaches its maximum. For each estimated scaling exponent q, the 95%-confidence intervals for the bootstrapped estimate of q with 10,000 replications are depicted.
Source: Own illustration based on data by United States Census Bureau.

However, we also find exemptions from this behavior. In particular, we cannot reject the exact Zipf's law ($q = 1$) for the years 1860 and 1870. In the years 1880 and 1890 the value of goodness of fit R^2 did not reach its maximum for $k = 0$, but for $k = 1$. Hence, the U.S. city size distribution followed the three-parameter Zipf model in these years. For the dataset from 1850, $k = 1$ is optimal and $q = 1$ cannot be rejected. We receive a special form of the two-parameter model. As an example, Figure 2.2 compares the estimation results for $k = 0$ and $k = 1$ in the years 1880 and 2016. For 1880, the city size distribution is most accurately described by a three-parameter model ($k = 1$), as the largest city is too small to dominate the remaining cities. Hence, there is a gap between the real largest city in the data and the possible largest city predicted by the model. For the year 2016, on the contrary, the two-parameter model ($k = 0$) fits the data set most accurately. We find evidence for a leading city dominating the remaining largest cities.

Table 2.2: Results of OLS estimation

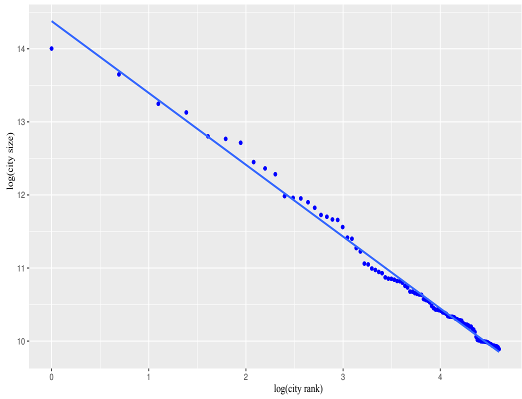
year	scaling range	Zipf model	k	P_{1-k}	q	rse	R^2	D
1840	100	*two*	0	12.4715	0.8937***	0.0192	0.9911	1.1090
1850	100	*two*	0	13.1002	0.9174***	0.0192	0.9888	1.0779
1850	100	*two′*	1	13.4222	0.9933***	0.0235	0.9895	0.9962
1860	100	*one*	0	13.6844	0.9904***	0.0138	0.9944	1.0040
1870	100	*one*	0	14.0502	0.9779***	0.0171	0.9921	1.0146
1880	100	*one*	0	14.3784	0.9829***	0.0186	0.9920	1.0093
1880	100	*three*	1	14.7240	1.0644***	0.0093	0.9929	0.9328
1890	100	*two*	0	14.7002	0.9510***	0.0233	0.9921	1.0432
1890	100	*three*	1	15.0385	1.0310***	0.0131	0.9951	0.9652
1900	100	*two*	0	14.9145	0.9461***	0.0150	0.9936	1.0502
1910	100	*two*	0	15.1230	0.9119***	0.0189	0.9912	1.0870
1920	100	*two*	0	15.3544	0.9078***	0.0146	0.9949	1.0959
1930	100	*two*	0	15.5427	0.9032***	0.0121	0.9948	1.1014
1940	100	*two*	0	15.6018	0.9075***	0.0135	0.9940	1.0954
1950	100	*two*	0	15.7199	0.9000***	0.0142	0.9896	1.0996
1960	100	*two*	0	15.6553	0.8381***	0.0161	0.9850	1.1753
1970	100	*two*	0	15.6316	0.8083***	0.0173	0.9800	1.2124
1980	100	*two*	0	15.5046	0.7688***	0.0172	0.9857	1.2822
1990	100	*two*	0	15.4963	0.7445***	0.0173	0.9873	1.3261
2000	100	*two*	0	15.5892	0.7463***	0.0157	0.9889	1.3252
2010	100	*two*	0	15.5968	0.7256***	0.0161	0.9820	1.3563
2016	100	*two*	0	15.6694	0.7268***	0.0147	0.9902	1.3624
				Larger datasets				
1990	587(601)	*two*	0	15.5562	0.7650***	0.0045	0.9996	1.3026
2000	601(601)	*two*	0	15.5728	0.7423***	0.0042	0.9981	1.3446
2010	299(300)	*two*	0	15.6133	0.7294***	0.0071	0.9961	1.3656
2016	300(300)	*two*	0	15.6779	0.7298***	0.0065	0.9968	1.3659

Notes: k: scaling parameter, P_{1-k}: size of the (1-k)th-city or size of the largest city P_1 for $k = 0$, q: scaling exponent, rse: robust standard error, R^2: value of goodness of fit, D: fractal dimension of urban hierarchies, two': special form of the two-parameter Zipf model, ***: statistically significant at a 1% level

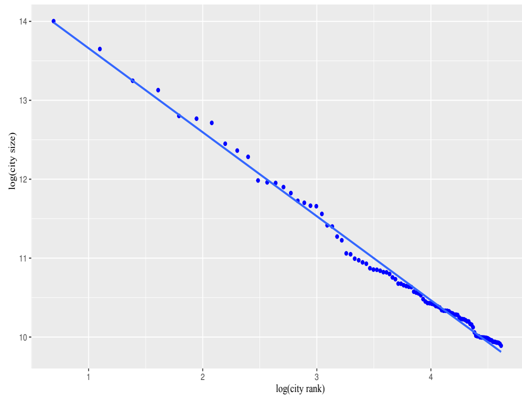
To check the robustness of our findings, we considered larger datasets, yielding rather similar estimation results.[11] We find that the scaling parameter $k = 0$ is optimal and the scaling exponent is slightly decreasing from $q = 0.7650$ in 1990 to $q = 0.7298$ in 2016.

[11] For the year 1990, the 587 largest cities and for 2000 all of the 601 cities are within the scaling range. In 2010, 299 cities and in 2016 all 300 cities are included in the estimation. Section 4.1 precisely describes how to determine the scaling range.

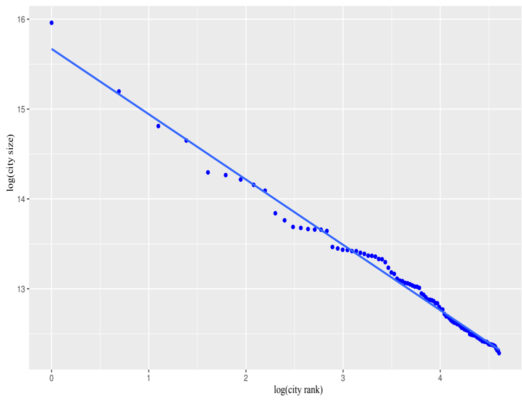
Figure 2.2: Estimation Results for the 100 largest U.S. cities in 1880 and 2016



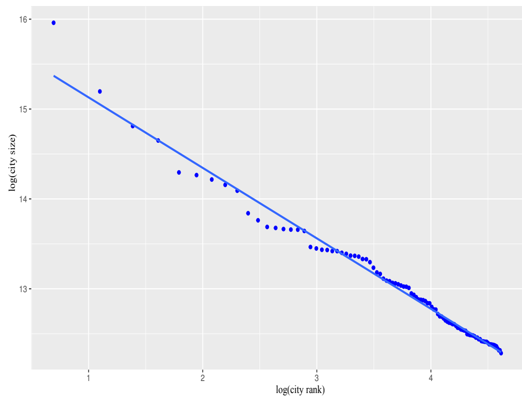
(a) Two-parameter Zipf model for the year 1880 ($k = 0$)



(b) Three-parameter Zipf model for the year 1880 ($k = 1$)



(c) Two-parameter Zipf model for the year 2016 ($k = 0$)



(d) Three-parameter Zipf model for the year 2016 ($k = 1$)

Notes: The city size distribution in 1880 can be described most accurately by means of the three-parameter model ($k = 1$). The city size distribution in 2016 can be described most accurately by means of the two-parameter model ($k = 0$).
Source: Own illustration based on data by United States Census Bureau.

To sum up, and with the exemption of the years 1850 to 1890, the U.S. city size distribution significantly (at a 5% level of significance) follows a two-parameter Zipf model in the years 1840-2016 even when considering larger samples. Hence, we can clearly reject the exact form of Zipf's law for U.S. city data. We find that for most of the years the city size distributions are more equally distributed than expected by the exact Zipf's law and that they have become more equally distributed over time.

2.6.2 The evolution of the hierarchy of the U.S. cities

We exploit the above mentioned dual relationship between the hierarchical scaling and Zipf's law to obtain a more precise understanding of the structure of urban hierarchies. In particular, we want to explore whether or not the existence of primate cities is a time invariant pattern that describes the U.S. city size distribution. In order to answer this question, we have to make sure that the city structure follows a hierarchical scaling law.[12]

According to the city number law (2.5), the cities are ranked into 7 levels. If the one- or the two-parameter Zipf model fits the data, the first level in the hierarchical structure consists of the largest city. The next level comprises the second and third largest cities, the third level consists of the fourth to the seventh largest cities and so on. The last level is supposed to comprise 64 cities, but because our dataset only contains 100 cities, the last level comprises 37 cities. Hence, it is not included in the estimation (see Table 2.3).

Table 2.3: Classification of 100 largest cities in levels (1880 and 2016)

	1880		2016	
Level	City Number	Average City Size	City Number	Average City Size
1			1	8537673.00
2			2	3340640.00
3	4	780829.25	4	1744720.25
4	8	258424.13	8	1020569.88
5	16	99894.31	16	678376.50
6	32	41654.16	32	410409.28
7	40	24470.05	37	256965.24

The estimation of a three-parameter or the special form of the two-parameter model with $k > 0$ suggests an absence of leading cities. That is why the first two levels are absent when constructing the city hierarchy. So, the four largest cities are classed with the third level. Again, the last level comprising the forty smallest cities is not included in the estimation, as it is a lame-duck class.[13]

Looking at Table 2.4, we see that the city structure follows a hierarchical scaling law from 1840 to 1950 as well as for 1990-2016 when larger datasets are used. For these years, our estimated Zipf models can be validated. We can confirm the absence of leading cities for

[12] Detailed information on the construction of urban hierarchies and the validation procedure are given in section 2.4.3.

[13] The classification for the years 1880 and 2016, when a three-parameter model and a two-parameter model hold, is presented in Table 2.3.

the years 1850, 1880 and 1890, in which a three- or special form of the two-parameter model was estimated ($k > 0$), by finding that the hierarchical scaling law fits an urban structure without top levels. For the remaining years, in which we estimated a Zipf model with $k = 0$, we find that the hierarchical scaling law fits an urban structure with the largest cities at the top levels. Hence, we can confirm the existence of leading cities for most of the years. For the 100 largest cities, we observe a pronounced divergence from the hierarchical scaling law starting with 1960 until the year 2016. This can also be seen by comparing the log-rank/log-size plot with the hierarchical scaling relation between the average sizes in the hierarchies of the U.S. cities and the city numbers.

Table 2.4: Validation Results

year	range	q^{OLS}	R^2	D^{OLS}	D^{RMA}	$D^{RMA} - D^{OLS}$
1840	1 to 6	1.0032	0.9860	0.9829***	0.9899	0.0069
1850	3 to 6	1.2783	0.9914	0.7756***	0.7789	0.0033
1860	1 to 6	1.0935	0.9992	0.9137***	0.9141	0.0004
1870	1 to 6	1.0238	0.9960	0.9728***	0.9748	0.0019
1880	3 to 6	1.4057	0.9971	0.7094***	0.7104	0.0010
1890	3 to 6	1.3299	0.9899	0.7443***	0.7481	0.0038
1900	1 to 6	1.0670	0.9874	0.9253***	0.9313	0.0059
1910	1 to 6	1.0509	0.9782	0.9308***	0.9411	0.0103
1920	1 to 6	1.0285	0.9844	0.9571***	0.9647	0.0075
1930	1 to 6	1.0299	0.9893	0.9606***	0.9658	0.0052
1940	1 to 6	1.0283	0.9879	0.9607***	0.9666	0.0059
1950	1 to 6	0.9942	0.9877	0.9935***	0.9996	0.0062
1960	1 to 6	0.9422	0.9806	1.0408***	1.0510	0.0103
1970	1 to 6	0.9147	0.9730	1.0637***	1.0784	0.0147
1980	1 to 6	0.8919	0.9773	1.0957***	1.1083	0.0127
1990	1 to 6	0.8800	0.9747	1.1076***	1.1218	0.0143
2000	1 to 6	0.8777	0.9766	1.1128***	1.1260	0.0132
2010	1 to 6	0.8540	0.9752	1.1419***	1.1564	0.0145
2016	1 to 6	0.8448	0.9775	1.1572***	1.1704	0.0132
			Larger datasets			
1990	1 to 9	0.8225	0.9893	1.2028***	1.2093	0.0065
2000	1 to 9	0.8120	0.9891	1.2181***	1.2248	0.0067
2010	1 to 8	0.8052	0.9856	1.2240***	1.2329	0.0089
2016	1 to 8	0.8022	0.9875	1.2309***	1.2387	0.0078

Notes: q^{OLS}: scaling exponent (OLS-estimation), R^2: value of goodness of fit, D^{OLS}: fractal dimension of urban hierarchies (OLS-estimation), D^{RMA}: fractal dimension of urban hierarchies (RMA-estimation), ***: statistically significant at a 1% level

To sum up, for the 100 largest cities and for most of the time span 1840-2016 we find evidence for leading cities dominating the remaining largest U.S. cities and we find a divergence from the hierarchical scaling law.

2.7 Discussion and Conclusion

This paper reveals the following aspects of the evolution of the U.S. city size distribution: (1) The 100 largest U.S. cities can mostly be described by a two-parameter Zipf model between 1840 and 2016. (2) For most of the years, the examined scaling exponent q is lower than one and it has decreased, especially during the second half of the twentieth century. (3) The U.S. city size distribution has become more even over time and diverged from the exact Zipf's law. (4) For most of the years, we find evidence for leading cities dominating the remaining largest U.S. cities.

When relating our findings to the existing relevant literature, it is striking that the great majority of studies uses cross-sectional data to check whether or not Zipf's law holds exactly.[14] For instance, Krugman (1996) and Gabaix (1999) use data for U.S. Metropolitan Statistical Areas (MSAs) and find that the one-parameter Zipf model holds exactly for a minimum threshold of 280,000 inhabitants. These findings are recently confirmed by Schmidheiny and Suedekum (2015) using novel data from an EC-OECD project. Zipf's law also occurs when applying other city definitions, like economic areas (Berry and Okulicz-Kozaryn, 2012), natural cities (Jiang and Jia, 2011) or geographic clusters (Rozenfeld et al., 2011). Some studies found opposing results for the U.S. city size distribution (Eeckhout, 2004) or found that Zipf's law only holds for the upper tail of the distribution while the body and lower tail are lognormal (Levy, 2009, Malevergne, Pisarenko and Sornette, 2011 and Ioannides and Skouras, 2013). Using U.S. census data, Soo (2005) found that the largest cities are more evenly and the largest urban agglomerations are more unevenly distributed than predicted by the exact Zipf's law (also see Gan, Li and Song, 2006 and Ioannides and Overman, 2003).

Focusing on the long-term perspective of Zipf's law, again, the results depend on the employed city definition. For Metropolitan Statistical Areas, Black and Henderson (2003) or Dobkins and Ioannides (2000) and Dobkins and Ioannides (2001) find an increasing urban concentration, which is higher than predicted by the exact Zipf's law. Other authors focused on states (Soo et al., 2012), counties (Beeson, DeJong and Troesken, 2001 and Desmet and Rappaport, 2017) or minor civil divisions (Michaels, Rauch and Redding, 2012) in the U.S.. Closest to our study is González-Val (2010). Comparing U.S. incorporated places, the author finds that the city sizes are lognormally distributed and more unequally distributed than

[14] A detailed literature review on the theoretical and empirical findings on Zipf's law is given by Arshad, Hu and Ashraf (2018).

predicted by the exact Zipf's law. Regarding the upper tail of the city size distribution, the author finds that the cities become more equally distributed over time.

This is in line with our results, which clearly show that since 1960, the scaling exponent significantly drops year by year until 2016, indicating more evenly distributed city sizes and a departure from Zipf's law for the 100 largest cities. In contrast to González-Val (2010), who explains the convergence of the city sizes with a loss of importance of the largest cities, for most of the time span, we find evidence for leading cities dominating the remaining largest cities. Our results indicate that the growth of the smaller cities plays the main role in the convergence process.

At the same time, Black and Henderson (2003) and Dobkins and Ioannides (2000) found that U.S. MSAs have become more unequally distributed during the twentieth century. Connecting these results to the convergence of city sizes, we found, confirms an increasing suburbanization in the growth process of the largest U.S. urban areas starting in the 1960s (Soo, 2005).[15]

The main point this paper makes is that the U.S. city size distribution has moved away from the exact Zipf's law, especially in the second half of the twentieth century. While for the years 1850, 1880 and 1890, leading cities are missing, they exist for each Census year from 1900 onwards. The scaling exponent decreased, indicating more equally distributed city sizes. In turn, different regimes of Zipf models imply different conditions of city development. Thus, the main deficiency of this paper is that we cannot identify which are the driving forces leading to this evolutionary development away from the exact Zipf's law over time. A more elaborated investigation is definitely needed, but beyond the scope of this paper. Besides that, the rather subjective definition of a city might influence the results this paper makes. Insofar, the results cannot be generalized to other countries or to the same country but with a different city definition. However, this problem is common to every study dealing with city-level data.

[15] According to Boustan and Shertzer (2013), a large portion of suburbanization in the U.S. over the twentieth century can be explained by factors associated with the natural evolution process of urbanization, like rising incomes, which led to a larger demand for housing and land, as well as transportation improvements, especially the growing network of interstate highways. Furthermore, the authors state that factors associated with the flight-from-blight theory of suburbanization, like school quality, taxes, crime-rates and socioeconomic factors of the population, reinforced the spatial dispersion. Also see Mieszkowski and Mills (1993), Bayoh, Irwin and Haab (2006) and Kim (2000).

References

Arshad, Sidra, Shougeng Hu, and Badar Nadeem Ashraf. 2018. "Zipf's law and city size distribution: A survey of the literature and future research agenda." *Physica A: Statistical mechanics and its applications*, 492: 75–92.

Auerbach, Felix. 1913. "Das gesetz der bevölkerungskonzentration." *Petermanns Geographische Mitteilungen*, 59: 74–76.

Bayoh, Isaac, Elena G Irwin, and Timothy Haab. 2006. "Determinants of residential location choice: How important are local public goods in attracting homeowners to central city locations?" *Journal of Regional Science*, 46(1): 97–120.

Beeson, Patricia E, David N DeJong, and Werner Troesken. 2001. "Population growth in US counties, 1840–1990." *Regional Science and Urban Economics*, 31(6): 669–699.

Berry, Brian, and Adam Okulicz-Kozaryn. 2012. "The city size distribution debate: Resolution for US urban regions and megalopolitan areas." *Cities*, 29: 17–23.

Black, Duncan, and Vernon Henderson. 2003. "Urban evolution in the USA." *Journal of economic geography*, 3(4): 343–372.

Boustan, Leah, and Allison Shertzer. 2013. "Population trends as a counterweight to central city decline, 1950–2000." *Demography*, 50(1): 125–147.

Chen, Yanguang. 2012*a*. "The mathematical relationship between Zipf's law and the hierarchical scaling law." *Physica A: Statistical Mechanics and its Applications*, 391(11): 3285–3299.

Chen, Yanguang. 2012*b*. "Zipf's law, 1/f noise, and fractal hierarchy." *Chaos, Solitons & Fractals*, 45(1): 63–73.

Chen, Yanguang. 2015. "The distance-decay function of geographical gravity model: Power law or exponential law?" *Chaos, Solitons & Fractals*, 77: 174–189.

Chen, Yanguang. 2016. "The evolution of Zipf's law indicative of city development." *Physica A: Statistical Mechanics and its Applications*, 443: 555–567.

Chen, Yanguang, and Yixing Zhou. 2008. "Scaling laws and indications of self-organized criticality in urban systems." *Chaos, Solitons & Fractals*, 35(1): 85–98.

Desmet, Klaus, and Jordan Rappaport. 2017. "The settlement of the United States, 1800–2000: The long transition towards Gibrat's law." *Journal of Urban Economics*, 98: 50–68.

Dobkins, L.H., and Y. M. Ioannides. 2000. "Dynamic evolution of the US city size distribution." In *The economics of cities.* , ed. J. M. Huriot and J. F. Thisse, 217–260. Cambridge University Press.

Dobkins, Linda Harris, and Yannis M Ioannides. 2001. "Spatial interactions among US cities: 1900–1990." *Regional science and urban Economics*, 31(6): 701–731.

Eeckhout, Jan. 2004. "Gibrat's law for (all) cities." *American Economic Review*, 94(5): 1429–1451.

Gabaix, Xavier. 1999. "Zipf's law for cities: an explanation." *The Quarterly journal of economics*, 114(3): 739–767.

Gan, Li, Dong Li, and Shunfeng Song. 2006. "Is the Zipf law spurious in explaining city-size distributions?" *Economics Letters*, 92(2): 256–262.

González-Val, Rafael. 2010. "The evolution of US city size distribution from a long-term perspective (1900–2000)." *Journal of Regional Science*, 50(5): 952–972.

Ioannides, Yannis, and Spyros Skouras. 2013. "US city size distribution: Robustly Pareto, but only in the tail." *Journal of Urban Economics*, 73(1): 18–29.

Ioannides, Yannis M, and Henry G Overman. 2003. "Zipf's law for cities: an empirical examination." *Regional science and urban economics*, 33(2): 127–137.

Jiang, Bin, and Tao Jia. 2011. "Zipf's law for all the natural cities in the United States: a geospatial perspective." *International Journal of Geographical Information Science*, 25(8): 1269–1281.

Kim, Sukkoo. 2000. "Urban Development in the United States, 1690-1990." *Southern Economic Journal*, 66(4): 855–880.

Krugman, Paul. 1996. "Confronting the mystery of urban hierarchy." *Journal of the Japanese and International economies*, 10(4): 399–418.

Levy, Moshe. 2009. "Gibrat's law for (all) cities: Comment." *American Economic Review*, 99(4): 1672–75.

Malevergne, Yannick, Vladilen Pisarenko, and Didier Sornette. 2011. "Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities." *Physical Review E*, 83(3): 036111.

Mandelbrot, Benoit B. 1982. *The fractal nature of geometry.* W. H. Freeman and Company, New York.

Michaels, Guy, Ferdinand Rauch, and Stephen J Redding. 2012. "Urbanization and structural transformation." *The Quarterly Journal of Economics*, 127(2): 535–586.

Mieszkowski, Peter, and Edwin S Mills. 1993. "The causes of metropolitan suburbanization." *Journal of Economic perspectives*, 7(3): 135–147.

Newman, Mark EJ. 2005. "Power laws, Pareto distributions and Zipf's law." *Contemporary physics*, 46(5): 323–351.

Rosen, Kenneth T, and Mitchel Resnick. 1980. "The size distribution of cities: an examination of the Pareto law and primacy." *Journal of urban economics*, 8(2): 165–186.

Rozenfeld, Hernán D, Diego Rybski, Xavier Gabaix, and Hernán A Makse. 2011. "The area and population of cities: New insights from a different perspective on cities." *American Economic Review*, 101(5): 2205–25.

Schmidheiny, Kurt, and Jens Suedekum. 2015. "The pan-European population distribution across consistently defined functional urban areas." *Economics Letters*, 133: 10–13.

Soo, Kwok Tong. 2005. "Zipf's Law for cities: a cross-country investigation." *Regional science and urban Economics*, 35(3): 239–263.

Soo, Kwok Tong, et al. 2012. "The size and growth of state populations in the United States." *Economics Bulletin*, 32(2): 1238–1249.

Zipf, George Kingsley. 1949. *Human Behaviour and the Principle of Least Effort.* Addison-Wesley, Cambridge MA.

CHAPTER

THREE

HOUSE PRICE CONVERGENCE ACROSS GERMAN REGIONS

Author: Angelina Hackmann[1]

Abstract

This paper analyses long-run house price dynamics across German labour market regions over the years 2007 to 2017. By means of a log t convergence test and clustering algorithm by Phillips and Sul (2007), we find no overall house price convergence across regions as well as within pre-determined subgroups based on historical/geographical linkages or demographic similarities. We can instead endogenously identify multiple convergence clusters. While lower house price clubs mostly comprise regions in the middle, the east (except regions around Berlin) and in the south west of Germany, convergence clubs with higher average house price growth mostly consist of regions in southern Germany as well as the seven largest cities and their surrounding regions. These high house price clusters around the largest cities indicate a diminishing monocentric structure and a gentrification process in the cities' fringes. Furthermore, we detect population developments and housing supply as key drivers of convergence club membership.

JEL Classification: C33, O18, R11, R21, R31, R52

Keywords: Convergence, House price, Relative transition, House price dynamics, German regions

[1]University of Bremen. E-mail: angelina.hackmann@uni-bremen.de.

3.1 Introduction

After a period of stagnation that lasted for almost two decades, German house prices have grown at an accelerated pace in recent years. While the sharp increase in prices starting in the year 2010 was largely concentrated in urban areas, it only took off gradually in non-urban regions. As of roughly 2015, the large upward pressure on residential prices has become more widespread across German regions (Deutsche Bundesbank, 2020). This severe acceleration of house price growth has caused the European Systemic Risk Board (ESRB) to issue a warning to Germany given its systemic risk to financial stability. The board noticed a significant overvaluation of house prices in urban areas for the years 2016, 2017 and 2018, reflecting a shortage of housing supply relative to demand (European Systemic Risk Board, 2019).[2] According to Dustmann, Fitzenberger and Zimmermann (2018), such increases in housing expenditures lead for low-income households to a decrease in saving rates and a loss in real disposable income. For high-income households, these changes are reversed and the share spent on housing decreases. Hence, increases in house prices, lead to inequality in disposable income, as well as a divergence in consumption and saving patterns across income groups.[3] With regard to its implications for income and wealth inequality, it is crucial to closely examine potential regional house price divergence as well as driving forces behind house price growth.

A theoretical foundation for house price dispersion is given by Rosen (1979) and Roback (1982) who modeled the relationship between incomes, amenities, and housing costs across metropolitan areas. According to these models, population, wages, and housing costs will adjust in such a way that the marginal resident in each location receives identical utility from amenities, housing, and wages, so that no resident will wish to move. Consequently, house price differences reflect a diversity in local amenities and wages. This model was extended by Van Nieuwerburgh and Weill (2010) who provide a spatial equilibrium model, in order to analyse the joint dynamics of the entire cross-section of house prices, construction, and wages. While keeping the dispersion of ability and housing supply regulation constant, diverging

[2] While, for example, the population in the Top 7 cities has grown on average by 1.0% per annum between 2015 and 2019, the remaining NUTS-3 regions have grown on average by 0.4% per annum. During the same time period the number of new residential buildings has increased in the seven largest cities by 5.0% per annum and in the remaining NUTS-3 regions by 14.3% per annum. These numbers give a first impression of the rising excess demand for housing in the urban areas over the last years.

[3] On the inter-linkages of house markets and households' wealth distribution, see also Kuhn, Schularick and Steins (2018), Albouy, Ehrlich and Liu (2016), Backhaus, Gebers and Schröder (2015), Scoccianti (2010).

productivity, and therewith, wages result in households moving towards high-productivity metropolitan areas. As a consequence of this influx combined with limited housing supply, local house prices in these areas increase, while the prices in low-productivity areas decrease. Overall house price dispersion is enhanced. Their model illustrates that an increase in regional income inequality is an essential part in explaining the increased house prices dispersion in the U.S.. Housing supply regulations seem to have a quantitatively small impact on the inequality of house prices.

Regarding the house price distribution within an urban region, from the well-known Alonso-Mills-Muth model (Alonso, 1964, Mills, 1967 and Muth, 1969), we expect higher house prices in the center and lower prices towards the fringe and the periphery of a city. This model assumes a city with a central business district and a fixed population with a given income level. The further away a worker lives from the center, the higher the commuting costs. Since not everyone can live in the center, the prices for land and housing are more expensive the more central the area is. Hence, a household can choose between living centrally in a small and expensive home or living further towards the city fringe with larger and less expensive housing. These assumptions result in a monocentric urban structure with high house prices in the central area of a city and lower prices towards the fringe and the city's periphery. This model can explain divergent house price developments even within an urban region.

Against this theoretical background, the aim of this study is to examine the evidence for regional house price convergence in Germany. As a first step, we employ a regression based convergence test by Phillips and Sul (2007) in order to determine whether house prices across all 141 labour market regions converge to a common steady state. If overall convergence has to be rejected, we test whether subgroups of regions, which are determined on the basis of historical/geographical linkages as well as the seven largest cities, show a convergent house price behavior. If we cannot find house price convergence within these pre-determined subgroups, a club convergence algorithm by Phillips and Sul (2007), which endogenously forms subgroups of regions with converging house prices, will be applied. As these clubs do not necessarily coincide with commonly known classifications of German regions, as a last step, we study general characteristics of these subgroups as well as possible factors driving convergence club membership.

The log t convergence test and the clustering algorithm by Phillips and Sul (2007) have been applied to various research fields, such as the convergence of consumer prices

(Phillips and Sul, 2007), GDP (Phillips and Sul, 2009, Bartkowska and Riedl, 2012, Lyncker and Thoennessen, 2017) or carbon dioxide emissions (Burnett, 2016). With regard to the housing market, this procedure was used to examine various regions worldwide. Encouraged by the inconsistent fall of house prices across the U.S. during the 2000s, Kim and Rous (2012) examine house price convergence in a panel of U.S. states and three panels of metropolitan areas. With their results, the authors reject overall convergence within these panels. Applying the clustering algorithm proposed by Phillips and Sul (2007), the revealed club memberships do not align with commonly defined regions in the U.S.. As drivers of club membership, the authors find housing supply regulation as well as climate conditions to be most important. Besides the U.S. housing market, this procedure was also used to examine regional house price developments in the UK (Montagnoli and Nagayasu, 2015), Spain (Blanco, Martín and Vazquez, 2016), Poland (Mateusz, 2019, Matysiak and Olszewski, 2019), Australia (Awaworyi Churchill, Inekwe and Ivanovski, 2018), South Africa (Apergis, Simo-Kengne and Gupta, 2015) and a sample of European countries (Tsai, 2018). Besides these cross-regional studies, another focus of research is on intra-regional house price convergence. By means of the pairwise approach, Abbott and De Vita (2012) reject overall multidistrict long-run convergence in Greater London. Furthermore, the authors find that the boroughs, which are contiguous to the City of London district, show the highest rate of convergence. Holmes, Otero and Panagiotidis (2019) studied intra-regional house price convergence for England and Wales between 1995 and 2017. Using the log t regression and the clustering algorithm proposed by Phillips and Sul (2007), the authors do not find overall convergence, but they detect four convergence clubs. When focusing on 32 London boroughs and the City of London, Holmes, Otero and Panagiotidis (2019) also reject overall convergence and for this regional subsample, they identify four convergence clusters.

To the best of our knowledge, no existing study has examined either intra- or inter-regional house price convergence for Germany. With regard to the strong and broadly based increase of house prices in recent years, we intend to fill this gap by answering the following questions: Do we find overall house price convergence for Germany between 2007 and 2017? Are there subgroups of regions based on historical/geographical linkages or demographic similarities with a convergent house price behavior? Which regional factors drive endogenously determined house price convergence clubs?

The remainder of the paper proceeds as follows: Section 3.2 describes the methodology, more precisely, the log t convergence test and clustering algorithm. Furthermore, it derives drivers of regional house prices from the inverted demand equation and presents an ordered logit model, which determines the importance of these drivers for the convergence club membership. In Section 3.3 our dataset is described. Section 3.4 reports our empirical results regardig the clustering analysis and the drivers of club membership. Section 3.5 concludes.

3.2 Methodology

A log t convergence test by Phillips and Sul (2007) is applied in order to find out whether the concept of relative convergence applies to the German housing market. For this purpose, we examine whether the cross-sectional dispersion of house prices decreases over time. The applied method allows for a wide range of transitional dynamics and individual heterogeneity. Furthermore, it does not depend on any particular assumptions concerning trend stationarity or stochastic non-stationarity in the common component. If the overall hypothesis of convergence has to be rejected, we apply a clustering algorithm proposed by Phillips and Sul (2007) , which allows to detect convergence clubs as well as diverging groups. The corresponding R code for these estimations and the algorithm can be found in the Appendix.

3.2.1 Log t Convergence Test

Let X_{it} be an observable time series that represents the growth of logarithmized house prices in region $i = 1, ..., N$ at time $t = 1, ..., T$. It is given by the sum of systematic components g_{it} and transitory components a_{it}:

$$X_{it} = g_{it} + a_{it}. \tag{3.1}$$

As both elements, g_{it} and a_{it}, can contain common and idiosyncratic components, we transform the time series in order to separate these components:

$$X_{it} = (\frac{g_{it} + a_{it}}{\mu_t})\mu_t = \delta_{it}\mu_t, \tag{3.2}$$

where δ_{it} represents the idiosyncratic part that varies over time. It measures the distance between common growth path μ_t and X_{it} for region i. This time-varying factor representation allows for individual heterogeneity as well as periods of transition in a path, which is ultimately governed by a common long run stochastic trend. Given these characteristics in the data as well as a short panel, conventional cointegration tests might suggest the absence of cointegration even though two time series converge. Hence, instead of analyzing their difference or linear combinations, Phillips and Sul (2007) propose to define convergence between two time series X_{it} and X_{jt} as their ratio. The authors state that a relative long run equilibrium among the time series exists if

$$\lim_{k\to\infty} \frac{X_{it+k}}{X_{jt+k}} = 1 \qquad \text{for all } i \text{ and } j, \tag{3.3}$$

which is equivalent to

$$\lim_{k\to\infty} \delta_{it+k} = \delta. \tag{3.4}$$

To trace out an individual trajectory for region i, the relative transition parameter h_{it} can be constructed. It measures the time-varying factor-loading coefficient δ_{it} relative to the panel average at time t and therewith the relative departure from the common growth path μ_t for region i:

$$h_{it} = \frac{X_{it}}{(1/N)\sum_{i=1}^{N} X_{it}} = \frac{\mu_t \delta_{it}}{(1/N)\sum_{i=1}^{N} \mu_t \delta_{it}} = \frac{\delta_{it}}{(1/N)\sum_{i=1}^{N} \delta_{it}}. \tag{3.5}$$

The cross-sectional mean of h_{it} is unity and the cross-sectional variance is defined as:

$$H_t = \frac{1}{N}\sum_{i=1}^{N}(h_{it}-1)^2. \tag{3.6}$$

When the panel units converge, then δ_{it} converges to δ, h_{it} to unity and H_t to zero. The latter condition is utilized in the following log t test procedure in order to test the null hypothesis of convergence. For this purpose, we estimate the regression

$$\log(\frac{H_1}{H_t}) - 2\log(\log(t)) = \hat{a} + \hat{b}\log t + \hat{u}_t \quad \text{for } t = [rT], [rT]+1, ..., T, \qquad r > 0 \tag{3.7}$$

and test the null hypothesis

$$\mathcal{H}_0 : \delta_i = \delta \quad \text{ and } \alpha \geq 0$$

against the alternative

$$\mathcal{H}_{\mathcal{A}} : \delta_i \neq \delta \quad \text{for all } i, \text{ or } \alpha < 0.$$

Using the fitted coefficient $\hat{b} = 2\hat{\alpha}$, an autocorrelation and heteroskedasticity robust one-sided t test is applied to test the inequality null hypothesis $\alpha \geq 0$. Under convergence ($H_t \to 0$ for $t \to \infty$) $\log(H_1/H_t)$ diverges to ∞, when $\alpha \geq 0$. The convergence hypothesis is rejected if $t_{\hat{b}} < -1.65$.

3.2.2 Clustering Algorithm

Even if we reject the null hypothesis of full panel convergence, we might identify converging subgroups. We use an algorithm proposed by Phillips and Sul (2007), which sorts regions into subgroups based on repeated log t regressions. Additionally, we present alterations of the algorithm proposed by Phillips and Sul (2009), Schnurbus, Haupt and Meier (2017), and Lyncker and Thoennessen (2017).[4]

Step 1: Cross Section Ordering

The dataset is ordered by sorting the last observation in the panel by its size, beginning with the highest entry.

Step 2: Core Group Formation

The subgroup $G_k = \{1, 2, ..., k\}$ with $2 \leq k < N$ is formed starting with the k highest individuals in the panel. For each subgroup, the convergence test statistic $t_k = t(G_k)$ is computed and the core group size k^* is chosen by maximizing t_k over k according to:

$$k^* = arg \max_{2 \leq k \leq \bar{K}} \{t_k\}, \quad \text{where } \bar{K} = arg \min_{2 \leq k \leq N} \{t_k > -1.65\}. \tag{3.8}$$

If the condition $t_k > -1.65$ does not hold for $k = 2$, then the highest individual in G_k is dropped and will be considered again for subsequent convergence groups. Step 2 is repeated starting with the second highest individual. This procedure is reiterated until a sequential

[4]The R code in the Appendix also has the option to apply these alterations of the clustering algorithm. For the estimations in this paper the algorithm by Phillips and Sul (2007) together with alterations by Schnurbus, Haupt and Meier (2017) are employed.

pair or more individuals with the highest entries are found, for which the convergence test statistic is larger than -1.65.

Step 3: Sieve Individuals for Club Membership

The individuals, that are not in the core group, are added one by one and the corresponding t-value t_c is computed. Each individual, for which the t-value is larger than the critical value $c^* = 0$, is added to the club and an overall log t test is performed for the club. If the t-value for the expanded group is larger than -1.65, then a convergence club is found. If not, we differentiate between two methods.

3.1) Phillips and Sul (2007), Phillips and Sul (2009): The authors propose to reiterate Step 3 with a higher value for c^*. The increase of c^* is repeated until the overall log t test for the core group and the new cluster candidates results in a t-value larger than -1.65 and thus a convergence club is found.

3.2) Instead of raising the critical value by a manual intervention, Schnurbus, Haupt and Meier (2017) propose to order the club candidates with respect to a decreasing t-value t_c obtained in Step 3. The club candidate with the highest test statistic is added to the core group and new candidates are sieved for this expanded core group. The candidates are again ordered according to a decreasing t-value t_c and the candidate with the highest t-value is added to the core group. This procedure is repeated until the highest t-value is less than -1.65. Then the expanded core group equates the definite convergence club.

Step 4: Stopping Rule

The individuals that have not been chosen to be members of a convergence club (Step 2 and 3) form a subgroup. If the overall log t test for this subgroup leads to a t-value larger than -1.65, the last cluster is found. If not, Steps 1 to 3 are repeated in order to find out whether this group can be further subdivided into convergence clubs. In case no two individuals can be found to build a core group in Step 2, the remaining individuals are ordered into the divergence group.

Step 5: Merging of Clubs

Steps 1 to 4 can lead to conservative clustering with more clubs than necessary. While Phillips and Sul (2007) stop at Step 4, there are three different approaches to merge clubs and thereby decrease the number of clubs.

5.1) Phillips and Sul (2009) propose to run the log t regression for all pairs of subsequent initial clubs and test whether they fulfill the convergence hypothesis.

5.2) Schnurbus, Haupt and Meier (2017) implement multiple iterations of the merging procedure by Phillips and Sul (2009). That means, after having merged all subsequent clubs, for which the overall log t test results in a t-value larger than -1.65, the procedure of building pairs of subsequent clubs and testing the convergence hypothesis is reiterated until no more two clubs can be found that can be merged.

5.3) Lyncker and Thoennessen (2017) decide that two subsequent clubs n and $n+1$ can be merged if the overall log t test of these two clubs is larger than -1.65 and if the t-value of these two clubs is larger than the overall t-value of the clubs $n+1$ and $n+2$. If these two conditions hold, then the clubs n and $n+1$ can be merged. Thereupon, these two conditions are checked for the newly merged club n and its adjacent club $n+1$, and so on. If one of the conditions is not fulfilled, then the next two clubs $n+1$ and $n+2$ are analysed with regard to the given conditions. This procedure ends, when no more two clubs can be found that can be merged.

Step 6: Merging of Divergence Group

Even though the diverging individuals could not be merged with any existing club in Step 2 to 3, Lyncker and Thoennessen (2017) see the opportunity for diverging individuals to be merged with the merged clubs from Step 5. For that purpose, the log t test is performed for each member of the diverging group together with each of the merged clubs at a time. If the highest t-value of these combinations is larger than -1.65, then the respective diverging individual is added to the respective merged club. The procedure is continued by performing a log t test with all diverging individuals and newly merged clubs at a time until no more combination can be found, for which the log t test results in a t-value larger than -1.65. All regions, which are left, form the divergence group.

3.2.3 The inverse demand approach

The convergence algorithm by Phillips and Sul (2007) endogenously sorts regions into clubs based on the regions' house price convergence to a similar steady state. As the algorithm is merely based on the regions' house price growth, the question arises, which are the driving factors behind the convergence club memberships and how do these regional characteristics affect the likelihood for a region to be a member of a certain convergence club. The theoretical background on drivers of house prices is presented by the inverted demand equation. Empirically, an ordered logit model is applied to investigate the role of certain region-specific characteristics in determining club membership (Section 3.2.4).

The drivers of regional house prices are derived from an inverted demand equation for housing based on Kajuth, Knetsch and Pinkwart (2013) by inverting and rearranging the housing demand equation so that house prices are the dependent variable.

The house price equation can be derived as follows. The housing stock at the beginning of period t (hs_t) is given by the sum of the housing stock at $t-1$, which is adjusted for depreciation at the rate δ, and residential investment in the previous period b_{t-1}

$$hs_t = (1-\delta)hs_{t-1} + b_{t-1}. \tag{3.9}$$

The demand for housing hd_t may be specified as:

$$\ln hd_t = \alpha_0 + \alpha_1 \ln y_t - \alpha_2 \ln m_t + \alpha_3 \ln x_t, \tag{3.10}$$

where y_t is income, m_t is the housing rent and x_t are other factors, such as demographic and labour market characteristics (Cameron, Muellbauer and Murphy, 2006*b*). A standard intertemporal arbitrage equation yields the connection between house prices hp_t and rent m_t

$$ry_t = \frac{\mathbb{E}_t hp_{t+1} + m_t - hp_t}{hp_t}, \tag{3.11}$$

in which the ex-ante real yield from renting out one housing unit ry_t and the expected future gains have to be equalized. The expected house price in the next period $(t+1)$ is given by $\mathbb{E}_t hp_{t+1}$. In order to receive house prices as dependent variable, Equation (3.11) has to be re-arranged, iterated forward and the transversality condition is imposed. As a consequence, house prices can be expressed as sum of discounted expected future rent payments

$$p_t = \mathbb{E}_t \sum_{k=0}^{\infty} \frac{m_{t+k}}{\prod_{l=0}^{m}(1+ry_{t+l})}. \tag{3.12}$$

Imposing (1) the assumption of rents being a constant fraction of income and growing at an average long-term rate g_t^e

$$m_{t+k} = (1+g_t^e)^k m_t \qquad \text{for } k = 1, 2, ...\infty \tag{3.13}$$

as well as (2) the assumption that future rent payments can be discounted at an average long-term real interest rate r_t

$$(1+r_t)^k = \prod_{l=0}^{m}(1+ry_{t+l}) \qquad \text{for } k = 1, 2, ...\infty \tag{3.14}$$

and (3) the assumption that $r_t - g_t^e > 0$, yields an equation for the price-to-rent ratio

$$\frac{hp_t}{m_t} = \frac{1}{r_t - g_t^e}. \tag{3.15}$$

As a last step, by substituting out m_t by means of Equation (3.10), by applying the equilibrium condition $\ln hd_t \equiv \ln hs_t$ and by taking logs, we receive the inverted demand equation for housing:

$$\ln hp_t = \frac{\alpha_0}{\alpha_2} - \frac{1}{\alpha_2}\ln hs_t + \frac{\alpha_1}{\alpha_2}\ln y_t + \frac{\alpha_3}{\alpha_2}\ln x_t - \ln(r_t - g_t^e). \tag{3.16}$$

According to Equation (3.16), house prices hp_t depend negatively on housing stock hs_t as well as the difference of the average long-term real interest rate and the average long-term growth rate of rents $(r_t - g_t^e)$, positively on income y_t and on other factors x_t. The direction of the impact of x_t depends on the specific characteristic.

3.2.4 The ordered logit model

By means of an ordered logit model, we empirically analyse how certain regional characteristics affect the likelihood of a region to be a member of each convergence club. The set of regional characteristics is theoretically based on the inverted demand equation (3.16).

Let Y be an ordinal outcome with I categories and $X = (x_1, x_2, ..., x_p)'$ be a vector of p explanatory variables (see Section 3.3). The cumulative probability $P(Y \leq i)$ is the probability that Y falls at or below a particular point given by the category $i = 1, ..., I-1$. The odds of being less than or equal to a particular category can be defined as

$$\frac{P(Y \leq i|X)}{P(Y > i|X)} \qquad \text{for } i = 1, 2, ...I-1 \tag{3.17}$$

with $P(Y \leq 1|X) \leq P(Y \leq 2|X) \leq ... \leq P(Y \leq I|X) = 1$ and $P(Y > I|X) = 0$. The logits of the cumulative probabilities (log odds) are given by

$$log\frac{P(Y \leq i|X)}{P(Y > i|X)} = logit(P(Y \leq i|X)) \tag{3.18}$$

and the ordinal logistic regression model is parameterized as follows

$$logit(P(Y \leq i|X)) = \alpha_{i0} - \beta_1 x_1 - ... - \beta_p x_p. \tag{3.19}$$

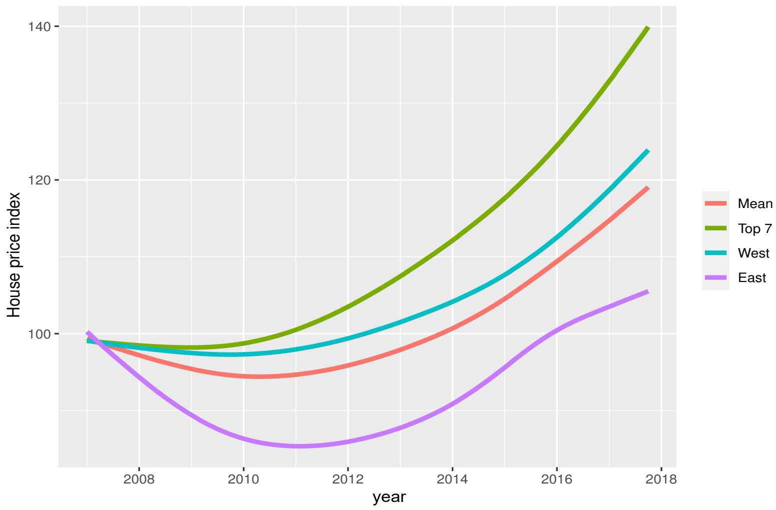
Furthermore, we estimate the marginal effects of the predicted probabilities in order to assess how the probability of belonging to a club changes for a one unit change in one of the explanatory variables while holding all other variables fixed at their sample averages (Agresti, 2003).

3.3 Data

For our estimations we employ the RWI-GEO-RED dataset providing a monthly house price index between 2007 and 2017 for German labour market regions by the FDZ Ruhr (research data center at the RWI - Leibniz Institute for Economic Research). This house price index was calculated by means of a hedonic price regression and a granular dataset based on real estate offers published on the largest German listing website ImmobilienScout24. It contains information, inter alia, on the size of the house, its facilities, features, energy consumption and regional information to the $1km^2$ grid level. [5] To the best of our knowledge, the present index is the best publicly avaibable house price index on a disaggregated level (labour market regions) at a monthly frequency.

[5]For a more detailed description on the granular dataset and the hedonic price regression, see Bauer et al. (2013).

Figure 3.1: Heterogeneous house price development in German regions between 2007 and 2017



Sources: Research Data Center FDZ Ruhr (RWI Essen) and author's calculations.
Notes: "East" represent the mean house price index of all labour market regions located in the federal states Brandenburg, Mecklenburg-West Pomerania, Saxony, Saxony Anhalt, Thuringia and Berlin and "West" the mean house price index of all regions located in the states Baden-Wurttemberg, Bavaria, Bremen, Hamburg, Hesse, Lower Saxony, Nordrhein-Westfalen, Rhineland Palatinate, Saarland, and Schleswig-Holstein. "Top 7" shows the mean house price index of the seven largest cities (Hamburg, Berlin, Düsseldorf, Cologne, Frankfurt am Main, Stuttgart and Munich). "Mean" represents the mean of all 141 labour market regions in the panel.

Figure 3.1 plots the monthly average house price development for all labour market regions in our panel between 2007 and 2017 (red line). The average house price index decreased from January 2007 to May 2010 and increased afterwards. Considerable differences appear, when dividing the labour market regions into East and West German regions.[6] While the West German house price index grows from October 2009 onwards, the East German house price index shows positive growth rates starting in March 2011. Furthermore, between June 2016 and October 2017 the growth rates of the West German price index are about twice as large as the growth rates of the East German index. Outstanding is the house price development in the Top 7 cities in Germany, namely Hamburg, Berlin, Düsseldorf, Cologne, Frankfurt am Main, Stuttgart and Munich. Their average house price index in October 2017

[6]East Germany comprises the new states of Germany, which are five re-established states of the former German Democratic Republic (GDR) that unified with the Federal Republic of Germany (FRG) in 1990. The new states are Brandenburg, Mecklenburg-West Pomerania, Saxony, Saxony Anhalt, and Thuringia. The state of Berlin, the result of a merger between East and West Berlin, is also ordered into the subgroup of East Germany given its geographic location. West Germany comprises 11 states, namely Baden-Wurttemberg, Bavaria, Bremen, Hamburg, Hesse, Lower Saxony, Nordrhein-Westfalen, Rhineland Palatinate, Saarland, and Schleswig-Holstein.

is 139.9, while the average house price index of all regions is 119.1 in the same month. This chart indicates an increase in house price dispersion between East and West German regions over time, but an even larger discrepancy to the Top 7 cities.

In order to improve the finite sample power and size of the log t test, we eliminate the cyclical component of the data by means of the Hodrick and Prescott filter (Hodrick and Prescott, 1997). In contrast to house price level data, the convergence of our price indices is subject to the base month, which is January 2007. Hence, as suggested by Phillips and Sul (2007), the fraction $r = (1/3)$ of the time series is discarded in order to rule out initial effects created by the base year initialization.

The set of regional characteristics X, which represents the set of explanatory variables in the ordered logit model in Equation (3.19), is based on the inverted demand equation (3.16). Given the fact that we are merely interested in explaining the determination of clubmembership based on regional house price growth, we only consider the region-specific explanatory variables in Equation (3.16).

Based on the theoretical and empirical literature on house price determinants (Table 3.1), the variable x_t - other factors - comprises key drivers of regional house prices in addition to the other explanatory variables in equation (3.16). Hence, the set of explanatory variables comprises the housing stock per capita hs_{it} as supply side determinant. Regarding the demand side of housing, we include real GDP per capita (y_{it}). Demographic and labour market characteristics are population density and growth, accounting for a scale effect that regions with higher population density and growth show higher house prices. These characteristics also include employment per capita as an indicator of region-specific income perspectives as well as the share of young population capturing life-cycle motives (Kajuth, Knetsch and Pinkwart, 2013).[7] We collect the regional indicators for employment, GDP and population from the ARDECO database, population density from Eurostat and housing stock from GENESIS, the German Federal Statistics Office data platform.

[7]As theoretically presented by Glaeser, Gyourko and Saks (2006), house prices can have an impact on income and population growth. Hence, in order to avoid endogeneity problems, the explanatory variables, real GDP per capita, employment per capita, population growth and density, are given as averages between the years 2000 and 2006. The variable housing stock is given in per capita levels in 2007 and young population is calculated as share of population at the age of 1 to 64 years of the total population in 2007. In order to make the variables comparable in size, average real GDP per capita is given by $10,000*$ GDP per capita between 2000 and 2006 and average population growth in % between 2000 and 2006. Population density is calculated as the average of $1,000*$ population per square kilometre between 2000 and 2006.

Table 3.1: Literature overview on drivers of house prices

Explanatory variable	Country	Literature
Housing stock	Germany	Kajuth, Knetsch and Pinkwart (2013)
		Belke and Keil (2018)
	US/UK	Giussani and Hadjimatheou (1992)
		Muellbauer and Murphy (1997)
		Meen (2001)
		Cameron, Muellbauer and Murphy (2006*b*)
		Caldera and Johansson (2013)
Income	Germany	Maennig and Dust (2008)
		Koetter and Poghosyan (2010)
		Bischoff (2012)
		Kajuth, Knetsch and Pinkwart (2013)
	US/UK	Abraham and Hendershott (1996)
		Muellbauer and Murphy (1997)
		Meen (2001)
		Jud and Winkler (2002)
		Cameron, Muellbauer and Murphy (2006*b*)
		Holly, Pesaran and Yamagata (2010)
		White (2015)
Population growth	Germany	Maennig and Dust (2008)
		Koetter and Poghosyan (2010)
		Bischoff (2012)
		Kajuth, Knetsch and Pinkwart (2013)
	US/UK	Reichert (1990)
		Murphy and Muellbauer (1994)
		Jud and Winkler (2002)
		Cameron, Muellbauer and Murphy (2006*a*)
		Cameron, Muellbauer and Murphy (2006*b*)
		Holly, Pesaran and Yamagata (2010)
		Sivitanides (2018)
Population density	Germany	Kajuth, Knetsch and Pinkwart (2013)
	US/UK	Miles (2012)
Employment	Germany	Kajuth, Knetsch and Pinkwart (2013)
	US/UK	Reichert (1990)
		Murphy and Muellbauer (1994)
		Abraham and Hendershott (1996)
		Baffoe-Bonnie (1998)
		Hyclak and Johnes (1999)
		Cameron, Muellbauer and Murphy (2006*a*)
Active Population	Germany	Kajuth, Knetsch and Pinkwart (2013)
	US/UK	Cameron, Muellbauer and Murphy (2006*b*)

3.4 Results

The first step of our empirical study is to estimate the log t regression given by Equation (3.7) for the whole panel of 141 regions. It yields a $\hat{b}$ value of -1.4 and a t-statistic of -18.7 rejecting overall convergence with the t-statistic being lower than -1.65 (5% significance level). These results indicate house price divergence within the full panel, but subgroups of regions may form convergence clubs. Hence, we will examine whether subgroups of German regions with historical/geographical linkages as well as the seven largest cities, show a convergent house price behaviour.

3.4.1 Clustering Analysis

We start the clustering analysis by employing a geographical classification into East and West Germany. Despite substantial improvements in recent decades, there is still an economic gap between East and West Germany with the East lagging behind the West regarding economic measures, such as unemployment and productivity (Bundesministerium für Wirtschaft und Energie, 2020). By means of a spatial equilibrium model Van Nieuwerburgh and Weill (2010) illustrate that an increase in income inequality is an essential part in explaining the increased house prices dispersion in the U.S.. Hence, we want to find out whether the rejection of overall house price convergence among the whole panel of 141 German regions can be explained by the economic gap between East and West Germany by finding overall convergence within these two subgroups. As shown in Table 3.2, the log t test yields a $\hat{b}$ value of -1.4 and a t-statistic of -14.0 rejecting overall house price convergence for East Germany. The same holds true for West Germany with a $\hat{b}$ value of -2.4 and a t-statistic of -42.3.

As a second step of the clustering analysis, we employ a geographical classification of our 141 labour market regions into 16 German federal states. If a labour market region is located in more than one state, then it is assigned to the state with its highest population share. Regions within a federal state are geographically neighboring, which might result in house price ripple effects within a state. Meen (1999) explains the diffusion of house prices by means of the so-called ripple effect. That means that shocks to a local housing market can spread out to the surrounding markets, which leads to house prices moving together in the long run. The author finds that house prices rise at first in the south-east of the UK and then gradually spread out to the rest of the country. As shown in Table 3.2, for each of the states, which comprise two regions or more, the log t test yields a t-value much lower than -1.65, so overall house price convergence within each German federal state has to be rejected.

Table 3.2: Convergence club classification for subgroups

Subgroup	No. of regions	$\hat{b}$	$t_{\hat{b}}$	No. of clubs	Divergent regions
All regions	141	-1.439	-18.735	7	0
East	35	-1.396	-14.04	5	1
West	106	-2.361	-42.333	7	0
BW	20	-1.657	-217.351	4	1
BY	28	-2.065	-40.899	6	1
BE	1				
BB	11	-1.618	-19.066	4	0
HB	2	-5.084	-25.471	0	2
HH	1				
HE	7	-2.736	-51.168	2	2
MV	5	-1.946	-19.1	2	1
NI	14	-1.521	-46.281	2	2
NW	18	-2.727	-93.814	3	4
RP	11	-2.834	-22.249	2	2
SL	1				
SN	4	-1.906	-54.037	1	2
ST	6	-1.013	-5.62	1	3
SH	4	-0.277	-5.626	1	1
TH	8	-0.763	-63.453	2	1
Top 7	7	-4.057	-40.007	1	5

Notes: The log t test is applied to all regions within each subgroup and it is distributed as a one-sided t-statistics with a 5% critical value of -1.65. States: Baden-Wurttemberg—BW; Bavaria—BY; Berlin—BE; Brandenburg—BB; Bremen—HB; Hamburg—HH; Hesse—HE; Lower Saxony—NI; Mecklenburg-West Pomerania—MV; North Rhine Westfalia—NW; Rhineland Palatinate—RP; Saarland—SL; Saxony—SN; Saxony Anhalt—ST; Schleswig-Holstein—SH; Thuringia—TH.

In contrast to these subgroups, which are determined by geographical proximity, we now turn to the seven largest regions within our data set based on the population size in 2017. Once again, overall house price convergence within this subgroup has to be rejected with a t-statistic of $-40.0 < -1.65$. While the cities Stuttgart and Hamburg converge to one steady state, the cities Düsseldorf, Cologne, Berlin, Munich and Frankfurt am Main are detected as divergent regions.

In sum, we cannot find evidence of house price convergence for subgroups of regions based on historical, geographical or population characteristics. Thus, an alternative grouping method, in particular the club convergence algorithm by Phillips and Sul (2007), which endogenously forms subgroups of regions with converging house prices, will be applied.

Table 3.3: Convergence club classification

Club clustering (Phillips and Sul, 2007)						
Club	Number of regions	$\hat{b}$	Std. err.	$t_{\hat{b}}$	House price index 10/2017	Average house price growth
1	2	3.394	0.755	4.495	173.908	0.446
2	8	1.142	0.184	6.197	153.607	0.340
3	7	0.123	0.230	0.537	148.519	0.317
4	10	1.277	0.044	28.759	139.910	0.261
5	33	0.446	0.023	19.742	131.792	0.221
6	20	0.466	0.025	18.869	119.327	0.144
7	10	0.400	0.122	3.285	116.795	0.130
8	10	0.326	0.166	1.970	103.694	0.028
9	13	0.484	0.017	28.192	108.859	0.068
10	6	0.296	0.019	15.727	94.488	-0.045
11	9	0.145	0.126	1.153	101.162	0.018
12	5	0.052	0.119	0.441	86.756	-0.124
13	5	0.376	0.053	7.091	83.686	-0.157
Div	3				113.703	0.045

Merging of clubs (Lyncker and Thoennessen, 2017)							
Club	Merging	Number of regions	$\hat{b}$	Std. err.	$t_{\hat{b}}$	House price index 10/2017	Average house price growth
1	1	2	3.394	0.755	4.495	173.908	0.446
2	2+3	15	0.765	0.040	18.977	151.233	0.329
3	4+5	43	0.274	0.032	8.500	133.680	0.231
4	6+7	30	0.389	0.045	8.645	118.483	0.139
5	8+9	23	0.481	0.097	4.979	106.613	0.051
6	10+11	15	0.322	0.030	10.662	98.492	-0.007
7	12+13	10	0.177	0.095	1.869	85.221	-0.140
	Div	3				113.703	0.045

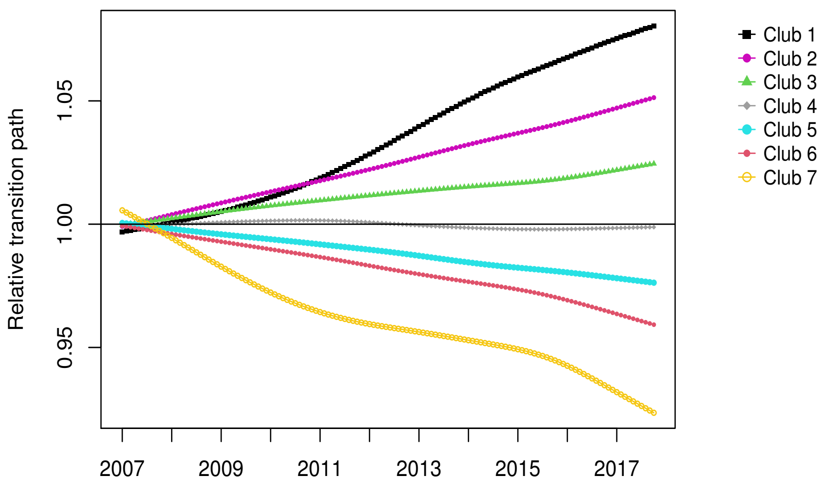
Merging of divergence group (Lyncker and Thoennessen, 2017)							
Club	Merging	Number of regions	$\hat{b}$	Std. err.	$t_{\hat{b}}$	House price index 10/2017	Average house price growth
2	2+3+Div	16	0.601	0.033	18.056	151.765	0.332
4	6+7+Div	31	0.387	0.047	8.287	118.386	0.139
7	12+13+Div	11	0.001	0.159	0.008	83.465	-0.159

Notes: Average house price growth between 2007 and 2017. The log t test is distributed as a one-sided t-statistics with a 5% critical value of -1.65.

As presented in Table 3.3, the algorithm reveals 13 convergence clubs with an estimated

$\hat{b} \geq 0$ and three divergent regions.[8] The speed of convergence $\hat{\alpha} = \hat{b}/2$ is the largest for club 1, which comprises the labour market regions Munich and Ingolstadt. The estimated $\hat{b} = 3.394$ indicates level convergence of the regions' house prices within this club ($\hat{b} \geq 2$). For the other clubs, the clustering procedure yields $2 > \hat{b} \geq 0$, which corresponds to conditional convergence, meaning the growth rates of the regions' house prices within these clubs converge over time. The descriptive statistics show that the first club consists of regions with the highest average house price growth between 2007 and 2017 and the largest house price index in 2017. With few exceptions, these two descriptive statistics decrease with an increasing club number. Hence, Club 13 shows the lowest house price index in 2017 and on average negative growth between 2007 and 2017.

Figure 3.2: Relative transition paths



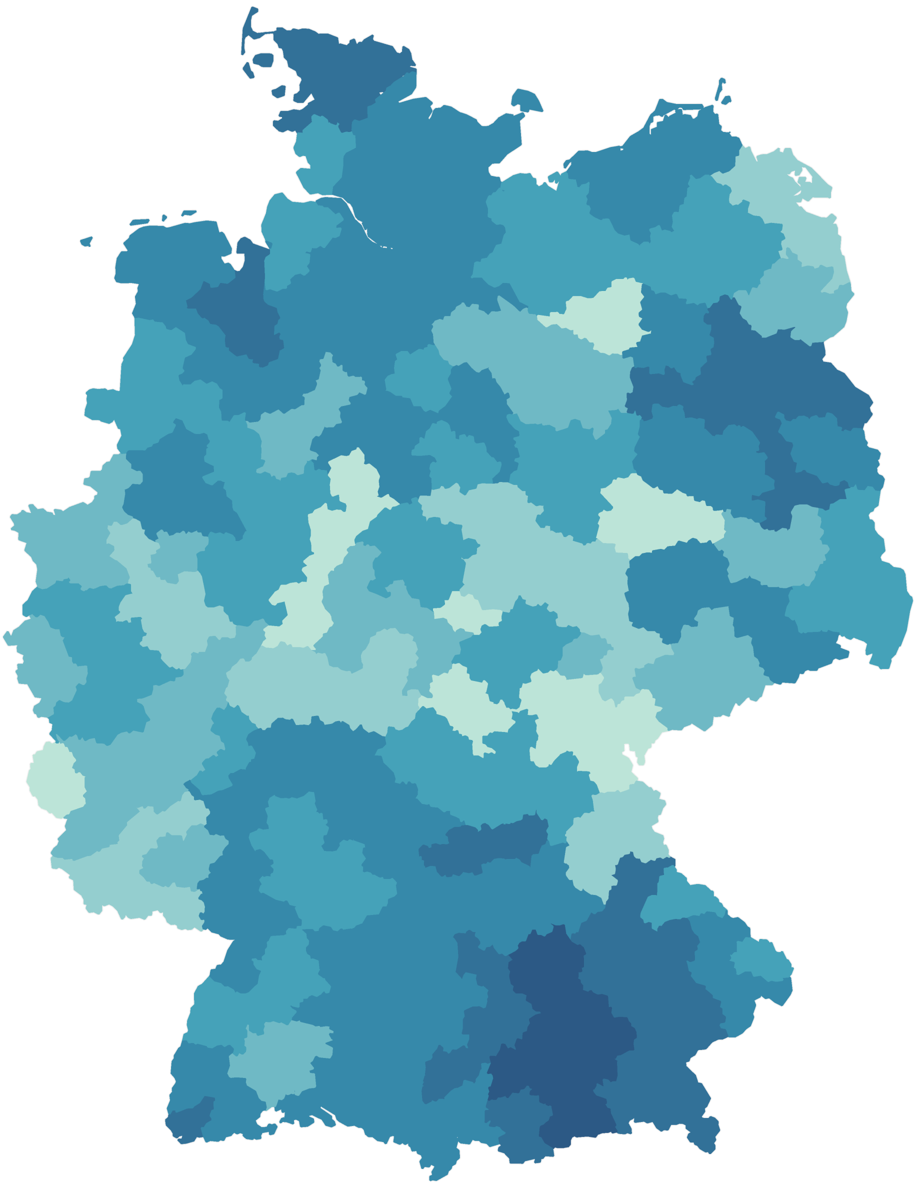
Notes: The relative transition paths of the seven different clubs are calculated as the cross-sectional mean of the relative transition paths of all members of each club.

In order to avoid conservative clustering with more clubs than necessary, the merging algorithm proposed by Lyncker and Thoennessen (2017) is applied. Clubs 2 and 3 can be merged to a new club with $2 > \hat{b} \geq 0$ indicating conditional convergence for the members of the newly merged club. The same holds true for clubs 4 and 5, 6 and 7, 8 and 9, 10 and 11, as well as 12 and 13, which can also be merged to new convergence clubs. Taken together, the original 13 convergence clubs are merged to 7 convergence clubs, and still, three divergent regions. As a last step, we test whether the divergent regions can be merged with the existing

[8] Phillips and Sul (2007) show that the log t regression test is consistent even in the (boundary) case where $\hat{\alpha} = \hat{b}/2 = 0$ (see Phillips and Sul, 2007, p. 1790).

convergence clubs. The divergent region Augsburg can be merged with club 2 (consisting of the original clubs 2 and 3), Coburg can be merged with club 4 (consisting of the original clubs 6 and 7) and Prignitz can be merged with club 7 (consisting of the original clubs 12 and 13), yielding $2 > \hat{b} \geq 0$ and therefore conditional convergence, respectively.

Figure 3.3: Map of final convergence clubs



Notes: labour market regions on the map are filled with different shades of blue indicating the clubmembership. The darkest blue represents the first club and, therefore, highest house price growth regions and the lightest blue represents club 7 with the lowest average house price growth.

Figure 3.2 illustrates the cross-sectional mean of the relative transition paths of all members of each club. The relative transition paths are given by equation (3.5), measuring the transition path of house prices in region i relative to the panel average. If the relative transition paths tend to unity, all regional house prices converge to the same level. Assuming club convergence, on the other hand, the transition paths of the members of each club tend to different levels. As shown in Figure 3.2, the latter holds true for our data set. While house prices in clubs 1, 2 and 3 develop above the average, clubs 5, 6 and 7 display a clear downward trend below the average. Club 4 appears to develop close to the panel average. In general, no evidence of a convergence process between clubs can be observed, rejecting overall convergence as well as confirming the estimated club membership.

Figure 3.3 provides a graphic illustration of house price club membership across the German labour market regions. A large proportion of members of the first two clubs, which are by construction regions with high house price growth, can be found in the south of Germany. The two labour market regions with the highest house price growth are Ingolstadt and Munich in Bavaria. Regarding the north of Germany, the labour market regions Oldenburg and Flensburg show the highest house price growth and are ordered into the second club. Another high house price growth club consists of Berlin and its surrounding labour market regions. The lower house price growth clubs mostly comprise regions in Central Germany, the east of Germany (except regions surrounding Berlin), and in the south-west (Rhineland Palatinate). To sum up, labour market regions in the south and in the north of Germany as well as Berlin and its surrounding regions show the highest house price growth, while house prices in Central Germany have grown at a lower rate.

Regarding the Top 7 cities in Germany, Munich belongs to the first club and Berlin is ordered into the second club. The labour market regions Stuttgart, Hamburg and Frankfurt am Main belong to the third and Cologne and Düsseldorf to the fourth club. It is observable that high house price regions cluster around Munich and Berlin. This is most pronounced for Munich - the region with the second largest house price growth in our panel -, which is surrounded by regions belonging to the first two clubs. The clustering of high house price growth regions around large cities is also observable for Hamburg, Frankfurt am Main and Stuttgart, but to a much lesser extent than for Munich and Berlin. These clusters of high house prices around the largest cities are in line with a gentrification process within regions described by Glaeser, Kolko and Saiz (2001) and Guerrieri, Hartley and Hurst (2013). They argue that a demand shock induces wealthier households to move to the city's fringe or periphery, which leads to an expansion of high-priced areas. As a consequence of this influx, the positive externality from living in that area, and thereby also its house prices, increase. This gentrification process in the city's fringe or periphery leads to a larger area of high house prices and therewith a more homogeneous house price distribution within the region. A former monocentric urban structure as described by the well-known Alonso-Mills-Muth model (Alonso, 1964, Mills, 1967 and Muth, 1969) diminishes through this gentrification process.

3.4.2 Drivers of club membership

So far, we have shown that overall house price convergence has to be rejected for all 141 labour market regions, as well as within the subgroups East and West Germany, the federal states and the Top 7 cities. However, by applying the clustering algorithm by Phillips and Sul (2007), we endogenously determined subgroups of regions within which house prices converge to their common price. As these convergence clubs do not coincide with commonly known classifications of German regions, it is of great interest to investigate the characteristics of these convergence clubs. For each house price convergence club determined in Section 3.4.1, Table 3.6 presents the mean and standard deviation of variables, which have been determined to be key drivers of house prices on the regional level (see Section 3.2.4). Regions in the first club have on average the highest real GDP per capita, employment per capita and population growth between the years 2000 and 2006 as well as the highest share of young population in 2007. With few exceptions, these variables decrease in size with an increasing club number. The average housing stock per capita in the year 2007, on the other hand, is lowest for the first club and - with exceptions - it increases with an increasing club number. Apart from average population density between 2000 and 2006, the table shows that the variable means increase/decrease with an increasing club number, which suggests linkages between the clubs' house prices and possible drivers of club membership. Hence, an ordered logit model is applied in order to analyse the complex interactions between club membership and the clubs' characteristics and possibly confirm these first considerations.

The estimation results from an ordered logit model are presented in Table 3.4. The dependent variable represents the club membership and it takes the value 1 for regions in club 7, 2 for regions in club 6,..., 7 for regions in club 1. As a result, the club with the highest number (club 7) shows on average the highest house price converging level, which will facilitate the interpretation of the results. In column (1) we report the parameter estimates of the ordered logit model and find a positive and significant coefficient (at the 1% level) for population growth indicating that a region with a higher value is more likely to belong to a club with a higher house price converging level. Regions with a higher population density and housing stock are more likely to belong to a club with a lower house price converging level (significant at the 1% level). Regions with lower employment per capita and regions with higher GDP per capita and share of young population are more likely to belong to a club with higher house prices, but these coefficients are not statistically significant, indicating that these variables do not appear to be key factors in determining club membership. These results contradict Kajuth, Knetsch and Pinkwart (2013), who find for German NUTS-3 regions that income per capita, unemployment, and the share of active population have significant effects on house prices.

Table 3.4: Estimation results from ordered logit model

	Coefficient	Std. err.	t statistics	p value
GDP	0.691	0.488	1.418	0.156
Employment	-7.785	5.817	-1.338	0.181
Population growth	2.299	0.479	4.796	0.000
Population density	-3.510	0.780	-4.500	0.000
Young population	9.215	14.309	0.644	0.520
Housing stock	-22.474	5.199	-4.323	0.000
1|2	-4.372	11.548	-0.379	0.705
2|3	-3.088	11.555	-0.267	0.789
3|4	-1.783	11.566	-0.154	0.878
4|5	-0.355	11.567	-0.031	0.976
5|6	1.895	11.550	0.164	0.870
6|7	4.559	11.544	0.395	0.693

In addition to the parameter estimates, Table 3.5 displays the marginal effects at the mean from the ordered logit model, which show the change in probability of belonging to a specific club for a small change in the explanatory variable after setting all other covariates at their means. A one unit increase in population growth significantly decreases the probability of belonging to club 1 to 3, while it increases the probability of being in club 5 and 6. For population density and the housing stock, the opposite holds true: A one point increase in each of these variables significantly increases the probability of being in club 1 to 3 and it decreases the likelihood of being in club 5 and 6. The marginal effects for club 7, which comprises the two regions with the highest house price growth, namely Ingolstadt and Munich, are not significant for any of the variables. Furthermore, an increase in GDP, employment and the share of young population does not significantly change the probability of belonging to a specific club.

Table 3.5: Marginal effects from ordered logit model

Club	GDP	Employment	Population growth	Population density	Young population	Housing stock
1	-0.019	0.213	-0.063***	0.096***	-0.252	0.614**
2	-0.04	0.454	-0.134***	0.205***	-0.538	1.312***
3	-0.08	0.896	-0.265***	0.404***	-1.061	2.588***
4	-0.025	0.277	-0.082	0.125	-0.328	0.799
5	0.124	-1.391	0.411***	-0.627***	1.647	-4.016***
6	0.037	-0.414	0.122***	-0.187***	0.49	-1.196***
7	0.003	-0.035	0.01	-0.016	0.042	-0.102

Notes: Statistical significance: *** p <0.01, ** p<0.05, * p<0.1

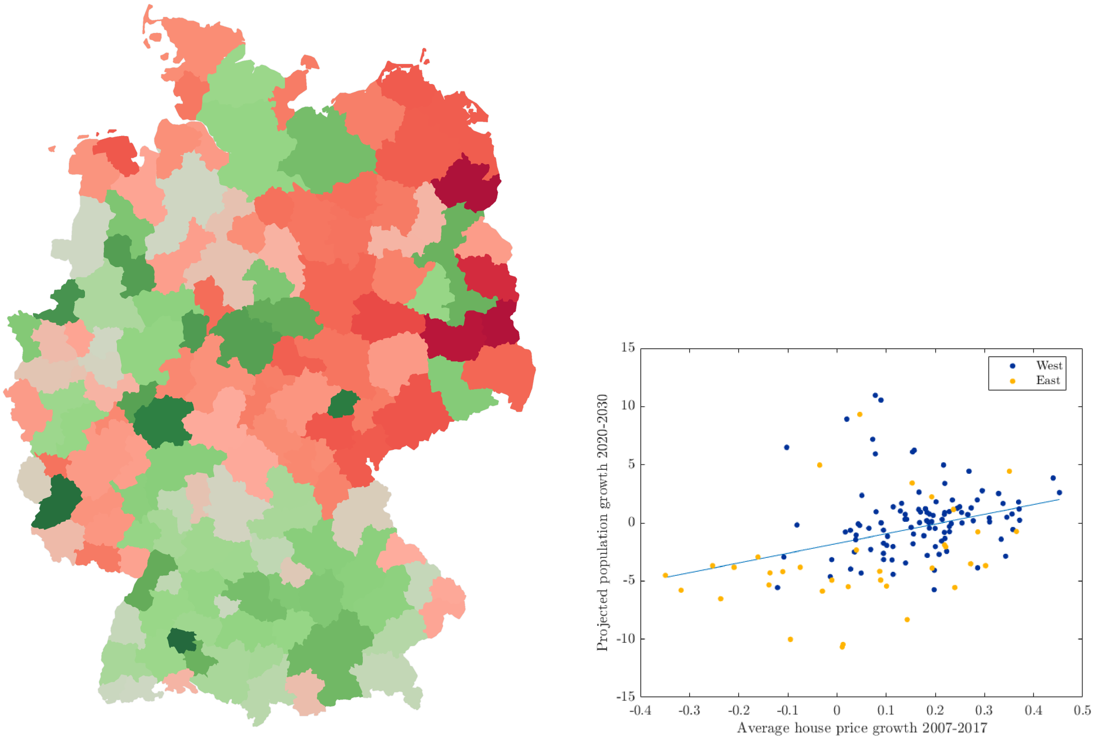
These results are not surprising and follow economic theory with the exception of the results for population density. It is striking that an increase in population density increases the probability of being in club 1 to 3 with the lowest house price converging level. On the one hand, this contradicts empirical results for Germany (Kajuth, Knetsch and Pinkwart, 2013) and theoretical considerations by Miles (2012). The author develops a model of the housing market with the evolution of population density as the major determinant of house price rises relative to incomes. For a densely populated country, the model predicts that house prices will ultimately rise faster than incomes as population density rises. On the other hand, Dunse, Thanos and Bramley (2013) find a non-linear quadratic relationship between house prices and population density for five cities in the UK. More precisely, the relationship is found to be concave for London meaning that house prices rise with increasing density until a point of inflection is reached and house prices decrease with increasing density. For other cities, the authors find a convex relationship. These contrasting results highlight the ambiguous relationship between house prices and density and show that density can be viewed both as amenity and as disamenity. There can be various reasons, why we find a negative relationship between house prices and density. First, our convergence algorithm uses a house price index, so that we only have information on the growth of house prices between 2007 and 2017, but not about the actual house price level. Highly populated regions in the year

2006 may only show moderate house price growth between 2007 and 2017, because they may already have extraordinary high house price levels before 2007, and therefore less potential for rising prices. As a result, we find a negative relationship between population density and house price growth, while there may be a positive underlying relationship between population density and actual house price levels. Furthermore, labour market regions are rather large entities, which can comprise very heterogeneous areas. The labour market region Hamburg, for example, consists of the highly populated city Hamburg, but also 5 surrounding NUTS-3 regions, which are connected to the city by commuter flows, but which are less dense. Hence, for future research it would be interesting to consider smaller entities than labour market regions in order to have a more accurate measure of population density as well as actual house price levels to examine the relationship between house prices and population density on a regional level.

The result that demographic developments, especially population growth, has an impact on future house price growth is of particular interest when regarding population forecasts for German labour market regions. As shown in Figure 3.4(a), Eurostat projects highly heterogeneous population developments on the regional level. While a maximum population growth of 10.96% is expected for the Zollernalb district in Baden-Wurttemberg, the population in the labour market region Uckermark in Brandenburg is projected to shrink by 10.68% between the years 2020 and 2030. In general, regions in southern Germany are mostly expected to grow, whereas regions in eastern Germany, with the exception of Berlin and its surrounding areas, are projected to decrease regarding their population. These first considerations indicate that regions with high house price growth between 2007 and 2017 are expected to grow further in their population size until 2030.

As shown in the scatter plot in Figure 3.4(b), the range of average house price growth between 2007 and 2017 and projected population growth between 2020 and 2030 is wide, suggesting a highly heterogeneous current house price and future population development across German labour market region. Moreover, the correlation between the two variables at the regional level is clearly positive, which indicates that many regions with high house price growth between 2007 and 2017 are expected grow in their population size until 2030. Connecting this consideration to our result that regions with high population growth are more likely to belong to a club with high house price converging level, indicates that the demand for housing will rise even further in regions with high house price growth while

Figure 3.4: Population growth and house prices



(a) Projected population growth between 2020 and 2030

(b) Average house price growth and projected population development

Source: Eurostat, Research Data Center FDZ Ruhr (RWI Essen) and author's calculations.
Notes: Panel (a) The map depicts projected population growth between 2020 and 2030 by Eurostat. The red color indicates a shrinking population and the green color indicates population growth. The stronger a region grows/shrinks, the more intense is the color. Panel (b) shows on the x-axis the average house price growth between January 2007 and October 2017. The y-axis reports projected population growth between 2020 and 2030. Each dot represents a labour market region.

the opposite holds true for regions with lower house price growth between 2007 and 2017. As a consequence, the divergence of regional house prices would become more pronounced, enhancing the inequality in regional housing affordability.

To sum up, population growth and population density as well as the supply side variable, namely housing stock, are key drivers in determining house price club membership. A one unit increase in population growth significantly increases the probability of belonging to a higher house price club, whereas an increase in population density and housing stock decreases this probability, and vice versa.

3.5 Conclusion

In this paper, we assessed the dynamics of the German housing market over the years 2007 to 2017. Rejecting the hypothesis of overall convergence, the results of the log t test propose a dispersion of house prices across the 141 German labour market regions. Given the overall divergent house price development, we analyse the convergence behaviour for subgroups of regions. As a first step, these subgroups are determined on the basis of historical/geographical linkages or similarities in their size, namely East and West Germany, the German states and the Top 7 cities. The hypothesis of overall house price convergence within each of these subgroups has to be rejected.

In contrast to these pre-determined subgroups of regions, as a next step, the convergence clustering algorithm by Phillips and Sul (2007) with alterations by Schnurbus, Haupt and Meier (2017) is applied, which endogenously forms 7 subgroups of regions with converging house prices. By construction of the algorithm, the members of the first club, Ingolstadt and Munich, show on average the highest house price growth between 2007 and 2017 as well as the highest house price index in 2017. These variables are decreasing in size with an increasing club number. Regarding the geographical distribution, lower house price clubs mostly comprise regions in the middle, the east (except regions around Berlin) and in the south west of Germany (Rhineland Palatinate and Saarland). Convergence clubs with a higher average house price growth mostly consist of regions in southern Germany, particularly in Bavaria and Baden-Wurttemberg. Furthermore, the seven largest cities and most of their surrounding regions belong to high house price clubs indicating a high house price cluster formation which is most pronounced for Munich and its surrounding regions. As described by Glaeser, Kolko and Saiz (2001) and Guerrieri, Hartley and Hurst (2013), this cluster formation around large cities, which goes in line with a diminishing monocentric urban structure, can be explained through a gentrification process in the city's fringe or periphery. It can be expected that this development will be enhanced by the COVID-19 crisis. Accompanying regulations, such as closed schools, offices, stores, and cultural facilities, added to the drawbacks of living in small and expensive homes in a central area while letting the amenities of central homes diminish. Even for the post-pandemic era, more flexible working arrangements, and consequently the acceptance of longer commuting times, may encourage more households to move to the cities' peripheries and enhance the ongoing gentrification processes for the largest cities and their surrounding regions. This development requires investments in infrastructure, broadband, public transport, and services for the cities' peripheries, which will need to be in the focus of policymakers.

As these convergence clubs are determined endogenously, the cluster formation does not depend on arbitrarily selected variables or thresholds, but as a consequence, its results are also somewhat atheoretical. Hence, we apply an ordered logit model in order to determine the key drivers of this club formation. The results indicate that population growth and density as well as the supply side variable, namely the housing stock, are key drivers in determining house price club membership. A one unit increase in population growth significantly increases the probability of belonging to a higher house price club, whereas an increase in population density and housing stock decreases this probability. The result that demographic developments, especially population growth, have an impact on future house price growth is of particular interest when regarding population forecasts for German labour market regions. Forecasts by Eurostat state that regions in southern Germany are mostly expected to grow, whereas regions in eastern Germany, with the exception of Berlin and its surrounding areas, are projected to decrease regarding their population size. Furthermore, we find a positive correlation between house price growth and future population developments at the regional level indicating that many regions with high house price growth between 2007 and 2017 are expected to grow in their population size until 2030. Connecting this consideration to our result that regions with high population growth are more likely to belong to a club with high house price converging level, indicates that the demand for housing will rise even further in regions with already high house prices while the opposite holds true for regions with lower house price growth between 2007 and 2017. As a consequence, the divergence of regional house prices would become more pronounced affecting regional housing affordability. Especially low income households are affected by these locally concentrated price increases resulting in growing disposable income inequality, as well as a divergence in consumption and saving patterns across income groups (Dustmann, Fitzenberger and Zimmermann, 2018).

At the same time, while population developments impact the demand side of housing, there are other factors on the supply side. The German government pledges in its coalition agreement to build 400,000 new housing units every year, 100,000 of which are to be publicly subsidized. Measures, as such, which focus on increasing the housing stock can be an important tool to locally decrease the price pressure and facilitate movements towards a single house price convergence club for an integrated German housing market.

Appendix

Tables

Table 3.6: Club-specific summary statistics

	1	2	3	4	5	6	7
GDP	45565.0	28191.7	30037.2	28276.5	24725.6	24674.9	22181.9
	(11982.9)	(5387.5)	(7155.7)	(6018.2)	(4278.6)	(4869.7)	(3906.7)
Employment	0.549	0.463	0.477	0.464	0.437	0.435	0.425
	(0.075)	(0.064)	(0.046)	(0.044)	(0.047)	(0.036)	(0.038)
Population growth	0.009	0.005	0.002	-0.001	-0.003	-0.005	-0.007
	(0.002)	(0.002)	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)
Population density	259.0	189.1	242.2	256.5	266.2	415.5	112.1
	(145.7)	(176.1)	(168.7)	(223.9)	(301.0)	(540.3)	(37.5)
Young population	0.823	0.806	0.802	0.798	0.789	0.783	0.774
	(0.003)	(0.009)	(0.016)	(0.015)	(0.018)	(0.017)	(0.015)
Housing stock	0.213	0.251	0.234	0.242	0.257	0.235	0.270
	(0.066)	(0.048)	(0.038)	(0.036)	(0.047)	(0.045)	(0.032)

Notes: Average real GDP per capita, employment per capita, population growth and density between 2000 and 2006; housing stock per capita and share of young population (share of population at the age of 1 to 64 years of the total population) in 2007.

Table 3.7: Club membership and housing market statistics by region

Region	Club	Merged club	House price index 10/2017	Average house price growth	State
Ingolstadt	1	1	175.028	0.452	BY
Munich	1	1	172.789	0.440	BY
Augsburg	Div	2	159.742	0.371	BY
Traunstein	2	2	159.403	0.371	BY
Berlin	2	2	158.280	0.351	BE
Oberhavel	2	2	158.277	0.365	BB
Weilheim-Schongau	2	2	154.949	0.358	BY
Memmingen	3	2	154.309	0.356	BY
Altötting	3	2	153.703	0.346	BY
Landshut	2	2	153.427	0.338	BY
Regensburg	3	2	153.055	0.369	BY
Loerrach	2	2	150.454	0.310	BW
Flensburg	2	2	150.322	0.343	SH
Oldenburg	3	2	150.223	0.334	NI
Donau-Ries	3	2	147.010	0.329	BY
Havelland	2	2	143.745	0.287	BB
Märkisch-Oderland	3	2	142.567	0.219	BB
Erlangen	3	2	138.768	0.266	BY
Rostock	4	3	148.698	0.302	MV
Kempten	4	3	146.424	0.310	BY
Vechta	4	3	143.950	0.296	NI
Potsdam-Mittelmark	4	3	143.919	0.272	BB
Nuremberg	5	3	140.691	0.285	BY
Stuttgart	5	3	140.503	0.273	BW
Ravensburg	4	3	139.413	0.266	BW
Sigmaringen	5	3	138.842	0.269	BW
Luebeck	5	3	138.841	0.286	SH
Bremen	5	3	138.120	0.254	HB
Deggendorf	5	3	137.957	0.277	BY
Ulm	4	3	137.954	0.249	BW
Schwäbisch Hall	5	3	137.368	0.253	BW
Weissenburg-Gunzenhausen	4	3	136.436	0.229	BY
Nordvorpommern	5	3	135.964	0.240	MV
Teltow-Fläming	4	3	135.605	0.237	BB
Hamburg	5	3	135.306	0.244	HH
Passau	4	3	134.263	0.219	BY
Frankfurt am Main	5	3	133.788	0.235	HE
Wuerzburg	4	3	132.437	0.233	BY
Constance	5	3	132.310	0.229	BW
Kiel	5	3	132.208	0.228	SH
Reutlingen	5	3	132.094	0.219	BW
Emden	5	3	131.895	0.223	NI
Ansbach	5	3	131.763	0.214	BY
Freiburg	5	3	131.186	0.218	BW
Wolfsburg	5	3	130.954	0.218	NI
Heidenheim	5	3	130.306	0.219	BW
Landau	5	3	130.177	0.213	RP
Karlsruhe	5	3	130.045	0.217	BW
Stade	5	3	129.217	0.200	NI
Frankfurt (Oder)	5	3	129.059	0.143	BB
Aschaffenburg	5	3	128.947	0.188	BY

Region	Club	Merged club	House price index 10/2017	Average house price growth	State
Dresden	5	3	128.855	0.193	SN
Göppingen	5	3	128.756	0.193	BW
Mainz	5	3	128.555	0.202	RP
Wilhelmshaven	5	3	128.339	0.198	NI
Ostprignitz-Ruppin	5	3	127.520	0.194	BB
Ludwigshafen	5	3	127.318	0.180	RP
Hannover	5	3	127.208	0.200	NI
Leipzig	5	3	126.588	0.221	SN
Muenster	5	3	126.561	0.196	NW
Waldshut	5	3	121.895	0.186	BW
Bremerhaven	6	4	128.167	0.208	HB
Heidelberg	6	4	125.545	0.188	BW
Dithmarsen	6	4	125.445	0.198	SH
Heilbronn	6	4	125.255	0.183	BW
Celle	7	4	124.084	0.184	NI
Brunswick	6	4	123.499	0.167	NI
Schweinfurt	6	4	123.393	0.177	BY
Cologne	6	4	123.119	0.170	NW
Ortenau district	6	4	123.029	0.183	BW
Emsland	6	4	122.475	0.168	NI
Osnabruck	6	4	122.199	0.157	NI
Cham	6	4	122.000	0.155	BY
Bonn	6	4	121.691	0.166	NW
Limburg-Weilburg	7	4	120.573	0.155	HE
Düsseldorf	6	4	120.520	0.151	NW
Schwerin	6	4	120.086	0.153	MV
Darmstadt	7	4	118.913	0.142	HE
Soest	7	4	117.648	0.115	NW
Bielefeld	7	4	117.523	0.131	NW
Pforzheim	7	4	117.431	0.138	BW
Boeblingen	7	4	117.253	0.129	BW
Bayreuth	7	4	116.604	0.140	BY
Goettingen	7	4	116.240	0.154	NI
Coburg	Div	4	115.462	0.112	BY
Bamberg	6	4	114.507	0.139	BY
Mecklenburg Lake district	6	4	113.322	0.101	MV
Freyung-Grafenau	6	4	111.991	0.103	BY
Bautzen	6	4	111.886	0.089	SN
Erfurt	6	4	105.601	0.039	TH
Magdeburg	6	4	102.816	-0.010	ST
Cottbus	7	4	101.681	0.012	BB
Uelzen	8	5	117.995	0.114	NI
Koblenz	8	5	115.598	0.114	RP
Aachen	9	5	113.362	0.101	NW
Minden	9	5	112.816	0.095	NW
Zollernalb district	8	5	112.652	0.078	BW
Kassel	8	5	112.384	0.094	HE
Essen	9	5	112.273	0.094	NW
Rottweil	9	5	112.135	0.080	BW
Dortmund	9	5	111.583	0.090	NW

Region	Club	Merged club	House price index 10/2017	Average house price growth	State
Siegen	9	5	111.476	0.078	NW
Altenkirchen	9	5	111.267	0.095	RP
Borken	9	5	108.725	0.073	NW
Trier	8	5	108.377	0.089	RP
Kaiserslautern	9	5	107.830	0.068	RP
Lüchow-Dannenberg	8	5	107.492	0.087	ST
Vulkaneifel	9	5	107.069	0.049	RP
Jena	9	5	106.630	0.046	TH
Kleve	9	5	106.497	0.051	NW
Uckermark	8	5	97.921	0.010	BB
Chemnitz	9	5	93.503	-0.030	SN
Eisenach	8	5	92.511	-0.075	TH
Elbe-Elster	8	5	87.640	-0.095	BB
Stendal	8	5	84.370	-0.136	ST
Wuppertal	11	6	108.303	0.065	NW
Bochum	10	6	105.618	0.038	NW
Amberg	10	6	104.604	0.046	BY
Hagen	11	6	104.520	0.043	NW
Fulda	11	6	104.093	0.039	HE
Giessen	11	6	104.029	0.020	HE
Saarbrucken	10	6	103.542	0.027	SL
Bad Kreuznach	11	6	103.167	0.036	RP
Südvorpommern	11	6	101.313	0.023	MV
Pirmasens	11	6	100.778	0.028	RP
Olpe	11	6	100.684	0.017	NW
Goslar	10	6	96.866	-0.035	ST
Halle	11	6	83.567	-0.110	ST
Nordhausen	10	6	78.815	-0.139	TH
Gera	10	6	77.484	-0.209	TH
Hameln	12	7	98.456	-0.014	NI
Waldeck-Frankenberg	12	7	97.295	-0.011	HE
Bitburg	13	7	94.175	-0.082	RP
Kronach	12	7	92.982	-0.108	BY
Hoexter	13	7	91.892	-0.102	NW
Hof	13	7	84.461	-0.121	BY
Suhl	13	7	79.959	-0.161	TH
Unstrut-Hainich	12	7	73.217	-0.253	TH
Dessau-Rosslau	12	7	71.827	-0.237	ST
Saalfeld-Rudolstadt	13	7	67.943	-0.317	TH
Prignitz	Div	7	65.905	-0.349	BB

Notes: The average house price growth is calculated between 01/2017 and 10/2017. States: Baden-Wurttemberg—BW; Bavaria—BY; Berlin—BE; Brandenburg—BB; Bremen—HB; Hamburg—HH; Hesse—HE; Lower Saxony—NI; Mecklenburg-West Pomerania—MV; North Rhine Westfalia—NW; Rhineland Palatinate—RP; Saarland—SL; Saxony—SN; Saxony Anhalt—ST; Schleswig-Holstein—SH; Thuringia—TH

R Code

The following Section presents the R Code, which is used to apply the convergence clustering algorithm by Phillips and Sul (2007). To the best of our knowledge, it goes beyond currently existing codes or packages in Stata or R by combining the algorithm by Phillips and Sul (2007) with proposed alterations by Phillips and Sul (2009), Schnurbus, Haupt and Meier (2017), and Lyncker and Thoennessen (2017) in one code.

Main Code

```
################################################################################
###########################
################# Convergence Clustering Algorithm based on Phillips and Sul
(2007) ######################
################# Implemented by Angelina Hackmann (2021)
#################################################
################################################################################
###########################


###########Install Packages##############
#install.packages("lmodel2")
#library(lmodel2)
#install.packages("foreign")
#library(foreign)
#install.packages("lmtest")
#library(lmtest)
#install.packages("sandwich")
#library(sandwich)
#install.packages("nnet")
#library(nnet)
##install.packages("xlsx")
##library(xlsx)
#install.packages("mFilter")
#library(mFilter)
##install.packages("ConvergenceClubs")
##library(ConvergenceClubs)
##install.packages("writexl")
#library(writexl)

# clear workspace and set working directory
rm(list = ls())
setwd("C:/Users/Hackmann/Desktop/HousingConvergence/R_SicheraPizzuto")

# Include functions
source("hac_ps.r")
source("filter_order.r")
source("logt_test.r")
source("logt_test1.r")
source("logt_test2.r")


# Choose which results to replicate:
# p<-1: Phillips and Sul (Econometrica, 2007); consumer price indices (CPI's)
for 19 U.S. MSAs
#p<-1
# p<-2: Phillips and Sul (Journal of Applied Econometrics, 2009); per capita
real income for 152 countries
# p<-2
# p<-3 house price indices postal code areas Germany
#p<-3
# p<-4 house price indices Arbeitsmarktkreise Germany
p<-4
```

```
# Choose which algorithm procedure to use:
#meth<-"PS2007"       # increasing critical value c* in Step 3
#meth<-"PS2009"       # increasing critical value c* in Step 3
                      # choose PS2009 for Lyncker Thoennessen (they also use
increasing c*)
meth<-"Schnurbus"     # ordering tvalues instead of increasing c*

# Choose which merging method to use or if no merging algorithm, then
mergemeth<-0
#mergemeth<-0
mergemeth<-"Schnurbus"    #merging as in PS2009, but more than one iteration
#mergemeth<-"PS2009"
#mergemeth<-"LynckerThoennessen"       #different argument for merging

if(meth=="PS2007"){
  mergemeth<-0
}

# Choose whether to use merging algorithm for divergence group by von Lyncker,
Thoennessen
divmeth<-TRUE
#divmeth<-FALSE

# read data
if(p==1){
  # "testvorHP.csv":logarithmized price indices with first value=log(100) (see
"PhilSulDataVorbereitung.xlsx")
  dat<- read.csv("C:/Users/Hackmann/Desktop/HousingConvergence/testvorHP.csv",
header=TRUE, sep=";",
                 dec=",")
  # time period is cut into half (only second half of the time period is
considered according to PS2007
  TT<-nrow(dat)/2
  # HP-Filter for annual data, lambda=400 (according to PS2007 & 2009)
  lambda<-400
}
if(p==2){
  # "PS2009iriginal.csv":original data set-> it has to be logarithmized
  dat<-
read.csv("C:/Users/Hackmann/Desktop/HousingConvergence/R_SicheraPizzuto/PS2009o
riginal.csv",
                 header=TRUE, sep=";", dec=",")
  TT<-nrow(dat)
  dat<-log(dat)
  lambda<-400
}

if(p==3){
  dat<-
read.csv("C:/Users/Hackmann/Desktop/HousingConvergence/R_SicheraPizzuto/HedMon_
forR_Berlin.csv",
                 header=TRUE, sep=";", dec=",")
  TT<-nrow(dat)
  # HP-Filter for monthly data, lambda=14400
```

```
  lambda<-14400
}

if(p==4){
  dat<-
read.csv("C:/Users/Hackmann/Desktop/HousingConvergence/R_SicheraPizzuto/HPrealA
MK.csv",
                 header=TRUE, sep=";", dec=",")
  #delete last column (Germany)
  dat<-dat[,-ncol(dat)]
  TT<-nrow(dat)
  # HP-Filter for monthly data, lambda=14400
  lambda<-14400
}

################# STEP 1: Last Observation Ordering
#######################################################
################################################################################
###########################
# the function "filter_order" filters the observations by means of a hodrick
prescott filter with lambda
# according to the frequency of the data and the cities/countries (from now on
called regions) are ordered
# according to their last observation or mean of last third of observations
# dat: trend component of data and ordered according to last observation
(decreasing)
dat <- filter_order(dat,(nrow(dat)-TT),lambda)


################# STEP 2: Core Group Formation
############################################################
################################################################################
###########################
# First Part: Find out, whether all regions converge
tt<-round((1/3)*TT)+1   # Convergence-test is carried out for the last 2/3 of
the time period
allclubs<-numeric(TT+1) # variable will include members of all clubs
n<-1                    # number of clubs set to 1
c_max<-50               # choose which ever upper limit you want to choose for
the critical value c
c_inc<-0.5             # choose by which amount c is supposed to be increased
after each iteration
div<-FALSE

# function "logt_test" tests for convergence of all regions in dat and has
t-value of this test as a result
tv<-logt_test(dat,TT,tt,ncol(dat))

# stop, if all cities converge
if(tv > (-1.65)) stop("All regions converge.")

# while loop: make sure that algorithm stops, when no cities are left behind
(number of columns of dataset dat,
# which contains the remaining cities after each iteration of algorithm, has to
```

```
be larger than 0)

while(ncol(dat)>2){

  # Second Part: Form Core Group
  # create empty vectors and matrices for upcoming calculations
  I<-numeric(ncol(dat))
  J<-numeric(ncol(dat))
  Q<-numeric(ncol(dat))
  F<-matrix(nrow=TT,ncol=ncol(dat))
  tv<-numeric(ncol(dat))

  # special for PS2007, they only use last third of time series for algorithm
  if(p==1){
    tt<-round(0.6*nrow(dat)+1)
  }

  # the function "logt_test1" adds one region at a time and calculates the
t-value of the accumulated
  # regions (starts with first 2 regions, then first 3 regions, and so on (last
t-value: all regions))
  X<-dat[,1]
  tv<-logt_test1(dat,TT,tt,X,ncol(dat))

  for(i in 1:ncol(dat)){
    I[i]<- -1.65
  }

  for(i in 2:ncol(dat)){
    if(tv[i]<= (-1.65)){
      break
    }
    I[i]<-tv[i]           # I includes all tvalues as long as tv>-1.65,
otherwise it has the value -1.65
  }

  # if first two regions directly lead to tvalue< -1.65, then first region will
be dropped and the
  # log t test is figured out starting with the second region that is now the
first column in dat
  # if these lead once again to tvalue< -1.65 for first two regions, then first
region is dropped again,
  # until first two regions converge (tvalue> -1.65)
  o<-2
  while(o<ncol(dat)){
    if(tv[2]<= (-1.65)){
      dd<-dat[,1]           # name and values of first region are saved
      nam<-colnames(dat[1])
      dat<-dat[,-1]         # first region is dropped from dataset

      # create empty vectors and matrices for upcoming calculations
      R<-numeric(ncol(dat))
      I<-numeric(ncol(dat))
      J<-numeric(ncol(dat))
```

```
      Q<-numeric(ncol(dat))
      F<-matrix(nrow=TT,ncol=ncol(dat))
      tv<-numeric(ncol(dat))

      # the function "logt_test1" adds one region at a time and calculates the
t-value of the accumulated
      # regions
      X<-dat[,1]
      tv<-logt_test1(dat,TT,tt,X,ncol(dat))

      for(i in 1:ncol(dat)){
        I[i]<- -1.65
      }

      for(i in 2:ncol(dat)){
        if(tv[i]<= (-1.65)){
          break
        }
        I[i]<-tv[i]         # I includes all tvalues as long as tv>-1.65,
otherwise it has the value -1.65
      }

      o<-o+1
      dat<-cbind(dat,dd)
      names<-numeric(ncol(dat))
      names[1:(ncol(dat)-1)]<-colnames(dat[1:(ncol(dat)-1)])
      names[ncol(dat)]<-nam
      colnames(dat)<-names    # values and name of dropped column is added as
last column of dataset
                              # this region can still be chosen as cluster
candidate later
    }else{
      o<-ncol(dat)            # leave while loop, when two converging regions
are found
    }
  }

  m<-which.is.max(I)        # m is the row of I (=number of region/column of
dat),
                            # for which the tvalue is maximal

  for(i in 1:ncol(dat)){
    if(i<=m){
      I[i]<-1               # I=1 from 1st to mth entry and 0 otherwise (for
core group I=1)
    }else{
      I[i]<-0
    }
  }

  ll<-TRUE
  mm<-TRUE
  d<-data.frame(dat)
```

```
################ STEP 3:Sieve Individuals for Club Membership
############################################
################################################################################
###########################
  while(ll==TRUE){
    J<-numeric(ncol(d))
    X<-numeric(TT)

    # members of core group (entries 1 to m) are addded, set J=1 for the
entries 1 to m (for core group)
    for(j in 1:m){
      for(t in 1:TT){
        X[t]<- X[t] + d[t,j]
        J[j]<-1
      }
    }

    pr<-sprintf("Core Group %i includes",n)
    print(pr)
    for(i in 1:ncol(d)){
      if(J[i]!=0){
        print(colnames(d)[i])
      }
    }

    c<-0
    tvn<-numeric(ncol(d))
    tv<-numeric(ncol(d))
    # the function "logt_test2" adds each region larger than m seperately to
the core gruop and
    # convergence test is carried out with core group plus this one additional
region
    # if tvalue with the additional region is larger than critical value c,
this region is a cluster
    # candidate
    tvn<-logt_test2(d,TT,tt,X,m)
    # for methods PS 2007 & 2009, the critical value is increased after each
iteration until core group +
    # cluster candidates have together tvalue > -1.65

    while(c< c_max){
      # I=1 for each region i for which tv[i]> critical value c and otherwise
it is 0
      # I=1 for the core group and all cluster candidates
      I<-numeric(ncol(d))

      for(i in 1:m){
        I[i]<-1
      }

      for(i in (m+1):ncol(d)){
        if(tvn[i]> c){
          I[i]<-1
        }else{
```

```
          I[i]<-0
        }
      }

      A<-0
      A<-sum(I)-m     # A is number of cluster candidates

      # test whether tv>-1,65 still holds when all cluster candidates are added
to core group
      dn<-data.frame(d)
      for(j in 1:ncol(d)){
        for(t in 1:TT){
          dn[t,j]<-0
        }
      }

      # cities of core group (cities 1 to m) are written in matrix dn in the
columns 1 to m
      for(j in 1:m){
        for(t in 1:TT){
          dn[t,j]<-d[t,j]
        }
      }
      j<-j+1

      # each city for which I=1 (from m+1 onwards), so all the cluster
candidates are written in matrix dn
      # in the columns m+1 onwards
      for(i in (m+1):ncol(d)){
        if(I[i]!=0){            # if I=1, column i of d is written into column
j of dn, then j<-j+1
          for(t in 1:TT){
            dn[t,j]<-d[t,i]
          }
          j<-j+1
        }else{dn[t,j]<-dn[t,j]+0}  # if I=0 the column of dn stays 0 and we use
it again for the next city i
      }
      dn<-dn[,-((m+A+1):ncol(dn))]  # all columns wihtout entries are deleted

      # now dn includes core group (columns 1 to m) and cluster candidates
(columns (m+1)to(m+A))
      # convergence test is carried out for all the cities from core group and
all cluster candidates
      tv<-logt_test(dn,TT,tt,(m+A))
      # cluster is found, if tv>-1.65
      rr<-FALSE

      # if tv<-1.65, then differentiate between method used: Schnurbus or
Phillips and Sul
      if(meth=="Schnurbus"){
        if(tv<= (-1.65)){
          rr<-TRUE
          while(rr==TRUE){
```

```
            # keep core group (1:m) and add other regions according to a
decreasing tvalue from logt_test2
            tm<-NULL
            tm<-which.is.max(tvn[(m+1):ncol(dn)])
            tm<-tm+m
            if(tvn[tm]<= (-1.65)){break}    # if max tvalue< -1.65, then
cluster includes only core group

            datn<-data.frame(dn[,tm])       # add region with max tvalue to
core group (add as column m+1)
            colnames(datn)<-colnames(dn[tm])
            dn<-dn[,-tm]
            dn<-cbind(dn[,1:m],datn,dn[,(m+1):ncol(dn)])
            m<-m+1
            d<-data.frame(dn)
            mm<-FALSE
            J<-numeric(ncol(d))
            X<-numeric(TT)

            # the entries 1 to m are added because they are the core group, set
J=1 for the entries 1 to m
            # (for core group), which now includes region with max tvalue
            for(j in 1:m){
              for(t in 1:TT){
                X[t]<- X[t] + d[t,j]
                J[j]<-1
              }
            }

            tvn<-numeric(ncol(d))
            tv<-numeric(ncol(d))
            # the function "logt_test2" adds each region larger than m
seperately to the core gruop and
            # convergence test is carried out with core group plus this one
additional region
            # if tvalue with the additional region is larger than critical
value, this region is a
            # cluster candidate
            tvn<-logt_test2(d,TT,tt,X,m)
          }

          c<-c_max        # make sure that while loop stops and cluster is
found
          ll<-FALSE
        }else{
          c<-c_max
          ll<-FALSE
        }

      }else{
        # if method is Phillips and Sul and if tvalue of core group + cluster
candidates is < -1.65, then
        # increase critical value c by c_inc; if tvalue> -1.65, then leave
loop, cluster is found
```

```
        if(tv> (-1.65)){
          c<-c_max
          ll<-FALSE
        }else{
          c<- c+c_inc
        }
      }
    }
  }

  # if before method Schnurbus and cluster candidates were ordered according to
max tvalue, then rr==TRUE
  # and some data adjustments have to be made in order to print cluster members
  if(rr==TRUE){
    A<-0
    d<-d[,-((m+1):ncol(d))]
    d<-d[,order(colnames(d),decreasing=FALSE)]
    dat<-dat[,order(colnames(dat),decreasing=FALSE)]
    I<-numeric(ncol(dat))
    j<-1

    for(i in 1:ncol(dat)){
      if(colnames(dat[i])==colnames(d[j])){
        I[i]<-1
        j<-j+1
        if(j==(ncol(d)+1)){break}
      }else{
        I[i]<-0
      }
    }
  }

  # print cluster members
  pr<-sprintf("Cluster %i includes",n)
  print(pr)
  j<-1
  for(i in 1:ncol(dat)){
    if(I[i]!=0){
      print(colnames(dat)[i])
      j<-j+1
    }
  }
  # print number of cluster members (j-1)
  pr<-sprintf("Cluster %i includes %i members",n,(j-1))
  print(pr)

################# Step 4: Stopping Rule
###################################################################
################################################################################
###########################
  # Convergence test for all regions, which are not included in clusters yet to
see whether they build
  # cluster themselves
  tv <-numeric(ncol(dat)-m-A)
```

```
  d<-matrix(0,nrow=TT,ncol=ncol(dat)-m-A)
  j<-1
  # loop goes through I starting with m+1 (not going through core group) and
writes in matrix d all
  # regions, which are not in the cluster (for which I=0)
  for(i in (m+1):ncol(dat)){
    if(I[i]==0){
      for(t in 1:TT){
        d[t,j]<-d[t,j]+dat[t,i]
      }
      j<-j+1
    }
  }

  # write all the members of the found cluster n in the matrix club
  club<-matrix(0,nrow=TT+1,ncol=m+A)
  memnames<-matrix(0,1,ncol=m+A)
  j<-1
  for(i in 1:ncol(dat)){
    if(I[i]==1){
      for(t in 1:TT){
        club[t,j]<-dat[t,i]
        memnames[j]<-colnames(dat[i])
      }
      club[TT+1,j]<-n
      j<-j+1
    }
  }

  # add club to matrix allclubs, which keeps all the clubs so far
  colnames(club)<-memnames
  allclubs<-cbind(allclubs,club)

  # if only 1 region is left, it is a divergent region
  if(ncol(d) < 2){
    memnames<-NULL
    for(i in 1:ncol(dat)){
      if(I[i]==0){
        dname<-colnames(dat[i]) # save names of the one divergent region
      }
    }
    colnames(d)<-dname
    dat<-d

  # if d includes more than 1 region, test whether all regions in d converge
  }else{
    # logt_test for all regions which are not in cluster, if tvalue>-1.65, then
last cluster is found
    tv<-logt_test(d,TT,tt,ncol(d))

    # matrix F includes all regions from dat, which are not in cluster and NA
for cluster members
    if(tv> -1.65){break}else{
      F<-matrix(nrow=TT,ncol=ncol(dat))
```

```
      for(i in 1:ncol(dat)){
        if(I[i]==0){
          F[,i]<-dat[,i]
        }else{F[,i]=NA}
      }

      # dat includes all regions, which are not in cluster, columns with NA are
deleted
      colnames(F)<-colnames(dat)
      dat<-F[,colSums(is.na(F)) != nrow(F)]
      dat<-data.frame(dat)
      dat<-dat[,order(dat[nrow(dat),],decreasing=TRUE)]
    }

    if(ncol(dat)==ncol(F)){break}

    # if dat is left with two regions, then test whether they diverge, if yes,
set div=TRUE
    if(ncol(dat)==2){
      tv<-logt_test(dat,TT,tt,ncol(dat))
      if(tv<= (-1.65)){
        div<-TRUE
      }
    }
    n<-n+1                              # increase number of cluster n for
next iteration
  }
}

if(ncol(d) < 2){
  div<-TRUE
}

j<-1
n<-n+1
club<-matrix(0,nrow=TT+1,ncol=ncol(d))
memnames<-matrix(0,1,ncol=ncol(d))
# if tvalue> -1.65, then last cluster is found
if((tv > (-1.65)) && (div==FALSE)){
  pr<-sprintf("Cluster %i includes",n)
  print(pr)
  for(i in 1:ncol(dat)){
    if(I[i]==0){
      print(colnames(dat)[i])
      for(t in 1:TT){
        club[t,j]<-dat[t,i]
        memnames[j]<-colnames(dat[i])
      }
      club[TT+1,j]<-n
      j<-j+1
    }
  }
  pr<-sprintf("Cluster %i includes %i members",n,(j-1))
  print(pr)
```

```
}else{
  # if tvalue< -1.65, then the regions do not converge and build a divergence
group
  print("the non-convergence group includes")
  if(ncol(d) < 2){
    print(dname)
    pr<-sprintf("The non-convergence group includes 1 member")
    print(pr)
  }else{
    for(i in 1:ncol(dat)){
      print(colnames(dat)[i])
      for(t in 1:TT){
        club[t,j]<-dat[t,i]
        memnames[j]<-colnames(dat[i])
      }
      club[TT+1,j]<-0           # last line in allclubs=0 if divergence group
      j<-j+1

    }
    pr<-sprintf("The non-convergence group includes %i members",(j-1))
    print(pr)
  }
}

if(ncol(d)< 2){
  dat<-data.frame(rbind(dat,0))
  allclubs<-cbind(allclubs,dat) #add last cluster or divergence group to matrix
allclubs
}else{
  colnames(club)<-memnames
  allclubs<-cbind(allclubs,club) #add last cluster or divergence group to
matrix allclubs
}

# save results in matrix "result", first row before and second row after
merging
allclubs<-allclubs[,-1]
if(meth=="PS2007"){
  #allclubs<-allclubs[,-1]
  mergeclubs<-matrix(NA,(TT+1),ncol(allclubs))
}
result<- matrix(0,2,ncol(allclubs))
result<- data.frame(result)
colnames(result)<-colnames(allclubs)

for(i in 1:ncol(allclubs)){
  result[1,i]<-allclubs[(TT+1),i]
}

################# Step 5: Merging Algorithm
###############################################################
###############################################################################
###########################
```

```
# 5.1 Merging Algorithm by Phillips and Sul (2009)
##########################################################
# Run the log t regression for all pairs of subsequent initial clubs and find
out whether they fulfill the
# convergence hypothesis
if(mergemeth=="PS2009"){
  # number of clubs is listed in last line of allclubs, first column is deleted
(includes only zeroes)
  #allclubs<-data.frame(allclubs[,-1])
  allclubs<-data.frame(allclubs)
  mergeclubs<-allclubs
  n<-1
  N<-max(allclubs[TT+1,])
  tv<-numeric(N-1)
  m<-1

  # while loop goes through number of clusters
  while(n<N){
    j<-1
    conclub<-matrix(NA,TT,ncol(allclubs))
    names<-matrix(NA,1,ncol(allclubs))
    # add members of cluster n and n+1 to conclub, for which convergence test
is figured out
    for(i in 1:ncol(allclubs)){
      if(allclubs[(TT+1),i]==n || allclubs[(TT+1),i]==(n+1)){
        for(t in 1:TT){
          conclub[t,j]<-allclubs[t,i]
        }
        names[j]<-colnames(allclubs[i])
      }
      j<-j+1
    }

    conclub<-data.frame(conclub)
    colnames(conclub)<-names[1:ncol(conclub)]
    conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

    # logt_test is figured out for members of cluster n and n+1
    # if tvalue>-1.65, then the two clusters can be merged
    tv[n]<-logt_test(conclub,TT,tt,ncol(conclub))
    for(i in 1:ncol(allclubs)){
      if(allclubs[(TT+1),i]==n){
        mergeclubs[(TT+1),i]<-m
      }
    }

    # members of club n and n+1 are saved as club m in matrix mergeclubs (which
includes all clubs)
    if(tv[n]> (-1.65)){
      for(i in 1:ncol(allclubs)){
        if(allclubs[(TT+1),i]==n+1){
          mergeclubs[(TT+1),i]<-m
        }
```

```
        if(allclubs[(TT+1),i] > (n+1)){
          mergeclubs[(TT+1),i]<-as.integer(mergeclubs[(TT+1),i])-1
        }
      }

      #matrix conclub includes members of club n and n+1
      conclub<-matrix(NA,TT,ncol(allclubs))
      names<-matrix(NA,1,ncol(allclubs))
      j<-1
      for(i in 1:ncol(allclubs)){
        if(mergeclubs[(TT+1),i]==m){
          for(t in 1:TT){
            conclub[t,j]<-mergeclubs[t,i]
          }
          names[j]<-colnames(mergeclubs[i])
        }
        j<-j+1
      }

      conclub<-data.frame(conclub)
      colnames(conclub)<-names[1:ncol(conclub)]
      conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

      # logt_test is figured out for members of cluster n and n+1
      # if tvalue>-1.65, then the two clusters can be merged
      tvn<-NULL
      tvn<-logt_test(conclub,TT,tt,ncol(conclub))

      if(tvn> (-1.65)){
        pr<-sprintf("Clubs %i and %i can be merged to a new convergence club.",
n, (n+1))
        print(pr)
        m<-m-1
      }else{
        for(i in 1:ncol(allclubs)){
          if(allclubs[(TT+1),i]==n+1){
            mergeclubs[(TT+1),i]<-m+1
          }
          if(allclubs[(TT+1),i] > (n+1)){
            mergeclubs[(TT+1),i]<-as.integer(mergeclubs[(TT+1),i])+1
          }
        }
      }
    }
    m<-m+1
    n<-n+1
  }

  # save last club in mergeclubs as club m
  if(n==N){
    for(i in 1:ncol(allclubs)){
      if(allclubs[(TT+1),i]==n){
        mergeclubs[(TT+1),i]<-m
      }
```

```
    }
  }

  N<-max(mergeclubs[TT+1,])
  m<-1
  # loop goes through number of merged clusters N
  while(m <= N){
    pr<-sprintf("After merging, Cluster %i includes",m)
    print(pr)
    j<-1
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==m){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }

    pr<-sprintf("Cluster %i includes %i members",m,(j-1))
    print(pr)
    m<-m+1
  }

  if(div==TRUE){
    pr<-sprintf("After merging, the non-convergence group includes")
    print(pr)
    j<-1
    for(i in 1:ncol(mergeclubs)){
      if(allclubs[(TT+1),i]==0){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }

    pr<-sprintf("The non-convergence group includes %i members",(j-1))
    print(pr)
  }
}


# 5.2 Merging Algorithm by Schnurbus et al. (2017)
##########################################################
# Run the log t regression for all pairs of subsequent initial clubs and find
out whether they fulfill the
# convergence hypothesis
# run more than 1 iteration until no clubs can be merged anymore
if(mergemeth=="Schnurbus"){
  # number of clubs is listed in last line of allclubs, first column is deleted
(includes only zeroes)
  #allclubs<-data.frame(allclubs[,-1])
  allclubs<-data.frame(allclubs)
  mergeclubs<-allclubs
  k<-TRUE

  while(k==TRUE){
```

```
    n<-1
    N<-max(allclubs[TT+1,])
    tv<-numeric(N-1)
    m<-1

    # loop goes through number of clusters
    while(n<N){
      j<-1
      # add members of cluster n and n+1 to conclub, for which convergence test
is figured out
      conclub<-matrix(NA,TT,ncol(allclubs))
      names<-matrix(NA,1,ncol(allclubs))
      for(i in 1:ncol(allclubs)){
        if(allclubs[(TT+1),i]==n || allclubs[(TT+1),i]==(n+1)){
          for(t in 1:TT){
            conclub[t,j]<-allclubs[t,i]
          }
          names[j]<-colnames(allclubs[i])
        }
        j<-j+1
      }

      conclub<-data.frame(conclub)
      colnames(conclub)<-names[1:ncol(conclub)]
      conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

      # logt_test is figured out for members of cluster n and n+1
      # if tvalue>-1.65, then the two clusters can be merged and saved in
mergeclubs as club m
      tv[n]<-logt_test(conclub,TT,tt,ncol(conclub))
      for(i in 1:ncol(allclubs)){
        if(allclubs[(TT+1),i]==n){
          mergeclubs[(TT+1),i]<-m
        }
      }

      if(tv[n]> (-1.65)){
        for(i in 1:ncol(allclubs)){
          if(allclubs[(TT+1),i]==n+1){
            mergeclubs[(TT+1),i]<-m
          }
          if(allclubs[(TT+1),i] > (n+1)){
            mergeclubs[(TT+1),i]<-as.integer(mergeclubs[(TT+1),i])-1
          }
        }

        # test whether new group still converges
        conclub<-matrix(NA,TT,ncol(allclubs))
        names<-matrix(NA,1,ncol(allclubs))
        j<-1
        for(i in 1:ncol(allclubs)){
          if(mergeclubs[(TT+1),i]==m){
            for(t in 1:TT){
              conclub[t,j]<-mergeclubs[t,i]
```

```
            }
            names[j]<-colnames(mergeclubs[i])
          }
          j<-j+1
        }

        conclub<-data.frame(conclub)
        colnames(conclub)<-names[1:ncol(conclub)]
        conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

        # logt_test is figured out for members of cluster n and n+1
        # if tvalue>-1.65, then the two clusters can be merged
        tvn<-NULL
        tvn<-logt_test(conclub,TT,tt,ncol(conclub))

        if(tvn> (-1.65)){
          pr<-sprintf("Clubs %i and %i can be merged to a new convergence
club.", n, (n+1))
          print(pr)
          m<-m-1
        }else{
          for(i in 1:ncol(allclubs)){
            if(allclubs[(TT+1),i]==n+1){
              mergeclubs[(TT+1),i]<-m+1
            }
            if(allclubs[(TT+1),i] > (n+1)){
              mergeclubs[(TT+1),i]<-as.integer(mergeclubs[(TT+1),i])+1
            }
          }
        }
      }
      m<-m+1
      n<-n+1
    }

    p<-as.integer(allclubs[(TT+1),])
    pp<-as.integer(mergeclubs[(TT+1),])

    # if no more clubs can be merged (clubs in mergeclubs and allclubs are the
same), then leave loop
    if(sum(pp-p)==0){
      k<-FALSE
    }
    allclubs<-mergeclubs
  }

  if(n==N){
    for(i in 1:ncol(allclubs)){
      if(allclubs[(TT+1),i]==n){
        mergeclubs[(TT+1),i]<-m
      }
    }
  }
```

```
  N<-max(mergeclubs[TT+1,])
  m<-1

  # loop goes through number of clusters
  while(m <= N){
    j<-1
    pr<-sprintf("After merging, Cluster %i includes",m)
    print(pr)
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==m){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }

    pr<-sprintf("Cluster %i includes %i members",m,(j-1))
    print(pr)
    m<-m+1
  }

  if(div==TRUE){
    j<-1
    pr<-sprintf("After merging, the non-convergence group includes")
    print(pr)
    for(i in 1:ncol(mergeclubs)){
      if(allclubs[(TT+1),i]==0){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }

    pr<-sprintf("The non-convergence group includes %i members",(j-1))
    print(pr)
  }
}


# 5.3 Merging Algorithm by von Lyncker, Thoennessen (2017)
#######################################################
# if tvalue>-1.65 and tv[n]>tv[n+1], then the two clusters can be merged, new
cluster m is used again to
# find out whether it merges with the next cluster n+2
if(mergemeth=="LynckerThoennessen"){
  #allclubs<-data.frame(allclubs[,-1])
  allclubs<-data.frame(allclubs)
  mergeclubs<-allclubs
  n<-1
  N<-max(allclubs[TT+1,])
  tv<-numeric(N-1)
  # loop goes through number of clusters
  while(n<N){
    j<-1
    # add members of cluster n and n+1 to conclub, for which convergence test
is figured out
```

```
    conclub<-matrix(NA,TT,ncol(mergeclubs))
    names<-matrix(NA,1,ncol(mergeclubs))
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==n || mergeclubs[(TT+1),i]==(n+1)){
        for(t in 1:TT){
          conclub[t,j]<-mergeclubs[t,i]
        }
        names[j]<-colnames(mergeclubs[i])
      }
      j<-j+1
    }

    conclub<-data.frame(conclub)
    colnames(conclub)<-names[1:ncol(conclub)]
    conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

    j<-1
    # add members of cluster n+1 and n+2 to conclub1, for which convergence
test is figured out
    conclub1<-matrix(NA,TT,ncol(mergeclubs))
    names1<-matrix(NA,1,ncol(mergeclubs))
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==(n+1) || mergeclubs[(TT+1),i]==(n+2)){
        for(t in 1:TT){
          conclub1[t,j]<-mergeclubs[t,i]
        }
        names1[j]<-colnames(mergeclubs[i])
      }
      j<-j+1
    }
    conclub1<-data.frame(conclub1)
    colnames(conclub1)<-names1[1:ncol(conclub1)]
    conclub1<-data.frame(conclub1[,colSums(is.na(conclub1)) != nrow(conclub1)])

    # logt_test is figured out for members of cluster n and n+1
    # logt_test is figured out for members of cluster n+1 and n+2
    # if tvalue>-1.65 and tv[n]>tv[n+1], then the two clusters can be merged
    tv[n]<-logt_test(conclub,TT,tt,ncol(conclub))
    tv[n+1]<-logt_test(conclub1,TT,tt,ncol(conclub1))
    if(tv[n]> (-1.65) && tv[n]>tv[n+1]){
      pr<-sprintf("Clubs %i and %i can be merged to a new convergence club.",
n, (n+1))
      print(pr)
      for(i in 1:ncol(mergeclubs)){
        if(mergeclubs[(TT+1),i]==n+1){
          mergeclubs[(TT+1),i]<-n
        }
        if(mergeclubs[(TT+1),i] > (n+1)){
          mergeclubs[(TT+1),i]<-as.integer(mergeclubs[(TT+1),i])-1
        }
      }

      n<-n-1
    }else{
```

```
      n<-n
    }
    n<-n+1
    N<-max(mergeclubs[TT+1,])
  }

  m<-1
  # loop goes through number of clusters
  while(m <= N){
    j<-1
    pr<-sprintf("After merging, Cluster %i includes",m)
    print(pr)
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==m){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }
    pr<-sprintf("Cluster %i includes %i members",m,(j-1))
    print(pr)
    m<-m+1
  }

  if(div==TRUE){
    j<-1
    pr<-sprintf("After merging, the non-convergence group includes")
    print(pr)
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==0){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
    }
    pr<-sprintf("The non-convergence group includes %i members",(j-1))
    print(pr)
  }
}


################# Step 6: Merging Algorithm for Divergence group (von Lyncker,
Thoennessen) ###############
################################################################################
###########################
# each member of divergence group is separately added to each convergence group
to see whether it
# converges with on of the clusters
if(divmeth==TRUE){
  if(div==TRUE){
    k<-TRUE
    while(k==TRUE){
      N<-max(mergeclubs[TT+1,])
      j<-1
      nD<-0
```

```
      for(i in 1:ncol(mergeclubs)){
        if(as.integer(mergeclubs[(TT+1),i])==0){
          nD<-nD+1
        }
      }

      # matrix D includes all members of divergence goup (for which
mergeclubs[TT+1,i]=0)
      D<-matrix(nrow=nrow(mergeclubs),ncol=nD)
      dnames<-matrix(nrow=1,ncol=nD)
      j<-1
      for(i in 1:ncol(mergeclubs)){
        if(as.integer(mergeclubs[(TT+1),i])==0){
          D[,j]<-mergeclubs[,i]
          dnames[j]<-colnames(mergeclubs[i])
          j<-j+1
        }

      }

      D<-D[-(TT+1),]
      if(is.null(ncol(D))){
        dnames<-NULL
        j<-1
        for(i in 1:ncol(mergeclubs)){
          if(as.integer(mergeclubs[(TT+1),i])==0){
            dnames<-colnames(mergeclubs[i])
            j<-j+1
          }
        }
        D<-data.frame(D)
        names(D)<-dnames

      }else{
        colnames(D)<-dnames[1:ncol(D)]
        D<-data.frame(D)
        tv<-logt_test(D,TT,tt,ncol(D))
        if(tv> (-1.65)){
          break
        }
      }

      # go through all clubs n and test for convergence with divergence group D

      tv<-matrix(0,ncol=ncol(D),nrow=N)
      n<-1
      while(n <= N){
        # conclub includes members of club n
        conclub<-matrix(NA,TT,ncol(mergeclubs))
        names<-matrix(NA,1,ncol(mergeclubs))
        for(i in 1:ncol(mergeclubs)){
          if(mergeclubs[(TT+1),i]==n){
            for(t in 1:TT){
              conclub[t,j]<-mergeclubs[t,i]
```

```
          }
          names[j]<-colnames(mergeclubs[i])
        }
        j<-j+1
      }

      conclub<-data.frame(conclub)
      colnames(conclub)<-names[1:ncol(conclub)]
      conclub<-data.frame(conclub[,colSums(is.na(conclub)) != nrow(conclub)])

      # tv includes tvalues for each combination of clubs n and members of D
      for(j in 1:ncol(D)){
        DD<-cbind.data.frame(conclub,D[,j])
        tv[n,j]<-logt_test(DD,TT,tt,ncol(DD))
      }
      n<-n+1
    }

    # if max tvalue> -1.65, then this member j of D is added to club n in
matrix mergeclubs
    tvm<- which(tv == max(tv), arr.ind = TRUE)
    if(ncol(D)==1){
      if(tv[tvm[1]] > (-1.65)){
        for(i in 1:ncol(mergeclubs)){
          if(colnames(mergeclubs[i])==names(D)){
            mergeclubs[(TT+1),i]<-tvm[1]
          }
        }
      }
      k<-FALSE
    }

    if(tv[tvm[1],tvm[2]] > (-1.65)){
      for(i in 1:ncol(mergeclubs)){
        if(colnames(mergeclubs[i])==colnames(D[tvm[2]])){
          mergeclubs[(TT+1),i]<-tvm[1]
        }
      }
    }else{
      k<-FALSE
    }
  }
  m<-1

  # loop goes through number of clusters
  while(m <= N){
    j<-1
    pr<-sprintf("After merging, Cluster %i includes",m)
    print(pr)
    for(i in 1:ncol(mergeclubs)){
      if(mergeclubs[(TT+1),i]==m){
        print(colnames(mergeclubs)[i])
        j<-j+1
      }
```

```
      }
      pr<-sprintf("Cluster %i includes %i members",m,(j-1))
      print(pr)
      m<-m+1
    }
    if(div==TRUE){
      pr<-sprintf("After merging, the non-convergence group includes")
      print(pr)
      j<-1
      for(i in 1:ncol(mergeclubs)){
        if(mergeclubs[(TT+1),i]==0){
          print(colnames(mergeclubs)[i])
          j<-j+1
        }
      }
      pr<-sprintf("The non-convergence group includes %i members",(j-1))
      print(pr)
    }
  }else{
    pr<-sprintf("There is no group of diverging regions, which the algorithm in
Step 6 can be applied to.")
    print(pr)
  }
}

# save results in matrix "result", first row before and second row after
merging
for(i in 1:ncol(allclubs)){
  result[2,i]<-mergeclubs[(TT+1),i]
}
write_xlsx(result,"C:/Users/Hackmann/Desktop/HousingConvergence/resultClubs.xls
x")
```

Function: filter and order data

```
filter_order <- function(x,y,z){
  # - the observations are filtered by means of a hodrick prescott filter with
a lamda according to the frequency
  #   of the data (400 for yearly data)
##the following lines are excluded only for HedMon_forR_Berlin.csv
hp.t<-matrix(nrow=nrow(x),ncol=ncol(x))
for(i in 1:ncol(x))
{
  hp<-hpfilter(x[,i], freq=z)
  hp.t[,i]<-hp$trend
}
##the cities are ordered according to their last observation or mean of last
third of observations
hp.t<-data.frame(hp.t)
colnames(hp.t)<-colnames(x)
#hp.t<-x
hp.t<-hp.t[,order(hp.t[nrow(hp.t),],decreasing=TRUE)]
dat<-matrix(nrow=y,ncol=ncol(hp.t))
dat<-hp.t[(y+1):nrow(hp.t),]
colnames(x)<-colnames(hp.t)

return(dat)
}
```

Function: Log t test (1)

```
logt_test <- function(x,y,z,xc){

  # for-loops to compute relative transition parameters h and cross-sectional
variance H
  h<-matrix(nrow=y,ncol=xc)
  g<-matrix(nrow=y,ncol=xc)
  k<-numeric(y)
  HH<-numeric(y)
  H<-numeric(y)
  y1 <-NA
  x1 <-NA
  xd<-x[,1]

  for(i in 2:xc){
    for(t in 1:y){
      xd[t]<- xd[t] + x[t,i]
    }
  }

  for(t in 1:y){
    for(i in 1:xc){
      h[t,i]<- x[t,i] / ((1/xc) * xd[t])
      g[t,i]<-(h[t,i]-1)^2
      k[t]<-k[t]+g[t,i]
    }
    H[t]<-(1/xc) *k[t]
    HH[t]<-H[1]/H[t]
  }

  # dependent variable (x1) and independent variable (y1) for linear regression
are computed
  n<-1
  for(l in z:y){
    y1[n]<-log(HH[l])-(2*log(log(l)))
    x1[n]<-log(l)
    n<-n+1
  }

  # linear regression, resulting tvalue is returned
  fm<-lm(y1~x1)
  mm<-model.matrix(fm)
  r<-residuals(fm)
  #calculate HAC standard errors (with function "hac_ps.R" which is constructed
as Gauss Code by Phillips and Sul (2007))
  hac<-hac_ps(r)
  se<-sqrt(diag(solve(t(mm) %*% mm))[2]*hac)
  b<-coef(fm)[2]
  tv<-b/se
  alpha<-(1/2)*b
  return(tv)
}
```

Function: Log t test (2)

```
logt_test1 <- function(x,y,z,xd,xc){
  tv<- numeric(xc)
  h<-matrix(nrow=y,ncol=xc)
  k<-numeric(y)
  HH<-matrix(0,nrow=y,ncol=xc)
  H<-matrix(0,nrow=y,ncol=xc)
  g<-matrix(0,nrow=y,ncol=xc)
  y1 <-NA
  x1 <-NA

  for(i in 2:xc){
    h<-matrix(nrow=y,ncol=xc)
    g<-matrix(nrow=y,ncol=xc)
    k<-rep(0,y)

    for(t in 1:y){
      xd[t]<- xd[t] + x[t,i]

      for(j in 1:i){
        h[t,j]<- x[t,j] / ((1/i) * xd[t])
        g[t,j]<-(h[t,j]-1)^2
        k[t]<-g[t,j] + k[t]
      }

      H[t,i]<-(1/i) *(k[t])
      HH[t,i]<-H[1,i]/H[t,i]
    }

    # dependent variable (x1) and independent variable (y1) for linear
regression are computed
    for(l in z:y){
      y1[l]<-log(HH[l,i])-(2*log(log(l)))
      x1[l]<-log(l)
    }

    # linear regression, resulting tvalue is returned
    fm<-lm(y1~x1)
    mm<-model.matrix(fm)
    r<-residuals(fm)
    #calculate HAC standard errors (with function "hac_ps.R" which is
constructed as Gauss Code by Phillips and Sul (2007))
    hac<-hac_ps(r)
    se<-sqrt(diag(solve(t(mm) %*% mm))[2]*hac)
    b <- coef(fm)[2]
    tv[i] <- b/se
    alpha<-(1/2)*b
  }
  return(tv)
}
```

Function: Log t test (3)

```
logt_test2 <- function(x,y,z,xd,xb){

  xc<-ncol(x)
  hh<-matrix(nrow=y,ncol=xc)
  gg<-matrix(nrow=y,ncol=xc)
  y1<-NA
  x1<-NA
 # x<-dat
#  y<-TT
#  z<-tt
#  xd<-X
#  xb<-m
  tv<-NA

  for(i in (xb+1):xc){



    XX<-numeric(y)
    g<-numeric(y)
    h<-matrix(nrow=y,ncol=xc)
    HH <-numeric(y)
    H <-numeric(y)

    # in this for-loop city i (and only city i) is added to the cities 1 to m
    for(t in 1:y){
      XX[t]<-xd[t]+x[t,i]
    }

    # in this for-loop H_t is calculated as a sum of the cities 1 to m
    for(j in 1:xb){
      for(t in 1:y){
        h[t,j]<- x[t,j] / ((1/(xb+1)) * XX[t])
        g[t]<-g[t] + (h[t,j]-1)^2
      }
    }

    # in this for-loop H_t is calculated for the cities 1 to m plus city i
    for(t in 1:y){
      hh[t,i]<- x[t,i] / ((1/(xb+1)) * XX[t])
      gg[t,i]<-g[t] + (hh[t,i]-1)^2
      H[t]<-(1/(xb+1)) *(gg[t,i])
      HH[t]<-H[1]/H[t]
    }

    # dependent variable (x1) and independent variable (y1) for linear
regression are computed
    for(l in z:y){
      y1[l]<-log(HH[l])-(2*log(log(l)))
      x1[l]<-log(l)
    }

    # linear regression, resulting tvalue is returned
    fm<-lm(y1~x1)
```

```
    mm<-model.matrix(fm)
    r<-residuals(fm)
    #calculate HAC standard errors (with function "hac_ps.R" which is
constructed as Gauss Code by Phillips and Sul (2007))
    hac<-hac_ps(r)
    se<-sqrt(diag(solve(t(mm) %*% mm))[2]*hac)
    b <- coef(fm)[2]
    tv[i] <- b/se
    alpha<-(1/2)*b

  }
  return(tv)
}
```

Function: HAC standard errors

```
hac_ps <- function(x){
  #this function calculates the hac standard errors according to Phillips and
Sul (2007, Gauss Code)
  t<-length(x)
  x1<-x[1:(t-1)]
  y1<-x[2:t]
  b1<-sum(x1*y1)/sum(x1*x1)
  ee<-y1-(x1*b1)

  a1 <- 4*(b1^2)/(((1-b1)^2)*((1-b1)^2))
  a2 <- 4*(b1^2)/((1-b1)^4)

  band1 <- 1.1447*(a1*t)^(1/3)
  band2<-1.3221*(a2*t)^(1/5)

  jb2<-seq_len(t-1)/band2
  jband2<-1.2*pi*jb2
  kern1<-((sin(jband2)/jband2)-cos(jband2))*(25/(((jb2*pi)^2*12)))

  tt<-length(ee)
  lam<-0
  for(i in 1:(tt-1)){
    ttp1<-(x[1:(tt-i)]%*% x[(1+i):tt])*kern1[i]/tt
    ttp<-(x[1:(tt-i)]%*% x[(1+i):tt])*kern1[i]/tt
    lam<- lam+ttp+ttp1
  }
  sigm<-(x %*% x)/tt
  lam<-sigm+lam
  #
  return(lam)
}
```

References

Abbott, Andrew, and Glauco De Vita. 2012. "Pairwise convergence of district-level house prices in London." *Urban Studies*, 49(4): 721–740.

Abraham, Jesse M., and Patric H. Hendershott. 1996. "Bubbles in Metropolitan Housing Markets." *Journal of Housing Research*, 7(2): 191–207.

Agresti, Alan. 2003. *Categorical data analysis.* John Wiley & Sons.

Albouy, David, Gabriel Ehrlich, and Yingyi Liu. 2016. "Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis." National Bureau of Economic Research, Inc NBER Working Papers 22816.

Alonso, William. 1964. *Location and land use: Toward a general theory of land rent.* Cambridge/Mass.: Harvard Univ. Press.

Apergis, Nicholas, Beatrice D. Simo-Kengne, and Rangan Gupta. 2015. "Convergence In Provincial-Level South African House Prices: Evidence From The Club Convergence And Clustering Procedure." *Review of Urban & Regional Development Studies*, 27(1): 2–17.

Awaworyi Churchill, Sefa, John Inekwe, and Kris Ivanovski. 2018. "House price convergence: Evidence from Australian cities." *Economics Letters*, 170(C): 88–90.

Backhaus, Teresa, Kathrin Gebers, and Carsten Schröder. 2015. "Evolution and Determinants of Rent Burdens in Germany." DIW Berlin, The German Socio-Economic Panel SOEP papers on Multidisciplinary Panel Data Research 806.

Baffoe-Bonnie, John. 1998. "The Dynamic Impact of Macroeconomic Aggregates on Housing Prices and Stock of Houses: A National and Regional Analysis." *The Journal of Real Estate Finance and Economics*, 17(2): 179–197.

Bartkowska, Monika, and Aleksandra Riedl. 2012. "Regional convergence clubs in Europe: Identification and conditioning factors." *Economic Modelling*, 29(1): 22–31.

Bauer, Thomas, Sven Feuerschütte, Michael Kiefer, Philipp an de Meulen, Martin Micheli, Torsten Schmidt, and Lars-Holger Wilke. 2013. "Ein hedonischer

Immobilienpreisindex auf Basis von Internetdaten: 2007–2011.” *AStA Wirtschafts- und Sozialstatistisches Archiv*, 7(1): 5–30.

Belke, Ansgar, and Jonas Keil. 2018. “Fundamental Determinants of Real Estate Prices: A Panel Study of German Regions.” *International Advances in Economic Research*, 24(1): 25–45.

Bischoff, Oliver. 2012. “Explaining regional variation in equilibrium real estate prices and income.” *Journal of Housing Economics*, 21(1): 1–15.

Blanco, Francisco, Victor Martín, and Guillermo Vazquez. 2016. “Regional house price convergence in Spain during the housing boom.” *Urban Studies*, 53(4): 775–798.

Bundesministerium für Wirtschaft und Energie. 2020. “Jahresbericht der Bundesregierung zum Stand der Deutschen Einheit.” Bundesministerium für Wirtschaft und Energie Institutional Report.

Burnett, J. Wesley. 2016. “Club convergence and clustering of U.S. energy-related CO2 emissions.” *Resource and Energy Economics*, 46(C): 62–84.

Caldera, Aida, and Asa Johansson. 2013. “The price responsiveness of housing supply in OECD countries.” *Journal of Housing Economics*, 22(3): 231–249.

Cameron, Gavin, John Muellbauer, and Anthony Murphy. 2006*a*. “Housing Market Dynamics and Regional Migration in Britain.” Centre for Economic Policy Research CEPR Discussion Papers 5832.

Cameron, Gavin, John Muellbauer, and Anthony Murphy. 2006*b*. “Was There A British House Price Bubble? Evidence from a Regional Panel.” University of Oxford Economics Series Working Papers 276.

Deutsche Bundesbank. 2020. “Monthly Report - October 2020.” Deutsche Bundesbank Institutional Reports 72/10.

Dunse, Neil, Sotirios Thanos, and Glen Bramley. 2013. “Planning policy, housing density and consumer preferences.” *Journal of Property Research*, 30(3): 221–238.

Dustmann, Christian, Bernd Fitzenberger, and Markus Zimmermann. 2018. "Housing Expenditures and Income Inequality." Institute of Labor Economics IZA Discussion Papers 11953.

European Systemic Risk Board. 2019. "Warning of the European Systemic Risk Board." ESRB Report.

Giussani, Bruno, and George Hadjimatheou. 1992. "House prices: An econometric model for the U.K." *Netherlands journal of housing and the built environment*, 7: 31–58.

Glaeser, Edward L, Jed Kolko, and Albert Saiz. 2001. "Consumer city." *Journal of economic geography*, 1(1): 27–50.

Glaeser, Edward L., Joseph Gyourko, and Raven E. Saks. 2006. "Urban growth and housing supply." *Journal of Economic Geography*, 6(1): 71–89.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst. 2013. "Endogenous gentrification and housing price dynamics." *Journal of Public Economics*, 100: 45–60.

Hodrick, Robert, and Edward Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking*, 29(1): 1–16.

Holly, Sean, M. Hashem Pesaran, and Takashi Yamagata. 2010. "A spatio-temporal model of house prices in the USA." *Journal of Econometrics*, 158(1): 160–173.

Holmes, Mark J, Jesús Otero, and Theodore Panagiotidis. 2019. "Property heterogeneity and convergence club formation among local house prices." *Journal of Housing Economics*, 43: 1–13.

Hyclak, Thomas, and Geraint Johnes. 1999. "House prices and regional labor markets." *The Annals of Regional Science*, 33(1): 33–49.

Jud, G. Donald, and Dan T. Winkler. 2002. "The Dynamics of Metropolitan Housing Prices." *Journal of Real Estate Research*, 23(1/2): 29–46.

Kajuth, Florian, Thomas A. Knetsch, and Nicolas Pinkwart. 2013. "Assessing house prices in Germany: Evidence from an estimated stock-flow model using regional data." Deutsche Bundesbank Discussion Papers 46/2013.

Kim, Young Se, and Jeffrey J. Rous. 2012. "House price convergence: Evidence from US state and metropolitan area panels." *Journal of Housing Economics*, 21(2): 169–186.

Koetter, Michael, and Tigran Poghosyan. 2010. "Real estate prices and bank stability." *Journal of Banking & Finance*, 34(6): 1129–1138.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2018. "Income and Wealth Inequality in America, 1949-2016." Federal Reserve Bank of Minneapolis Opportunity and Inclusive Growth Institute Working Papers 9.

Lyncker, Konrad, and Rasmus Thoennessen. 2017. "Regional club convergence in the EU: evidence from a panel data analysis." *Empirical Economics*, 52(2): 525–553.

Maennig, Wolfgang, and Lisa Dust. 2008. "Shrinking and growing metropolitan areas asymmetric real estate price reactions?: The case of German single-family houses." *Regional Science and Urban Economics*, 38(1): 63–69.

Mateusz, Tomal. 2019. "House Price Convergence on the Primary and Secondary Markets: Evidence from Polish Provincial Capitals." *Real Estate Management and Valuation*, 27(4): 62–73.

Matysiak, George A., and Krzysztof Olszewski. 2019. "A panel analysis of Polish regional cities: residential price convergence in the primary market." Narodowy Bank Polski NBP Working Papers 316.

Meen, Geoffrey. 1999. "Regional house prices and the ripple effect: a new interpretation." *Housing studies*, 14(6): 733–753.

Meen, Geoffrey. 2001. *Modelling spatial housing markets: Theory, analysis, and policy.* Springer Science & Business Media.

Miles, David. 2012. "Population Density, House Prices and Mortgage Design." *Scottish Journal of Political Economy*, 59(5): 444–466.

Mills, Edwin S. 1967. "An aggregative model of resource allocation in a metropolitan area." *The American Economic Review*, 57(2): 197–210.

Montagnoli, Alberto, and Jun Nagayasu. 2015. "UK house price convergence clubs and spillovers." *Journal of Housing Economics*, 30(C): 50–58.

Muellbauer, John, and Anthony Murphy. 1997. "Booms and Busts in the UK Housing Market." *Economic Journal*, 107(445): 1701–1727.

Murphy, Anthony, and John Muellbauer. 1994. "Explaining regional house prices in the UK." School of Economics, University College Dublin Working Papers 199421.

Muth, Richard F. 1969. *Cities and housing. The spatial pattern of urban residential land use.* The University of Chicago Press, Chicago and London.

Phillips, Peter C. B., and Donggyu Sul. 2007. "Transition Modeling and Econometric Convergence Tests." *Econometrica*, 75(6): 1771–1855.

Phillips, Peter C. B., and Donggyu Sul. 2009. "Economic transition and growth." *Journal of Applied Econometrics*, 24(7): 1153–1185.

Reichert, Alan K. 1990. "The Impact of Interest Rates, Income, and Employment upon Regional Housing Prices." *The Journal of Real Estate Finance and Economics*, 3(4): 373–391.

Roback, Jennifer. 1982. "Wages, rents, and the quality of life." *Journal of political Economy*, 90(6): 1257–1278.

Rosen, Sherwin. 1979. "Wage-based indexes of urban quality of life." *Current issues in urban economics*, 74–104.

Schnurbus, Joachim, Harry Haupt, and Verena Meier. 2017. "Economic Transition and Growth: A Replication." *Journal of Applied Econometrics*, 32(5): 1039–1042.

Scoccianti, Filippo. 2010. "Causes and Welfare Consequences of Real Estate Price Appreciation." In *Housing Markets in Europe*. 241–259. Springer.

Sivitanides, Petros Stavrou. 2018. "Macroeconomic drivers of London house prices." *Journal of Property Investment & Finance*, 36(6): 539–551.

Tsai, I-Chun. 2018. "House price convergence in euro zone and non-euro zone countries." *Economic Systems*, 42(2): 269–281.

Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill. 2010. "Why has house price dispersion gone up?" *The Review of Economic Studies*, 77(4): 1567–1606.

White, Michael. 2015. "Cyclical and structural change in the UK housing market." *Journal of European Real Estate Research*, 8(1): 85–103.

CHAPTER

FOUR

NAVIGATING THE HOUSING CHANNEL ACROSS EURO AREA REGIONS

Authors[1]**:** Niccolò Battistini[2], Matteo Falagiarda[3], Angelina Hackmann[4], Moreno Roma[5]

Abstract
This paper assesses the role of the housing market in the transmission of conventional and unconventional monetary policy across euro area regions. By exploiting a novel regional dataset on housing-related variables, a structural panel VAR analysis shows that monetary policy propagates effectively to economic activity and house prices, albeit in a heterogeneous fashion across regions. Although the housing channel plays a minor role in the transmission of monetary policy to the economy on average, its importance increases in the case of unconventional monetary policy. We also explore the determinants of the diverse transmission of monetary policy to economic activity across regions, finding a larger impact in areas with lower labour income and more widespread homeownership. An expansionary monetary policy can thus be effective in mitigating regional inequality via its stimulus to the economy.

JEL Classification: D31, E32, E44, E52, R31

Keywords: housing market, conventional and unconventional monetary policy, regional inequality, business cycles

[1]We would like to thank Paola Di Casola, Simon Hildebrandt, Fédéric Holm-Hadulla, Bartosz Maćkowiak, Klaus Masuch, Beatrice Pierluigi, Thomas Westermann and an anonymous referee for their helpful comments and suggestions. This paper has also benefited from discussions with participants at seminars at the ECB, the University of Bremen and the University of Neuchâtel, at the 6th Household Finance Workshop (Leibniz Institute SAFE), at the 16th CEUS Workshop on European Economics, at the 2022 ECHOPPE Conference on the Economics of Housing and Housing Policies. The views expressed in this paper are those of the authors and do not necessarily represent those of the European Central Bank.

[2]European Central Bank. E-mail: niccolo.battistini@ecb.europa.eu.

[3]European Central Bank. E-mail: matteo.falagiarda@ecb.europa.eu.

[4]University of Bremen. E-mail: angelina.hackmann@uni-bremen.de.

[5]European Central Bank. E-mail: moreno.roma@ecb.europa.eu.

4.1 Introduction

Profound economic and institutional differences across regions have long challenged the effectiveness of monetary policy in the euro area.[6] The unequal geography of the transmission of monetary policy has also stoked concerns about its possible side effects on regional inequality, especially owing to the unconventional measures conducted by the European Central Bank (ECB) over the last decade.[7] The ECB's large-scale asset purchases—critics maintain—have inflated the prices of assets, such as stocks and houses, unfairly favouring rich, wealthy households.[8] To the extent that similar households cluster geographically, monetary policy has, according to critics, further exacerbated regional inequality. In the transition of the ECB out of crisis-era stimulus, a crucial issue on the policy agenda has thus become the calibration of an appropriate monetary policy stance that can support the recovery while minimising economic divergence across regions. In this context, the housing market—in light of its role in the propagation of aggregate shocks, its distributional implications and its local dimension—[9] has often come to the front of the media and policy debate on the intended and unintended effects of monetary policy.[10]

Our paper contributes to the literature on this debate by assessing empirically the role of the housing market in the conventional and unconventional transmission of monetary policy across regions in the first two decades of the euro area. Our contribution is threefold. First, we construct a large dataset with a panel of 106 (mostly) NUTS2-level regions in eight euro area countries (Belgium, Germany, Spain, France, Ireland, Italy, the Netherlands and Portugal) covering the period 1999-2018. Most notably, we compile novel indicators for regional house prices and loan-to-value (LTV) ratios on the basis of loan-level data from the European DataWarehouse. We also collect regional indicators for aggregate and sectoral activity, labour market developments and housing market features from the ARDECO database and Eurostat. Our dataset features a high degree of within-country, besides cross-country, diversity pervading housing markets over the first twenty years of the euro area (Figure 4.1).[11]

[6] For a discussion of financial integration challenges in the euro area, see European Central Bank (2022). For the implications of regional heterogeneity for monetary policy in the euro area, see Cœuré (2019).

[7] See, for instance, The Economist (2016) and Cœuré (2018).

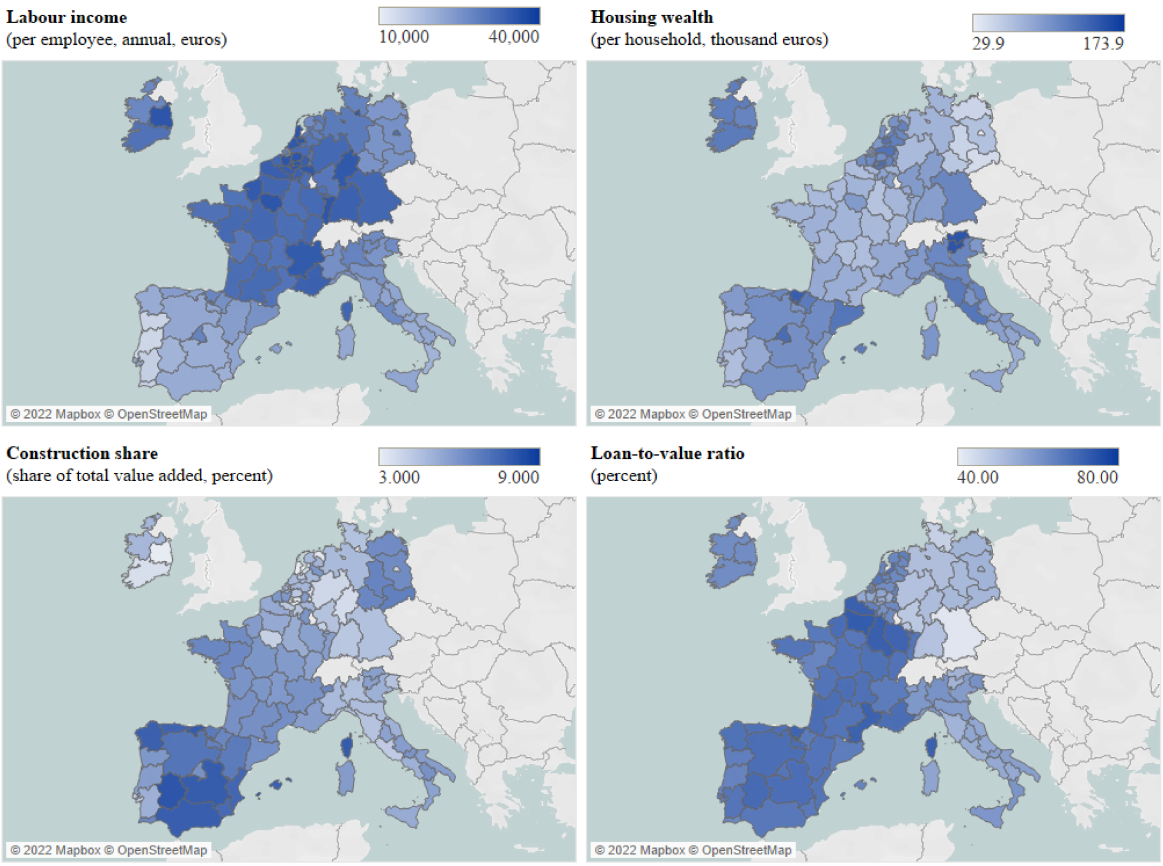
[8] Among the earliest concerns, see The Economist (2013) and The Financial Times (2015*a*).

[9] On the features of the housing market, see the comprehensive study by Piazzesi and Schneider (2016).

[10] To name a few recent examples, see, in the media, The Financial Times (2021) and, in policy circles, OECD (2020), Schnabel (2021), Battistini et al. (2021) and European Commission (2021).

[11] A simple measure of the information content specific to within-country (relative to cross-country) heterogeneity can be computed, for each variable, as the ratio of the cross-country average of the within-country

Figure 4.1: Regional heterogeneity in euro area housing markets



Source: ARDECO, Eurostat, European DataWarehouse and authors' calculations.
Notes: Labour income is measured as compensation of employees divided by number of employees (average 1999-2018). Housing wealth is computed as house price level (average 1999-2018) multiplied by homeownership rate (share of households living in owner-occupied dwellings) in 2011. Construction share is calculated as construction value added divided by total value added (average 1999-2018). Loan-to-value ratio is computed as the amount of the mortgage loan divided by the value of the underlying property (average 1999-2018).

This indicates that the information content of our regional dataset extends beyond that of a typical cross-country panel, confirming the pronounced local dimension of housing markets.

Our second contribution is to consider monetary policy through its conventional and unconventional transmission mechanisms. To this end, we tap the Euro Area Monetary Policy Database (Altavilla et al., 2019*b*) to construct a measure of monetary policy surprises. To isolate the impact of "genuine" monetary policy surprises, we adopt a high-frequency identification and impose sign and zero restrictions on high-frequency changes in OIS interest

standard deviations to the cross-country standard deviation of the within-country averages. All the considered variables exhibit a sizeable degree of relative within-country variation, especially construction share (85 percent), followed by homeownership rate and labour income (both 55 percent) and LTV ratio (36 percent).

rates and stock prices around the ECB's monetary policy announcements (Jarociński and Karadi, 2020). We assume that the conventional transmission mechanism of monetary policy has mainly operated through short-term rates, whereas long-term rates were primarily related to the unconventional transmission mechanism of monetary policy in the aftermath of the Global Financial Crisis.

Third, using our regional dataset and our measure of conventional and unconventional monetary policy, we design a methodology to assess the role of the housing market in the transmission of monetary policy to the real economy. Using a structural panel vector autoregression (SPVAR) model with regional GDP, employment and house prices as endogenous variables, and euro area monetary policy shocks as exogenous variable, we first assess the average impact of monetary policy on GDP, employment and house prices across regions. Accounting for the endogenous reaction of GDP to employment and house prices, we further quantify the role of the employment and the housing channels in conveying monetary stimulus.

Our results show a significant, positive impact of a monetary policy easing on GDP, employment and, to a lesser extent, house prices. Further, monetary policy stimulus to the overall economy transmits mainly through the employment channel, in line with Hauptmeier, Holm-Hadulla and Nikalexi (2020), with a rather limited role for the housing channel, consistently with findings in Slacalek, Tristani and Violante (2020) and Lenza and Slacalek (2021). However, unconventional monetary policy is estimated to induce significantly larger responses in house prices, relative to conventional monetary policy, thereby amplifying the housing channel.

Finally, we provide an anatomy of the long-term drivers of the diverse impact of monetary policy across euro area regions. The region-specific estimates of our benchmark SPVAR model allow us to dissect the role of several housing-related economic and institutional characteristics. We find that monetary policy has a larger impact on the economy of regions with lower labour income and a higher homeownership rate. This suggests that poorer regions stand to benefit the most from expansionary monetary policy, but can also be more negatively affected from a policy tightening.

Overall, our results point to an effective, yet widely heterogeneous transmission of monetary policy across the euro area, with monetary policy stimulating economic activity mainly through labour income, compared with housing wealth. Nevertheless, the housing channel becomes more relevant in the unconventional transmission of monetary policy. Moreover, as monetary policy is found to impact poorer regions the most, policy-makers should carefully monitor the risks of an increase in cross-regional inequality as monetary policy normalises, especially in the case of resurgent fragmentation risks. Our findings suggest that a proper assessment of the monetary policy transmission should not neglect the housing market, with its multiple sources of propagation and its pronounced local dimension.

The remainder of this paper is structured as follows. Section 4.2 describes the data. Section 4.3 lays out the theoretical and empirical frameworks. Section 4.4 presents a quantitative assessment of the housing channel of monetary policy. Section 4.5 analyses the role of economic and institutional characteristics in explaining the heterogeneous impact of monetary policy across regions. Section 4.6 conducts robustness tests on our main results. Section 4.7 draws concluding remarks.

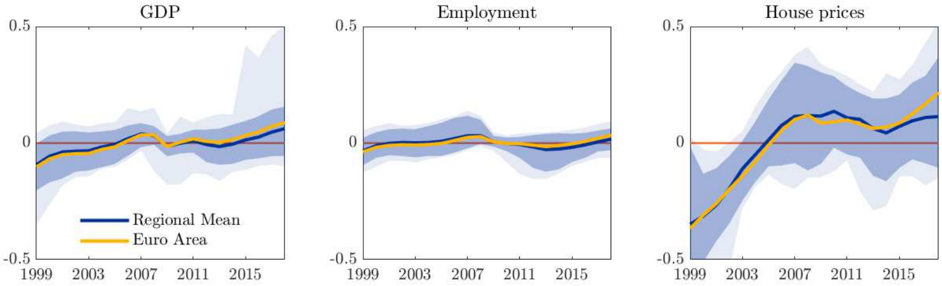
4.2 Data

4.2.1 Regional dataset

Our regional dataset has annual frequency and spans the period from 1999 to 2018. It covers 106 regions of eight euro area countries (Belgium, Germany, Spain, France, Ireland, Italy, the Netherlands and Portugal) accounting for around 90 percent of euro area gross domestic product (GDP). We consider NUTS2 regions for Belgium, Spain, France, Ireland, Italy, the Netherlands, and Portugal and NUTS1 regions for Germany.[12] Regional data on real GDP, real gross value added (GVA) for the construction and manufacturing sectors, real compensation of employees, as well as employment and population are obtained from the ARDECO database, which is maintained and updated by the Joint Research Centre of the European Commission. Moreover, we collect regional data on homeownership rate (share of households living in owner-occupied housing) and population density (persons per square

[12] Very small regions (Ceuta and Melilla in Spain; Madeira and Azores in Portugal; overseas departments in France) are excluded. In line with the Italian Constitution, we consider the provinces of Trentino and Alto Adige/Südtirol a single political region, although they are two different NUTS2 areas. Therefore, the variables available for these two provinces at the NUTS2 level are aggregated or averaged at the regional level. We consider NUTS1-level regions for Germany in order to have a number of regions (16) similar to that of France (22), Italy (20), and Spain (17). The use of NUTS2 regions for Germany (which are 38) would have led this country to be over-represented in the aggregate estimates. As regards the other countries, we consider 11 regions for Belgium, 12 for the Netherlands, 5 for Portugal and 3 for Ireland.

Figure 4.2: Key variables in our dataset



Notes: Demeaned log variables. The yellow line depicts euro area aggregate data, while the dark blue line the cross-regional mean of the variable. The dark (light) blue shading indicates 10th and 90th (1st and 99th) percentiles of the regional distribution.

kilometre) from Eurostat.[13]

Crucial for our analysis, house price indices, loan-to-value (LTV) ratios and the share of variable-rate mortgages at the regional level are derived via loan-level data provided by the European DataWarehouse (ED). The ED is a securitisation repository that collects, validates and makes available detailed, standardised and asset class-specific loan-level data for asset-backed securities (ABS) transactions. For our purposes, only residential mortgage-backed securities (RMBS) transactions are used. Note that ED data dictate our choice on the country coverage and the level of geographical disaggregation. First, within the euro area, ED data are available only for the countries included in our sample. Second, NUTS3-level geographical units (NUTS2 in Germany) would not ensure that the sample is sufficiently representative, as only a relatively small number of loans may be recorded at such granular level in some regions. For more details on how ED data are processed, see Appendix 4.7.

Our key variables (i.e. GDP, employment and house prices) are transformed as follows. We consider real GDP and employment in per capita terms. We do this to make our estimates comparable to other empirical studies and consistent with assessments based on standard DSGE models, where the population is typically normalised to unity and economic aggregates are thus in per capita terms. Moreover, we take the log of GDP, employment and house price indices. Finally, we demean these variables in order to remove region-specific fixed effects in the data.[14]

[13] Regional data on homeownership rates are available from Eurostat only for a few, distant years at irregular intervals. Hence, we only consider 2011 data, which broadly corresponds to the middle of our sample.

[14] Note that our methodology based on mean-group estimation deals with further potential fixed effects in the transmission of monetary policy by estimating region-specific parameters.

Table 4.1: Summary statistics of the key variables

		Mean	Median	Minimum	Maximum	Standard deviation
GDP	regional	29467	28052	14181	65785	9267
	national	32019	33480	17474	45529	8872
Employment	regional	43.63	43.00	31.24	65.17	6.84
	national	45.10	43.19	40.95	52.38	4.41
House prices	regional	146.01	145.84	97.68	193.39	23.23
	national	149.35	153.18	114.76	180.03	21.87

Notes: Real GDP and employment are in per capita terms. National GDP and employment are calculated as cross-regional aggregate of all regions within a country. National house prices are given by GDP-weighted cross-regional means of all regions within a country.

A closer look at our regional dataset confirms its suitability to investigate the role of the housing market in the euro area. Figure 4.2 shows indeed that the cross-region mean of each variable, computed across the 106 regions in the eight countries in our sample, tracks well the corresponding euro area aggregate over time. Moreover, the cross-region dispersion of house prices is significantly higher than that for the other variables, confirming that the housing market is indeed a regional phenomenon. Lastly, the dispersion across regions, especially between the 1st and 99th percentiles, seems to widen in the second half of the sample, possibly reflecting the impact of the Global Financial Crisis and the Sovereign Debt Crisis. This pattern is already documented in Hauptmeier, Holm-Hadulla and Nikalexi (2020) for GDP, while we observe similar dynamics for house prices.

Table 4.1 shows descriptive statistics on the cross-region and cross-country distributions of our variables over the sample period. For all variables, we find a higher degree of heterogeneity on the regional vis-à-vis the national level. On average over the entire period, GDP per capita ranges at the national level between 17,474 EUR in Portugal and 45,529 EUR in Ireland, while the regional minimum is 14,181 EUR in Norte (Portugal) and the maximum is 65,785 EUR in Région de Bruxelles-Capitale (Belgium). Regarding house prices, we also find a large cross-regional dispersion with a minimum house price index of 97.7 in Sachsen-Anhalt (Germany) and a maximum of 193.4 in País Vasco (Spain). The national house price indices range between 114.8 in Germany and 180.0 in Spain. Comparing these statistics over three different time periods (1999-2008, 2009-2012 and 2013-2018) reveals differences in the dispersion of the variables over time (see Table 4.4 in Appendix 4.7). While all variables show the lowest regional dispersion before the Global Financial Crisis, the standard deviation of

GDP and employment is the largest between 2013 and 2018. In contrast, the standard deviation of regional house prices is the largest during the Global Financial Crisis and decreases thereafter.

4.2.2 Monetary policy shocks

We identify monetary policy shocks by means of high-frequency changes in OIS interest rates and stock prices around the ECB's monetary policy decisions. A narrow time window around monetary policy events allows us to measure exogenous changes in the monetary policy stance (i.e. monetary policy surprises). For this purpose, we use the Euro Area Monetary Policy Database (EA-MPD) by Altavilla et al. (2019*b*) containing high-frequency movements in OIS interest rates and EURO STOXX 50 around the ECB's monetary policy announcements. The EA-MPD differentiates between three time windows: the publication of the press release, the press conference, and the union of these two windows, referred to as "monetary policy event". In our analysis, we consider the window of the monetary policy event as a reference period (Enders, Hünnekes and Müller, 2019, Holm-Hadulla and Thürwächter, 2021).[15]

4.2.2.1 Pros and cons of event-based monetary policy surprises

The use of an event-based identification of genuine monetary policy shocks comes with some caveats, but also clear advantages. On the one hand, as any event-based identification, this strategy is successful insofar as it captures all the relevant monetary policy events. During speeches, interviews and other public occasions, monetary authorities may partly signal policy shifts before the monetary policy events (i.e. press releases and conferences). The measured monetary policy surprises in our dataset ultimately reflect the changes in the risk-free yield curve and stock prices within a narrow event window due to deviations of the actual announcements from market expectations (Rostagno et al., 2021). Hence, this event-based identification strategy may over- or under-estimate monetary policy surprises taking shape in a period stretching beyond the event window if, for example, the relevant events are already "discounted" by market participants or if there are delayed market adjustments

[15] In our sample, Governing Council meetings took place in regular intervals of six weeks. At 13 : 45 CET a press release provides the policy decision and at 14 : 30 CET, the president explains the rationale of the decision in a press conference in more detail. The change of the financial market variables due to the monetary policy event is given as the change of the median value in the pre-release window (13 : 25 CET to 13 : 35 CET) and the median value in the post-conference window (15 : 40 CET 15 : 50 CET).

to the policy announcements.

On the other hand, this identification strategy is insulated from other problems afflicting conventional approaches (Ramey, 2016). Unlike empirical approaches relying on observed interest rates, monetary policy surprises identified from high-frequency event-studies are exogenous to economic conditions, which are already part of the market participants' information set at the time of the announcement. Further, unlike DSGE models or structural VAR models, the theoretical assumptions needed to capture monetary policy shocks in high-frequency event-studies are minimal. This comes with important benefits. First, the risk of estimation issues due to model misspecification is low. Second, any possible time dependence in the reaction function used by monetary authorities is already taken into account, at least to the extent that market participants have incorporated this variation when interpreting monetary policy announcements.

The identification of our monetary policy shocks poses two main challenges, namely the selection of "genuine" shocks and the aggregation of surprises from an event-based frequency to an annual frequency. We explain how we address both challenges in the next two subsections.

4.2.2.2 Identification of genuine monetary policy shocks

OIS interest rate changes around monetary policy events do not only reflect how market participants assess whether and how the ECB adjusts its policy instruments, but also their perception of potential superior information on the state and prospects of the economy the ECB might have. For instance, if the monetary authority announces an interest rate hike and market participants see it as a true monetary policy tightening, this will be accompanied by a negative stock price reaction. This is a so-called genuine monetary policy shock. Conversely, if market participants perceive this increase as a sign of buoyant economic prospects, this will have a positive impact on the stock price. This is a so-called central bank information shock (see Jarociński and Karadi, 2020).

We disentangle (genuine) monetary policy shocks and (central bank) information shocks by imposing sign and zero restrictions on high-frequency changes in OIS interest rates and stock prices. In line with Jarociński and Karadi (2020), high-frequency OIS interest rate changes are assumed to be uncorrelated with their own past values and with current and past values of other variables, since they are measured in a narrow time window around monetary policy announcements. We extend the same modelling assumption to stock price movements, as these are measured over the same narrow window.[16] Hence, we can use the series of OIS interest rate and stock price changes as reduced-form residuals and impose sign restrictions directly on their covariance matrix to identify monetary policy and information shocks.[17]

To capture the movements across the term structure, we use OIS interest rate changes at different points of the yield curve. We focus on the 3-month and 10-year maturities to ensure sufficient liquidity in the underlying instruments. Our focus on distant maturities (3 months and 10 years) is also justified by the fact that they are less prone to be affected by both conventional and unconventional monetary policy measures, compared with intermediate maturities.

Our identification strategy allows us to disentangle conventional and unconventional monetary policy shocks. We impose the following sign and zero restrictions.

	CMP_d	UMP_d	INF_d
$\Delta OIS3M_d$	+	0	+
$\Delta OIS10Y_d$	0	+	+
ΔSP_d	-	-	+

In the table above, $\Delta OIS3M_d$, $\Delta OIS10Y_d$ and ΔSP_d denote the change in the 3-month OIS interest rate, the 10-year OIS interest rate and the EURO STOXX 50 index at event date d, while CMP_d, UMP_d and INF_d refer to conventional monetary policy, unconventional monetary policy and information shocks, respectively. Finally, we compute total monetary policy shocks as the sum of conventional and unconventional monetary policy shocks. Our restrictions imply that a positive conventional (unconventional) monetary policy shock induces an increase in the 3-month (10-year) OIS interest rate, a decrease in the stock price and no movement in the 10-year (3-month) OIS interest rate, while a positive information shock is associated with an increase in all variables.[18]

[16] This assumption differs from other approaches in the literature, who measure other financial variables over a longer time span (e.g. a month) and thus cannot rule out their endogenous reaction to high-frequency interest rate changes. These studies typically impose further structure on the model to extract the shock from co-movements between interest rate changes and other financial variables (Jarociński and Karadi, 2020).

[17] We implicitly use flat priors on the covariance matrix of our reduced-form residuals. When comparing methods, Jarociński and Karadi (2020) argue that their results with a Bayesian approach are similar to the frequentist results by Gertler and Karadi (2015).

[18] Our main findings are largely unchanged if TMP_d is estimated directly by imposing a negative co-movement between the sum of the 3-month and the 10-year OIS interest rate changes and stock price changes, with information shocks inducing a positive co-movement between these two variables.

Our identification strategy warrants an explanation of how to interpret conventional and unconventional monetary policy shocks. On the one hand, the reaction of OIS interest rates at the short end of the yield curve should uniquely reflect conventional monetary policy measures up to 2008. Thereafter, as standard measures stopped affecting the short end of the term structure due to an effective lower bound on risk-free rates, the ECB sought to enhance the conventional transmission of its monetary policy through non-standard measures, such as fixed-rate tenders with full allotment, forward guidance, and negative interest rate policy (see, for example, the discussion in Gambacorta, Hofmann and Peersman, 2014, and Falagiarda and Reitz, 2015). On the other hand, the reaction of long-term OIS interest rates should primarily encompass the effects of several unconventional measures implemented since 2011, such as asset purchase programmes, longer-term refinancing operations and some types of forward guidance. Hence, our approach can capture the impact of monetary policy through its conventional and unconventional *transmission mechanisms*, rather than the impact of the conventional and unconventional *measures per se*.

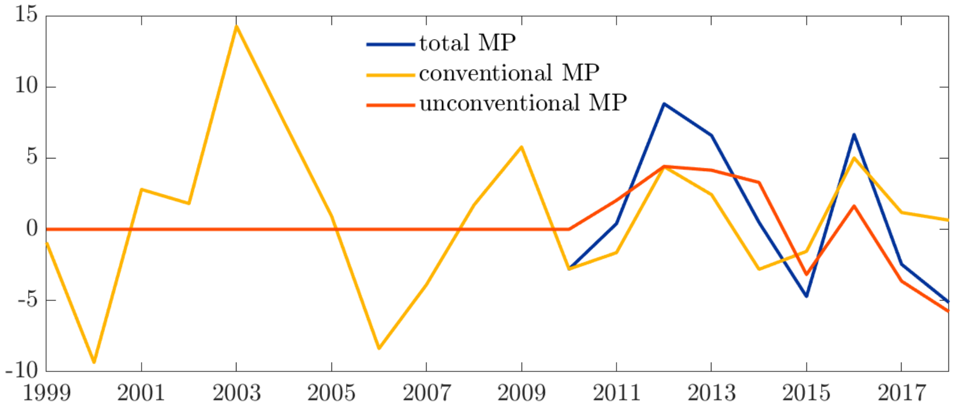
4.2.2.3 Temporal aggregation of event-based monetary policy shocks

To account for the annual frequency of our regional dataset, we apply a weighting procedure. Specifically, we assign theoretical weights to monetary policy shocks depending on the distance of the day of the event from the first day of the reference year. Formally, to calculate a monetary policy shock for year t, we consider all monetary policy shocks in year t and $t-1$ and give a higher weight to shocks at the beginning of year t and at the end of year $t-1$ compared with shocks at the end of year t and the beginning of year $t-1$, that is:

$$\begin{aligned} w_{d,t} &= 1 - \left| \frac{d_t - d_t^1}{365} \right| \\ W_{d,t} &= \frac{w_{d,t}}{\sum_{i=1}^{N} w_{i,t}} \\ MP_t &= N \sum_{d=1}^{N} W_{d,t} MP_d, \end{aligned} \tag{4.1}$$

where $w_{d,t}$ denotes the theoretical weight attached to the monetary policy event on day d in year t or $t-1$ given the reference year t, $W_{d,t}$ its normalised value such that $\sum_{d=1}^{N} W_{d,t} = 1$, N the number of monetary policy events in year t and $t-1$, d_t^1 the first day of year t and MP_t our final measure of (total, conventional or unconventional) monetary policy shock in

Figure 4.3: Monetary policy surprises



Notes: The chart shows the time series of the (genuine) monetary policy shocks at annual frequency resulting from the weighting procedure.

year t.

Intuitively, Equation (4.1) aligns the monetary policy surprises identified at high frequency with the concomitant economic developments, then building consistent low-frequency monetary policy shocks. To give an example, consider a monetary policy surprise in the fourth quarter of year $t-1$, such as the monetary tightening observed on 3 December 2015, reflecting financial markets' disappointment about the increase of the size of the ECB's asset purchase programme (The Financial Times, 2015*b*). To the extent that this monetary policy shock has a relatively larger impact on the contemporaneous growth rates of economic variables, this impact will be more visible in year t, i.e. 2016, than in year $t-1$, i.e. 2015.[19]

Figure 4.3 shows the implied time series for our total, conventional and unconventional monetary policy shocks. Looking at the total monetary policy shocks, monetary tightening starting in 2008 to curb rising inflation is followed by monetary accommodation in 2010 and 2011 to fight the Global Financial Crisis and then again in 2014 and 2015 as a reaction to the Sovereign Debt Crisis. As of 2015, when the large-scale APP are launched, the main impulse from monetary accommodation switches from the conventional to the unconventional transmission mechanism.[20]

[19]This follows from a simple accounting exercise, which implies that 25 and 75 percent of the quarterly growth rate of any economic variable in the fourth quarter of year $t-1$ contribute to its annual growth rates in years $t-1$ and t, respectively. Our theoretical weights, calculated at daily frequency, are largely consistent with these quarterly weights.

[20]Due to data availability in the EA-MPD, where monetary policy surprises for the 10-year tenure are recorded as of 7 July 2011, unconventional monetary policy shocks only start in 2011. Although a non-

4.3 Methodology

This section presents the theoretical framework and the empirical strategy adopted. First, we outline the channels of monetary policy that we aim to capture in our empirical assessment. Then, we describe our benchmark structural panel vector autoregression (SPVAR) model and discuss how we disentangle the channels of interest. We finally present a simple econometric framework to link the estimated monetary policy impact to housing-related economic and institutional characteristics at the regional level.

4.3.1 The transmission of monetary policy through the housing channel

Monetary policy propagates to the real economy through several direct and indirect channels. For illustrative purposes, we consider a closed economy with households, firms, financial intermediaries and a central bank. This framework is consistent with a broad class of general equilibrium models used to analyse the role of the housing market in the transmission of monetary policy, including models with collateral constraints (Iacoviello, 2005; Guerrieri and Iacoviello, 2017), non-rational expectations (Adam and Woodford, 2021) and household heterogeneity (Kaplan, Moll and Violante, 2018).

Let us assume that the central bank engenders an expansionary monetary policy shock, i.e. risk-free rates decline more than expected. This directly improves supply conditions on the credit market, inducing financial intermediaries to expand their lending to the private sector. This in turn supports households and firms' current spending decisions, thus stimulating aggregate demand across the consumption, housing, capital and labour markets. At the same time, as the central bank announces the monetary easing, private sector agents adjust their expectations to internalise the improved future economic prospects. Positive expectations exert upward pressures on financial and non-financial asset prices. In turn, house price increases boost homeowners' wealth, thus increasing private consumption. As house prices grow compared with construction costs, favourable Tobin's Q effects make housing investment more attractive. To the extent that housing is posted as collateral, an increase in house prices relaxes borrowing constraints and allows homeowners to smooth consumption

standard monetary policy tool, such as the the Securities Markets Programme (SMP), had already been activated for a year, we believe that this should not significantly affect our results. Indeed, the objective of the SMP was "to ensure depth and liquidity" and "restore an appropriate monetary policy transmission", thus clearly falling under our definition of a conventional transmission mechanism.

over the life cycle, further boosting aggregate demand. Overall, monetary accommodation expands the resources available for the private sector, generating positive income and wealth effects for both households and firms and supporting activity.

In a first step, our structural panel VAR (SPVAR) analysis identifies a subset of the various general equilibrium effects of monetary policy at play. Specifically, we consider household income sources, especially housing wealth, proxied by house prices and capturing the housing channel, and labour income, proxied by employment and capturing the employment channel. Our focus on the comparison between the housing and the employment channels is motivated by the growing evidence, both in the theoretical (Kaplan, Moll and Violante, 2018) and the empirical (Hauptmeier, Holm-Hadulla and Nikalexi, 2020; Lenza and Slacalek, 2021) literature, pointing to a larger role for labour income relative to housing wealth in transmitting monetary policy to the real economy. Given the scope of our analysis and the limited availability of regional data on other variables, the residual effect of monetary policy includes the net effect of several other channels identified in the literature, such as intertemporal substitution, net interest rate exposure, net nominal balance sheet positions, stock market wealth (Slacalek, Tristani and Violante, 2020), as well as other income sources supporting corporate, public and net foreign demand.

In a second step, our empirical analysis lays out an anatomy of the impact of monetary policy on economic activity across regions. By means of formal econometric regressions, we dissect the regional impact of monetary policy along several dimensions related to the housing market, such as labour income, housing wealth, the construction share of total value added and the share of variable-rate mortgages. The mean-group estimation used in our first step becomes instrumental to this analysis, as it provides us with region-specific impacts of monetary policy. This approach is different from subsample analysis or quantile (auto)regressions (Koenker and Hallock, 2001; Koenker and Xiao, 2006), as it fully exploits the heterogeneity in the data and does not impose additional structure.

4.3.2 A Structural Panel VAR for the housing channel

We first consider the following reduced-form VAR model in companion form:

$$Y_{i,t} = B_i Y_{i,t-1} + C_i X_t + u_{i,t}, \tag{4.2}$$

where $Y_{i,t}$ is a vector of unit-specific endogenous variables for region i at time $t = 1, ..., T$, X_t a vector of common exogenous variables (including a constant and a trend) and $u_{i,t}$ a serially uncorrelated vector of errors with zero mean and a constant positive definite variance-covariance matrix. Matrices B_i and C_i denote reduced-form parameters.

The equivalent representation in structural form is given by:

$$\mathrm{A}_i Y_{i,t} = \mathrm{B}_i Y_{i,t-1} + \Gamma_i X_t + \Delta_i \epsilon_{i,t}, \tag{4.3}$$

where A_i, B_i, Γ_i and Δ_i are matrices of structural parameters, which are related to the reduced-form parameters as follows:

$$\begin{aligned} \mathrm{A}_i^{-1}\mathrm{B}_i &= B_i \\ \mathrm{A}_i^{-1}\Gamma_i &= C_i \\ \mathrm{A}_i^{-1}\Delta_i\epsilon_{i,t} &= u_{i,t}. \end{aligned} \tag{4.4}$$

In our analysis, we focus on the effect of common exogenous variables Γ_i and the contemporaneous relationships among endogenous variables A_i, while we do not investigate the impact of region-specific structural shocks implied by Δ_i.

The benchmark structural panel VAR (SPVAR) model includes three endogenous variables $Y_{i,t} = [\mathrm{GDP}_{i,t}, \mathrm{Employment}_{i,t}, \mathrm{House\ prices}_{i,t}]$, where $\mathrm{GDP}_{i,t}$ is measured as real GDP divided by population, $\mathrm{Employment}_{i,t}$ as number of employees divided by population and $\mathrm{House\ prices}_{i,t}$ as average house price index. We include as exogenous variable X_t the relevant measure of monetary policy shock, either total monetary policy $X_t = TMP_t$ or, simultaneously, conventional and unconventional shocks $X_t = [CMP_t, UMP_t]$. Considering similar regional data, Beetsma, Cimadomo and Van Spronsen (2021) argue that common, national and regional factors all play an important role in explaining regional business cycles. In particular, they find that one common (euro area) factor, mostly related to monetary policy, one national factor and one idiosyncratic factor can account for regional dynamics. To the extent that the lagged endogenous variables net out the impact of country- and region-specific developments, our benchmark specification appropriately disentangles the impact of common (conventional and unconventional) monetary policy shocks. As a robustness check, we also include other explanatory variables, focusing on the part of cross-sectional averages unexplained by our total monetary policy shocks, and find broadly similar results (Section 4.6).

Note that the vector of reduced-form coefficients C_i represents the overall impact of a monetary policy shock on GDP, employment and house prices. To disentangle the contribution of the housing and employment channels, we need to identify the structural coefficients in A_i and Γ_i denoting the contemporaneous relationships among endogenous variables. Once we estimate the reduced-form parameters with standard OLS, we use the scoring algorithm (Amisano and Giannini, 1997) to impose the following identifying restrictions:

$$\mathrm{A}_i = \begin{bmatrix} 1 & \alpha_{i,12} & \alpha_{i,13} \\ 0 & 1 & \alpha_{i,23} \\ 0 & 0 & 1 \end{bmatrix} \tag{4.5}$$

and

$$\Gamma_i = \begin{bmatrix} \gamma_{i,1} \\ \gamma_{i,2} \\ \gamma_{i,3} \end{bmatrix}, \tag{4.6}$$

which imply a recursive structure, with the first variable as the most endogenous variable. Using Equation (4.4), we obtain the following vector of structural coefficients:

$$\begin{aligned} C_i = \mathrm{A}_i^{-1}\Gamma_i &= \begin{bmatrix} 1 & -\alpha_{i,12} & -\alpha_{i,13} + \alpha_{i,12}\alpha_{i,23} \\ 0 & 1 & -\alpha_{i,23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{i,1} \\ \gamma_{i,2} \\ \gamma_{i,3} \end{bmatrix} \\ &= \begin{bmatrix} \gamma_{i,1} - \alpha_{i,12}\gamma_{i,2} - (\alpha_{i,13} + \alpha_{i,12}\alpha_{i,23})\gamma_{i,3} \\ \gamma_{i,2} - \alpha_{i,23}\gamma_{i,3} \\ \gamma_{i,3} \end{bmatrix}, \end{aligned} \tag{4.7}$$

which allows us to disentangle the housing and employment channels from other direct and indirect channels. Specifically, looking at the impact of monetary policy on GDP in the first element of C_i, the three terms reveal the contribution from unidentified direct and indirect channels, $\gamma_{i,1}$, the contribution from the employment channel, $-\alpha_{i,12}\gamma_{i,2}$, and the contribution from the housing channel, $-(\alpha_{i,13} + \alpha_{i,12}\alpha_{i,23})\gamma_{i,3}$.

Note that our identification strategy only aims to disentangle the contribution of the employment and housing channels to the transmission of monetary policy to economic activity. As such, our identification affects neither the interpretation nor the estimated impact of monetary policy shocks. In our benchmark specification, we focus on a Cholesky structure among endogenous variables, with GDP ordered as the most endogenous variable and house prices as the most exogenous one. In this way, our estimates account for all the potential contemporaneous effects of the housing and the employment channels on the transmission of monetary policy to the business cycle. As the contemporaneous contributions tend to assign a larger weight to the less reactive (or more exogenous) variables, the estimates from our benchmark model should be considered as an upper bound of the contribution of the employment and the housing channels.[21]

We estimate our SPVAR model with one lag for each region i and apply the mean-group (MG) estimation procedure proposed by Pesaran and Smith (1995) to obtain an average response across regions. Our choice of the number of lags is standard considering the frequency of our data, and ensures the use of a consistent model across regions.[22]

4.3.3 Analysing the regional heterogeneity of housing markets

In a second step, we provide an anatomy of the diverse impact of monetary policy across euro area regions. More specifically, it is formally tested which housing-related economic and institutional characteristics contribute the most to explain the regional impact of monetary policy. To that purpose, we estimate the following regression:

$$y_i = \alpha + \sum_{n=1}^{N} \beta_{i,n} x_{i,n} + \sum_{m=1}^{M} \gamma_{i,m} z_{i,m} + \epsilon_i, \tag{4.8}$$

where the dependent variable y_i represents the region-specific long-term (5-year) cumulative monetary policy impact as estimated via the mean-group procedure, α, β_i and γ_i are parameters, $x_{i,n}$ corresponds to the nth explanatory variable ($n = 1, ..., N$), $z_{i,m}$ corresponds to the mth demographic, country and country-group control variable ($m = 1, ..., M$) and ϵ is an error term. The set of regional economic and institutional explanatory variables $x_{i,n}$ includes labour income (measured as compensation per employee), housing wealth (home-

[21] This is confirmed when we invert the ordering of the variables (see Section 4.6).

[22] Assuming two lags, the SPVAR model produces largely comparable results in qualitatively and quantitatively terms. However, the impulse response functions become less smooth and more volatile compared with our benchmark specification, hence impairing the interpretation of our findings.

ownership rate times average house price level), construction and manufacturing shares of total value added, the share of variable-rate mortgages and a measure of lending activity. The demographic controls include total employment and population density at the regional level. Consistently with the dependent variable, which reflects the average estimated impact of monetary policy, all regressors are averaged over the sample period, except for the homeownership rate, only available for 2011.

4.4 The housing channel of monetary policy

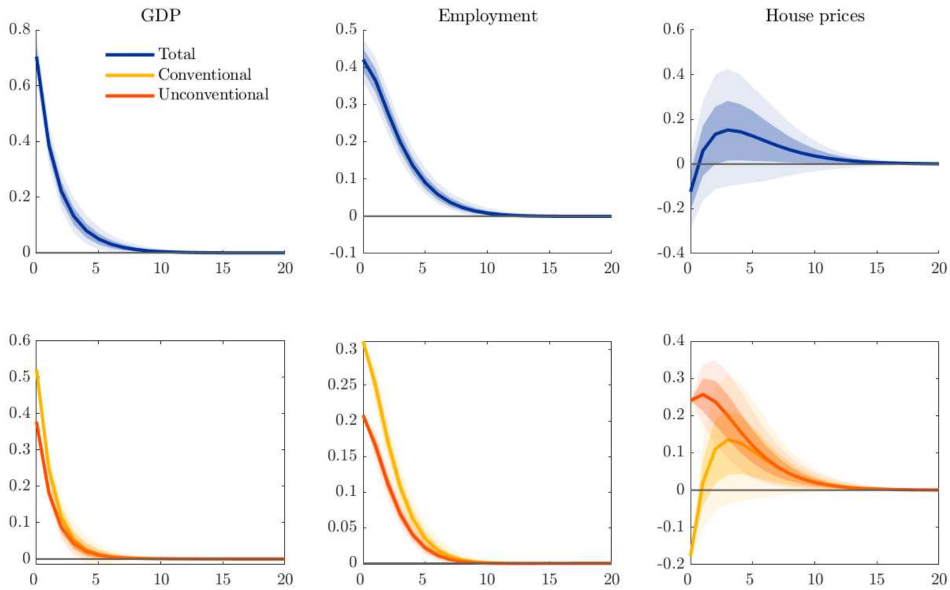
Based on the mean-group estimates of our SPVAR model, Figure 4.4 shows the responses of GDP, employment and house prices to an expansionary monetary policy shock, standardised to its mean absolute value.[23] We differentiate between responses to a total monetary policy shock (first row), conventional and unconventional monetary policy shocks (second row). As suggested by economic theory, GDP, employment, and house prices increase after a monetary policy easing shock, with the statistical significance at least at the 68 percent level. However, for house prices, the response on impact is not statistically different from zero. On average, total monetary policy shocks lead to an increase in (detrended) GDP and employment levels by 0.7 and 0.4 percent on impact, respectively, gradually declining over time. House prices exhibit instead a hump-shaped reaction, with a positive peak response of 0.15 percent after three years before fading out over the remainder of the horizon.[24]

The responses to conventional and unconventional monetary policy shocks are significantly different for all the variables. For GDP and employment, the effect of conventional monetary policy shocks is larger compared to unconventional shocks. For house prices the opposite occurs, with the peak response to unconventional shocks being around twice the response to conventional shocks (almost 0.3 after 1 year versus slightly more than 0.1 percent after three years, respectively). The impact of a conventional monetary policy shock on house prices reported in the literature generally varies between 0 and 0.6 percent, with our estimate being close to the lower end of this range (see, e.g., Musso, Neri and Stracca, 2011; Nocera

[23] We choose to set the size of the monetary policy shocks to their mean absolute value since, although their mean value is not necessarily zero over the sample, this metric is a better gauge of their average estimated impact.

[24] Corsetti, Duarte and Mann (2020) find a smaller difference in the impact of monetary policy on GDP and house prices (with the long-term impact on GDP being almost twice that on house prices), while a similar impact is documented in Rosenberg (2020).

Figure 4.4: Impulse response functions to an expansionary monetary policy shock



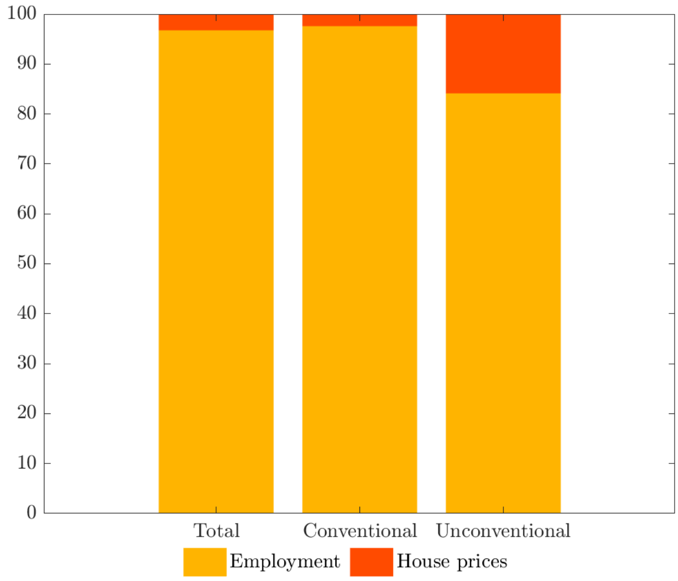
Notes: The size of the monetary policy shock is calculated as its mean absolute value, which is 5.2 basis points for the total, 4.6 basis points for the conventional and 1.7 basis points for the unconventional monetary policy shock. The y-axis reports the percentage change in (detrended) levels of each variable over the considered horizon. The x-axis reports the years. Solid lines denote point estimates and light (dark) shaded areas 95 percent (68 percent) confidence bands.

and Roma, 2017; Zhu, Betzinger and Sebastian, 2017; Huber and Punzi, 2020; Hülsewig and Rottmann, 2021).

By estimating the contemporaneous responses of our endogenous variables to a monetary policy shock on GDP as described in Equation (4.7), it is possible to examine the role of the housing and the employment channels. Figure 4.5 compares the share of the GDP response to a total, conventional and unconventional shock explained by house prices and employment at the 5-year horizon. With a share of less than 4 percent, the housing channel plays only a minor role in the transmission of a total and conventional monetary policy shock. In contrast, around 16 percent of the explained part in the transmission of unconventional monetary policy shocks to economic activity can be attributed to the housing channel.

A forecast error variance decomposition provides insight regarding the contribution of a monetary policy shock to fluctuations in GDP, employment and house prices at the 5-year horizon. As shown in Figure 4.6, total monetary policy shocks explain about 7 percent of the variation in both GDP and employment. When conventional and unconventional monetary policy shocks are included separately, conventional shocks account for about 4 percent of

Figure 4.5: Importance of housing and employment channels

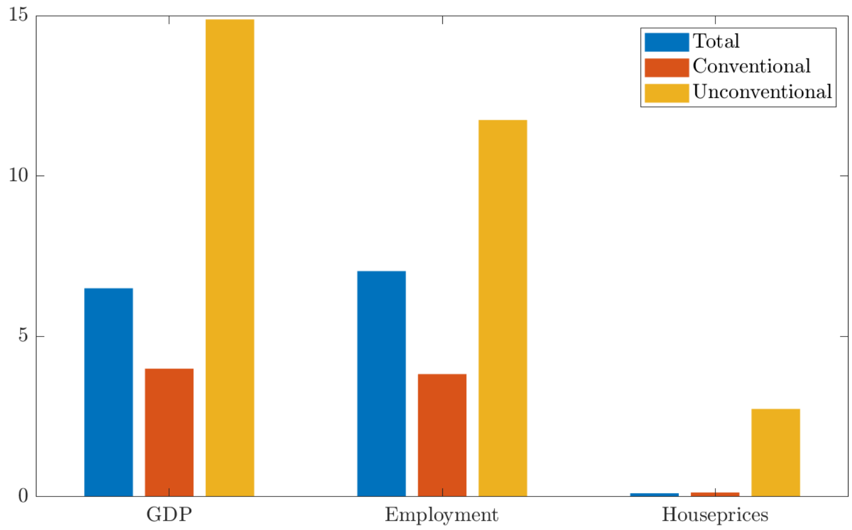


Notes: The y-axis shows the share of the contribution of employment and house prices out of the sum of their contributions to the GDP response to a total, conventional and unconventional monetary policy shock.

the variation in GDP and employment, while unconventional monetary policy shocks can explain 15 percent and 12 percent of fluctuations in GDP and employment, respectively. However, monetary policy shocks explain a relatively small share of house price fluctuations. Approximately 0.1 percent of the variations in house prices can be attributed to a total and conventional monetary policy shock and 2.7 percent to an unconventional shock.

These results are confirmed by a historical decomposition of GDP and house prices. As shown in Figure 4.7, contractionary monetary policy shocks played an important role in the development of GDP between the years 2003 and 2005 as well as between 2012 and 2014. By contrast, expansionary monetary policy shocks – in particular unconventional ones – are key factors supporting economic activity in the latter part of the sample. In particular, out of the total increase by 7.7 percent in the (detrended) level of (cross-regional average) GDP between 2013 and 2018, unconventional monetary policy contributed to 39 percent and conventional monetary policy only to 3 percent. House prices are instead affected only to a small extent by monetary policy throughout the sample period and their dynamics are mostly explained by other (non-identified) factors. However, monetary policy plays a larger role in the later years of the sample. Out of the total increase by 5.2 percent in the (detrended) level of (cross-regional average) house prices between 2013 and 2018, unconventional monetary policy

Figure 4.6: Forecast error variance decomposition

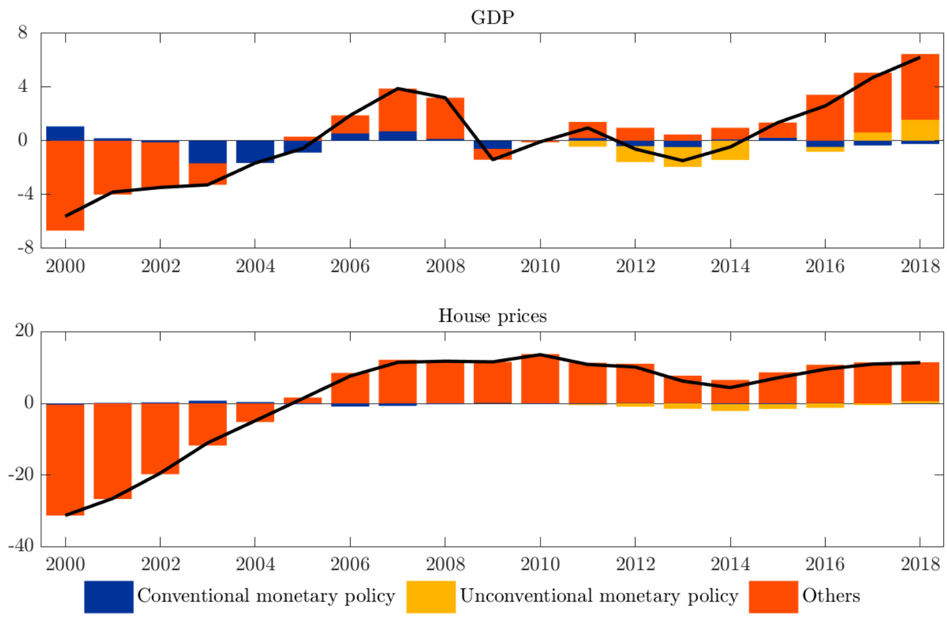


Notes: The y-axis reports the contribution of a total, conventional and unconventional monetary policy shock to variations in GDP, employment and house prices at the 5-year horizon.

contributed to 41 percent and conventional monetary policy induced a negative contribution by about 3 percent.

Overall, our results are in line with the small multipliers of house price changes on consumption typically found in the empirical macroeconomic literature. However, due to our use of a broad measure of economic activity and, hence, the presence of several other channels, our results hint to a less pronounced role for house prices in the transmission of monetary policy compared with other studies. Elbourne (2008) and Ozkan et al. (2017) state that 12-15 percent for the UK and 20 percent for the US of the drop in aggregate consumption after a contractionary interest rate shock can be attributed to changes in house prices. Moreover, Aladangady (2017) and Garbinti et al. (2020) estimate a consumption multiplier of about 5 percent in the US and between 1 and 4 percent across euro area countries, to changes in home values. Both studies report larger responses for households with little wealth, suggesting that looser borrowing constraints are a primary driver of the marginal propensity to consume (MPC) out of housing wealth.

Figure 4.7: Historical decomposition of GDP and house prices



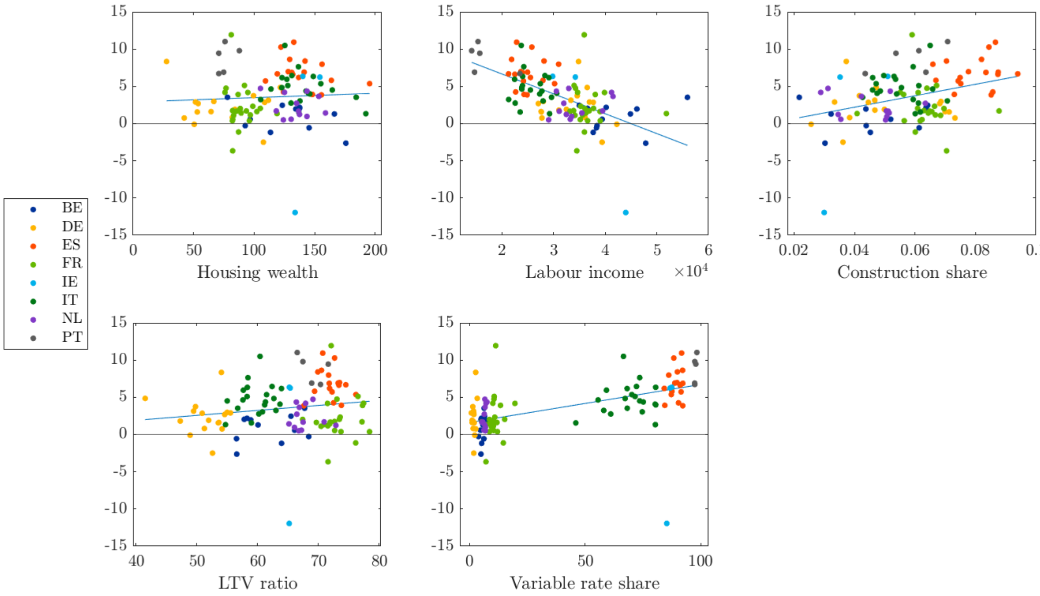
Notes: The y-axis reports the (detrended) level of (cross-regional average) GDP (upper chart) and house prices (lower chart) as well as the contributions of conventional and unconventional monetary policy shocks and other (unidentified) factors.

4.5 The regional heterogeneity of housing markets: An anatomy

A major advantage of the chosen estimation technique applied to our dataset is that it allows us to analyse the heterogeneous response of economic activity and house prices to monetary policy across regions and to link it to several economic and institutional features. To assess the role of the housing channel relative to other relevant channels, we explore how the effectiveness of monetary policy relates to different long-term characteristics across regions, including households' income levels (labour income and housing wealth), the production structure of the economy (in terms of construction and the manufacturing share of total value added),[25] and other key housing-related economic and institutional features, such as households' tenure status (homeownership rate), indebtedness (LTV ratio) and type of mortgages (share of variable-rate mortgages). The relationship between some of these factors (averaged over the sample period) and the estimated monetary policy impact is depicted in Figure 4.8. One can notice that the transmission of monetary policy to the economy is particularly heterogeneous

[25] In fact, sectors producing durable goods are key in the transmission of monetary policy via the user-cost-of-capital and interest-rate channels.

Figure 4.8: Monetary policy impact on real GDP and regional factors



Notes: The y-axis reports the cumulative percentage change in (detrended) levels for GDP 5 years after an accommodative monetary policy shock. The x-axis reports the regional housing wealth (thousand euros per household), labour income (euros per employee, at 2015 prices), construction share (percent of value added), LTV ratio (percent), share of variable-rate loans (percent of total loans). Each dot represents a region.

across euro area regions. This unequal geography of monetary policy transcends the cross-country perspective, as the range of monetary policy effects on GDP spanned by dots of the same colour is wide.[26]

A potentially important driver of the heterogeneous impact of monetary policy across euro area regions is households' income, most notably housing wealth and labour income. A significant relationship between the monetary policy impact and these two variables would allow us to infer whether an easing of monetary policy exacerbates or mitigates regional income inequality. As shown by the weakly positive correlation in the scatter plot in the upper left panel of Figure 4.8, monetary policy appears to be somewhat more effective at stimulating economic activity in regions with higher housing wealth.[27] At the same time, monetary policy seems to be more effective in lower-income regions, given the negative correlation shown in

[26]We focus on the long-term (5-year) impact of monetary policy on real GDP. Note that using a shorter (1-year) horizon would yield qualitatively similar results.

[27]This relationship is stronger when considering the effect of monetary policy on house prices, as shown in Figure 4.12 in Appendix 4.7.

the second panel of Figure 4.8. These results indicate that the ultimate impact of monetary policy on income inequality masks countervailing forces. On the one hand, a loosening of monetary policy may reduce regional inequality by stimulating activity more in regions at the bottom of the labour income distribution. On the other hand, it may also contribute to a larger regional dispersion by supporting activity in regions at the top of the housing wealth distribution. However, housing wealth reflects both the diffusion of wealth across the population (measured by the homeownership rate) as well as the concentration of wealth among owner-occupying households (measured by average house prices). In our econometric analysis below, we formally test the relative importance of each driver of housing wealth.

Moreover, we investigate the relationship between the impact of monetary policy and three further dimensions of the housing market. First, we consider the production structure of the economy and explore how the region-specific construction intensity, measured by the share of construction value added in total value added, affects the effectiveness of monetary policy. As shown in Figure 4.8, the share of the construction sector relative to the overall economy is positively correlated with the impact of monetary policy on real economic activity.[28]

Second, we investigate how households' indebtedness relates to the impact of monetary policy. Figure 4.8 suggests that the level of indebtedness, measured by the LTV ratio, is only weakly correlated with the impact of monetary policy across euro area regions.[29]

Third, the diverse impact of monetary policy across regions can be given by heterogeneous mortgage market characteristics, such as the share of variable-rate mortgages. In countries where most mortgages have adjustable rates, policy-induced changes in interest rates have an almost immediate effect on household cash flows. As illustrated in the last panel of Figure 4.8, the impact of monetary policy on GDP is indeed larger in regions with a higher share of variable-rate loans. These regions are concentrated in Italy, Spain, Ireland and Portugal. This result is in line with the model simulations by Calza, Monacelli and Stracca (2013), who document a stronger impact of monetary policy on consumption in

[28]This suggests a role for the construction sector in conveying monetary policy shocks to the overall economy, in line with evidence on the user-cost-of-capital and interest-rate channels of monetary policy in affecting the production of durable and capital goods (Dedola and Lippi, 2005; Peersman and Smets, 2005).

[29]The positive relationship with the LTV ratio at the regional level is consistent with the evidence pointing to a different transmission of monetary policy for liquidity-constrained and non-constrained households (Aladangady, 2017, Guerrieri and Iacoviello, 2017). By including an endogenously estimated threshold variable (i.e. the LTV ratio at the regional level) in our baseline model, we find indeed a non-linear transmission mechanism for monetary policy on housing and macroeconomic variables, with a significantly stronger impact when the LTV ratio is above a certain level. The results are available from the authors upon request.

those countries where mortgage contracts are predominantly of the variable-rate type, and Pica (2022), who finds that a higher share of adjustable-rate mortgages and a higher homeownership rate interact to amplify the effects of monetary policy on economic activity in the euro area. However, given the decrease in the share of variable-rate mortgages observed over the second half of the sample period (especially in those countries where variable-rate contracts are traditionally prevailing), homeowners' interest-rate sensitivity fell in recent years (see, for example, Bech and Mikkelsen, 2021).

We carry out a formal analysis in order to shed more light on the link between the monetary policy effectiveness and economic and institutional characteristics across euro area regions. Besides the variables mentioned above, we include controls commonly found to be important determinants of the transmission of monetary policy to the business cycle, such as the manufacturing share of value added and a measure of lending activity to households. Panel (a) of Table 4.2 reports the results of various regression specifications that link our estimated long-term impact of total monetary policy shocks on real GDP to the key variables discussed above. In the most parsimonious specifications, the regression coefficients of these variables have the expected sign (as in the graphical overview discussed above) and are found to be statistically significant, except for housing wealth and lending activity. The significance is robust to the inclusion of demographic factors. When housing wealth is replaced by its determinants, the homeownership rate is estimated to play a significant role.[30] When all variables are considered, labour income, the share of construction, the share of manufacturing and lending activity display a statistically significant coefficient. For labour income and the share of manufacturing the coefficient remains significant even after the inclusion of country and country-group dummies.[31] Similar findings are observed when considering the impact of conventional monetary policy (panel (b) in Table 4.2), except that the share of manufacturing is no longer significant. Focusing on unconventional monetary policy (panel (c) in Table 4.2),

[30] This result relates to the work by Paz-Pardo (2021), who shows that increases in labour income inequality and uncertainty are key drivers for a decrease in homeownership among younger households in several major advanced economies, suggesting that the evolution of homeownership rates is closely intertwined with labour markets, housing markets and financial conditions.

[31] The *Vulnerable* dummy variable splits the regions into two large groups according to a conventional assessment of "vulnerability". In the academic and policy literature, this assessment typically considers a certain type of macroeconomic imbalances, such as government debt-to-GDP ratios and current account deficits, and implies a division between more and less vulnerable countries (sometimes also referred to as periphery and core countries, respectively). The more vulnerable group contains all regions in Spain, Ireland, Italy and Portugal, and the less vulnerable group consists of all regions in Belgium, Germany, France and the Netherlands.

lending activity (proxied by the product of regional average house prices and LTV ratios) becomes statistically significant. This confirms the role of bank lending in supporting the effectiveness of (unconventional) measures and thus restoring the functioning of the monetary policy transmission mechanism after the Sovereign Debt Crisis (for more details, see Altavilla et al., 2019*a* and Adalid and Falagiarda, 2020).

We perform the same exercise considering the impact of monetary policy on house prices as dependent variable (Table 4.5 in Appendix 4.7). Besides confirming the importance of labour income, the results of these regressions highlight the role of housing wealth in the propagation of monetary policy, particularly in the case unconventional monetary policy shocks.

Overall, as the coefficient on compensation per employee remains significant across all specifications, our findings point to the effectiveness of monetary policy in reducing regional inequality by stimulating economic activity more in regions with lower labour income. Together with the absence of a clear predominance of one of the two determinants of housing wealth (diffusion of owner-occupying housing and home valuations), this suggests that monetary policy easing has an overall beneficial impact on cross-regional inequality.

Our results add to a growing literature on monetary policy and inequality. Most contributions examine the issue at the household or individual level. Some studies find that expansionary monetary policy can mitigate income inequality as lower-income households disproportionately benefit from positive effects via the stimulus to economic activity and employment, which outweigh those via financial markets (for the US, see Coibion et al., 2017; for the euro area, see Casiraghi et al., 2018, Lenza and Slacalek, 2021 and Altavilla et al., 2021). This stands in contrast to Amberg et al. (2021), who show that the income response to monetary policy in Sweden is U-shaped, and to Andersen et al. (2020), who find that monetary easing in Denmark raises income shares at the top of the income distribution while reducing them at the bottom, hence leading to higher income inequality. The impact of monetary policy on wealth inequality is also a subject of debate. Lenza and Slacalek (2021) state that monetary policy has only a negligible impact on wealth inequality. A U-shaped response of wealth inequality is found by Casiraghi et al. (2018), while according to Andersen et al. (2020) monetary easing is more beneficial to the net wealth of higher income households, thereby increasing wealth inequality.

Little attention has been given to the geographical dimension of inequality and how it is affected by monetary policy. An outstanding exception is the work by Hauptmeier, Holm-Hadulla and Nikalexi (2020), who focus on the heterogeneity of the impact of monetary policy across euro area regions. The authors find that monetary easing shocks have a significantly more pronounced and persistent effect on output in poorer than in richer regions, implying a mitigation of regional inequality. Besides confirming this result, our study differentiates between income sources, i.e. housing wealth and labour income. Focusing on the US, Beraja et al. (2019) examine the transmission of monetary policy via mortgage markets at the regional level. In contrast to previous recessions, they find that, during the Global Financial Crisis, depressed regions reacted less to interest rate cuts, thus increasing regional consumption inequality.

Table 4.2: Relationship between monetary policy impact on real GDP and regional factors

(a) Dependent variable: *Impact of TMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-4.934***	-4.316***				-4.253***	-4.470***	-4.060**
Housing wealth	0.639					-1.011	-0.733	0.704
Homeownership rate		0.028**						
House price level		0.299						
Share of construction in GVA			0.581***			0.214*	0.200*	0.014
Share of manufacturing in GVA			0.063**			0.057**	0.057**	0.069***
Share of variable-rate mortgages				0.027***		0.002	0.013	0.020
Lending activity					0.598	1.511*	1.328	-0.437
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.424	0.439	0.189	0.324	0.015	0.494	0.501	0.538
(b) Dependent variable: *Impact of CMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-3.903***	-3.932***				-3.494***	-3.412***	-3.747*
Housing wealth	0.187					1.141	1.036	1.612
Homeownership rate		0.006						
House price level		0.424						
Share of construction in GVA			0.560***			0.373***	0.378***	-0.022
Share of manufacturing in GVA			0.038			0.027	0.026	0.042
Share of variable-rate mortgages				0.018***		-0.003	-0.007	0.006
Lending activity					-0.094	-0.579	-0.510	-0.650
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.268	0.270	0.172	0.149	0.014	0.322	0.323	0.451
(c) Dependent variable: *Impact of UMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-1.874**	-1.934**				-3.792***	-4.051***	-4.741**
Housing wealth	1.010					-1.701	-1.368	-1.316
Homeownership rate		0.014						
House price level		1.214						
Share of construction in GVA			0.114			-0.090	-0.106	-0.134
Share of manufacturing in GVA			0.060**			0.075***	0.076***	0.076**
Share of variable-rate mortgages				0.010**		-0.014	-0.001	-0.006
Lending activity					1.032**	3.240***	3.021**	1.770
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.116	0.115	0.079	0.085	0.076	0.217	0.227	0.251

Notes: The table present regressions of the cumulative monetary policy impact on real GDP at the regional level (as estimated in section 4.4) on regional factors (compensation per employee in logs, housing wealth in logs, homeownership rate in percent, the average house price level in logs, the share of construction and manufacturing in GVA, the share of variable-rate mortgages in percent, and a proxy for lending activity). Housing wealth is computed as the product of the homeownership rate and the average house price level. The proxy for lending activity is computed as the product of housing wealth and the LTV ratio. Demographics controls include total employment and population density at the regional level. The Vulnerable dummy is a binary variable that takes value one for regions of Italy, Spain, Portugal and Ireland, and zero for regions of Germany, France, the Netherlands and Belgium. A constant is included. An outlier is excluded. *** p <0.01, ** p<0.05, * p<0.1

4.6 Robustness Checks

4.6.1 Additional common components

To check the robustness of our findings, we first extend the set of exogenous variables in our baseline VAR model. As shown, for example, by Vansteenkiste and Hiebert (2011) and Campos, Fidrmuc and Korhonen (2019), there are significant interlinkages among regional housing markets and business cycles in the euro area. Hence, the set of exogenous variables, which in the baseline model only includes the monetary policy shocks, is expanded to include the euro area GDP, employment and house prices. Following Chudik and Pesaran (2015), these euro area variables are calculated as cross-sectional means of all the regions within our dataset, namely $Y_t^* = N^{-1} \sum_{i=1}^{N} Y_{i,t}$, where $Y_{i,t}$ denotes the vector of endogenous variables in our SPVAR model defined in Equation (4.3). Insofar as these variables are endogenous to monetary policy changes, they incorporate to some extent our monetary policy shock. To avoid double-counting, we first regress the cross-sectional averages of GDP, employment and house prices on total monetary policy shocks. Formally, we posit the following linear relation between common components and total monetary policy shock:[32]

$$Y_t^* = \Omega_0 + \Omega_1 TMP_t + \omega_t \tag{4.9}$$

where $\omega_t \sim N(0, \sigma_\omega)$. The non-monetary policy common components are then extracted by subtracting the product of the estimated coefficient $\hat{\Omega}_1$ and the total monetary policy shock from the cross-sectional averages, namely $\tilde{Y}_t^* = Y_t^* - \hat{\Omega}_1 TMP_t$. Finally, we introduce these non-monetary policy common components as additional exogenous regressors in the SPVAR by augmenting the vector $X_t = [MP_t, \tilde{Y}_{t-d}^*]$ where MP_t denotes TMP_t or $[CMP_t, UMP_t]$.[33]

When including these additional exogenous variables, the results of the baseline SPVAR model estimation are broadly confirmed (Figure 4.13 in Appendix 4.7). An accommodative monetary policy shock has a positive impact on GDP and employment. The impact on house prices is initially negative, albeit insignificant, and fades to zero subsequently. Unlike

[32] We do not perform this regression on conventional and unconventional monetary policy shocks, since their combined information corresponds to the one contained in the total monetary policy shock.

[33] For the purpose of our analysis, we assume a delay parameter $d = 1$, aligning the timing of non-monetary policy common components with the lagged endogenous variables. Note that this approach differs from the common correlated effect (CCE) estimator proposed by Chudik and Pesaran (2015). However, the CCE estimator would not suit our purposes because it would only allow us to retrieve the coefficients of region-specific variables.

in the baseline, an unconventional monetary policy shock has a larger impact on GDP and employment than a conventional monetary policy shock. As in the baseline specification, an unconventional shock has a larger and statistically significant impact on house prices compared to a conventional one.

4.6.2 A pooled fixed-effects estimator

In order to check the robustness of our mean-group estimates, a pooled OLS regression is applied to the demeaned regional dataset, resulting in a fixed-effects estimator. Figure 4.14 in Appendix 4.7 displays the impulse response functions to an accommodative monetary policy shock under this specification. In line with the mean-group estimation results, the impact of a monetary policy easing shock on GDP, employment and house prices is positive, but slightly larger in size. In addition, the impact on house prices is statistically significant.

4.6.3 An alternative structural identification strategy

The estimated contributions from the housing and the employment channel in our benchmark SPVAR model depend on the ordering of the endogenous variables. As the contemporaneous contributions tend to assign a larger weight to the less "reactive" (or more exogenous) variables, we consider the estimates from our benchmark SPVAR model as an upper bound of the contribution of the employment and, especially, the housing channels. In fact, both theoretical and empirical arguments would suggest an alternative ordering to model the contemporaneous relationships among the endogenous variables in our benchmark SPVAR model.

On theoretical grounds, asset prices are typically placed as the most endogenous variables, as they are highly sensitive to contemporaneous and expected economic news or shocks (see, for instance, Stock and Watson, 2016). Moreover, employment typically lags GDP, as labour market frictions impede an immediate adjustment to the business cycle (Mortensen and Pissarides, 1994). In line with these considerations, there are papers in the literature imposing a recursive structure in VARs in which house prices react to GDP in the same period (Nocera and Roma, 2017, Musso, Neri and Stracca, 2011, Giuliodori, 2005).

From an empirical perspective, pairwise Granger (1969) causality tests on comparable euro area aggregate data at quarterly frequency confirm these theoretical predictions. According to the results of the tests, shown in Table 4.6 in Appendix 4.7, GDP (Granger) causes both employment and house prices. Employment causes only house prices, while house prices cause neither GDP nor employment. Hence, as a robustness exercise, we invert our preferred ordering and consider GDP as the most exogenous variable and house prices as the most endogenous one. This alternative ordering implies nil contemporaneous contributions from the housing and the employment channels, which appear restrictive assumptions, especially at annual frequency.

Figure 4.15 in Appendix 4.7 shows the variance decomposition when we order the endogenous variables as follows: $Y_{i,t} = [\text{House prices}_{i,t}, \text{Employment}_{i,t}, \text{GDP}_{i,t}]$. Confirming our results, the contribution of unconventional monetary policy shocks to the variation in GDP and employment is more than three times larger than the contribution of conventional shocks. Variations in house prices can be explained by the different monetary policy shocks to a much smaller extent. Moreover, this alternative ordering confirms our results on the limited role of the house price channel as a conveyor of monetary policy shocks to economic activity.[34]

4.7 Conclusion

By means of a structural panel VAR estimated with novel regional data, this paper investigates the role of the housing market in the transmission of conventional and unconventional monetary policy in the euro area. We show that the housing channel plays a limited role in the propagation of monetary policy to the economy, but its contribution is amplified in the case of unconventional monetary policy.

The transmission of monetary policy to the economy is found to be heterogeneous across regions, with a larger impact in areas with lower labour income and higher homeownership rates. This suggests that poorer regions stand to benefit the most from monetary policy accommodation. While the easing of monetary policy is found to mitigate regional inequality through its stimulus to the economy, the unintended consequences of the ongoing monetary policy normalisation warrant close monitoring by policy-makers, particularly in the case of resurgent fragmentation risks.

[34]Results for the role of the employment and housing channels from the alternative ordering are available from the authors upon request.

Appendix

House prices at the regional level: The ED database

Regional house prices are derived from the loan-level database of the European DataWarehouse (ED), a securitisation repository that collects, validates and makes available detailed, standardised and asset class-specific loan-level data for asset-backed securities (ABS) transactions.[35] The data are collected in the context of the ABS loan-level initiative, which establishes specific loan-by-loan information requirements for ABS accepted as collateral in Eurosystem credit operations. This initiative was launched in 2012 and aimed to improve transparency in ABS markets and facilitate the risk assessment of these instruments used by Eurosystem counterparties as collateral in monetary policy operations. Banks are required to submit at least at quarterly frequency detailed information regarding the loans backing the ABS, including loan, borrower and collateral characteristics.

For the purpose of this analysis, we only consider the loans underlying residential mortgage-backed securities (RMBS). The reporting templates are populated with information on the loan (e.g. original and outstanding balance, date of origination, maturity, purpose, interest rate, repayment type, performance), the borrower (e.g. employment status, annual gross income, age), and the property (e.g. valuation, property type, geographic location—with the first two digits of the postcode typically available). These fields can be either static (reported at origination) or dynamic (updated at each submission), as well as mandatory (always populated) or optional (whereby missing values can be found). Eight euro area countries are covered in the ED database: Germany, France, Italy, Spain, the Netherlands, Belgium, Portugal and Ireland.

The raw data are processed and cleaned as follows. First, imputation techniques are used for the main static variables whenever we observe for each loan (i) missing values in one or more submissions; (ii) inconsistent values across submissions. This imputation procedure allows us to keep a large number of loans that would have otherwise been discarded and therefore to increase the coverage of the sample. Second, we drop outliers by considering only loans used for the purchase of a property with a price below EUR 5 mn and above EUR

[35] ED loan-level data has been used by Ertan, Loumioti and Wittenberg-Moerman (2017), Amzallag et al. (2019), Gianinazzi, Pelizzon and Plazzi (2018), van Bekkum, Gabarro and Irani (2018), Gaudêncio, Mazany and Schwarz (2019), Kang, Loumioti and Wittenberg-Moerman (2020), Klein, Mössinger and Pfingsten (2021).

10,000. Third, we exclude loans with missing information on the key variables used in the analysis. Third, as multiple loans can be used to purchase the same property, especially in the Netherlands, we aggregate loans originated at the same time by a single borrower for the purchase of a single property, as in Gianinazzi, Pelizzon and Plazzi (2018). The summary statistics of some of the key variables included in the cleaned loan-level dataset are reported in Table 4.3.

Table 4.3: Summary statistics of the ED dataset (over the period 1999-2018)

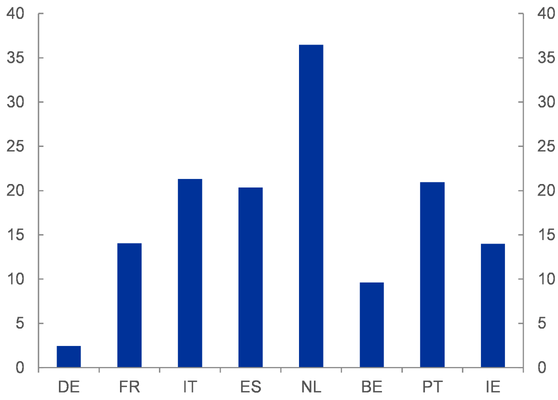
	DE	FR	IT	ES	NL	BE	PT	IE
Number of loans (in thousand)	687.2	3381.6	1814.6	1886.6	2799.6	1125.9	496.6	291.6
Loan size (median, in EUR thousand)	87.3	87	100	120	160.4	100	68.6	180
Maturity (median, in years)	20	17	20	30	30	19.3	30.4	25
Share of fixed-rate loans (in %)	98.0	89.3	27.2	10.8	93.5	94.5	2.4	14.1
Borrower's income (median, in EUR thousand)	43.5	37.4	25.1	27.5	50	48.1	17.1	54.4
Property valuation (median, in EUR thousand)	183	137.2	170	177.5	238.4	175	109.7	260

A graphical illustration of the coverage of the dataset is provided in Figure 4.9. The overall volume of the loans in our dataset is a significant share of total loan origination in all countries, except Germany. This is due to the fact that mortgages in this country are much more commonly pooled into covered bonds than RMBS. The coverage varies significantly over time in all countries in our sample, reaching a peak in the aftermath of the Global Financial Crisis, when banks started to retain securitised products on their balance sheets in order to use them as collateral for Eurosystem's credit operations. The coverage of our data has decreased thereafter, reflecting the contraction in the securitisation markets observed in many euro area countries and the concomitant pick-up in mortgage credit.

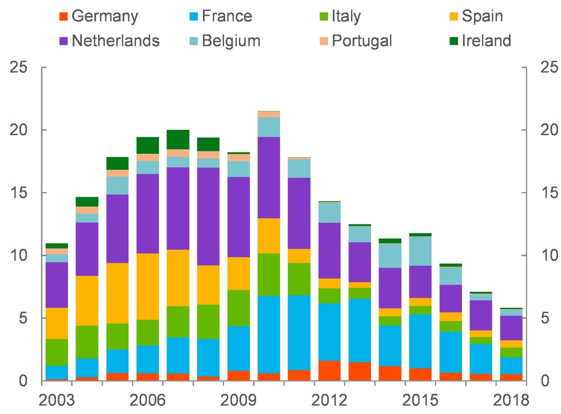
The property valuation contained in the ED data is used to derive house price indexes for euro area regions and countries. The resulting country aggregates are then compared with the correspondent official series (Figure 4.10). A graphical inspection of the two series shows that the implied house price indexes closely resemble the official ones for all countries, suggesting that our sample is well representative of house price dynamics at the national level.

A similar exercise is conducted for mortgage rates in order to check whether our data is representative of credit dynamics. As the ED database does not provide information on the interest rate of floating-rate mortgages at origination, this exercise can be only performed for countries where fixed-rate mortgages have been more popular over the sample period (i.e. Germany, France, the Netherlands and Belgium). The implied mortgage rates closely

Figure 4.9: Share of mortgage loans covered by ED data (percent)



(a) Over the full period 1999-2018

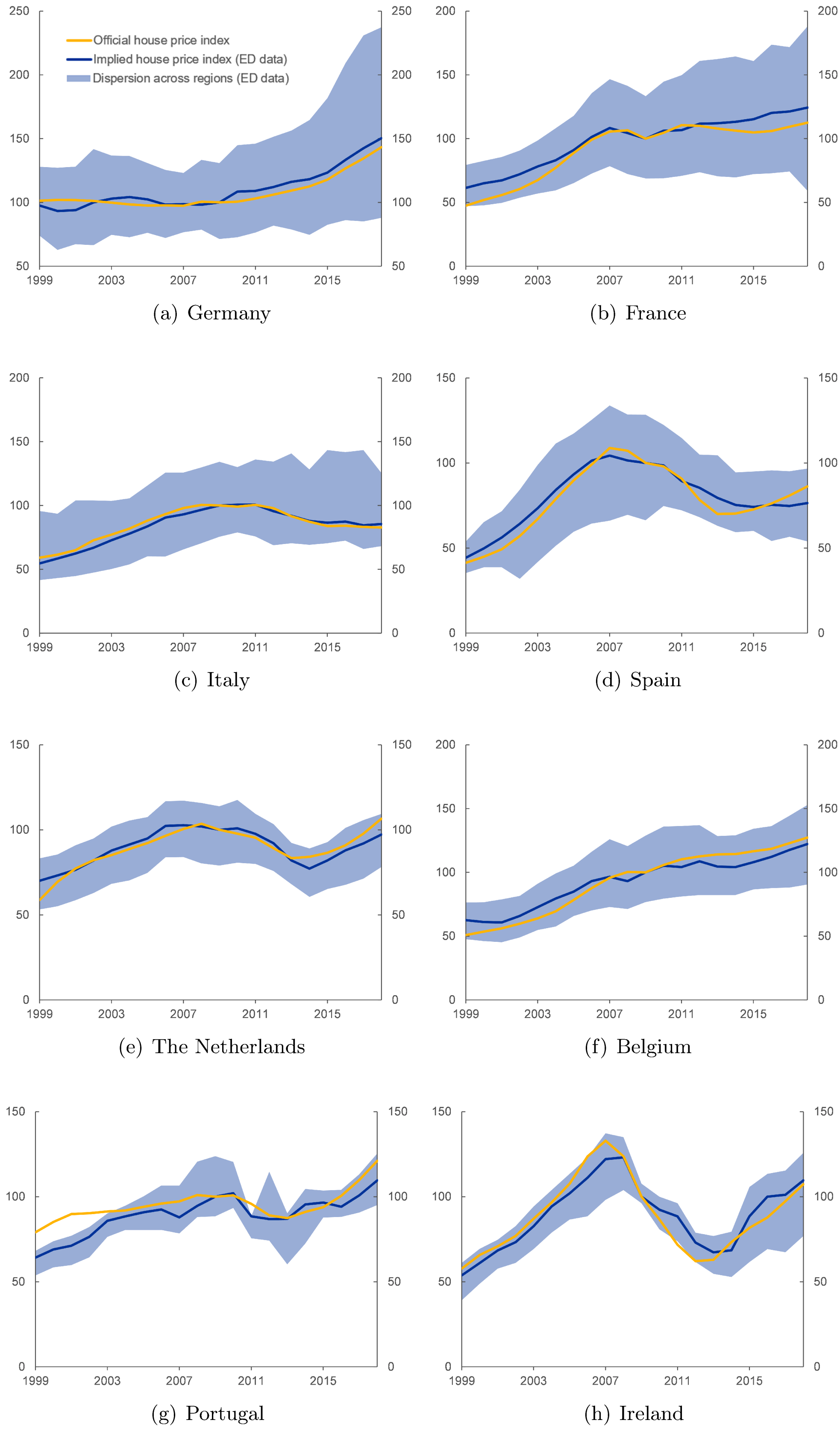


(b) By year of origination

Notes: Sum of original balance of loans of the ED dataset over total new business volumes from the MFI Interest Rate Statistics of the ECB.

follow the official country rates over time (Figure 4.11). The results also point to very similar developments across regions, suggesting that bank lending policies tend to be uniform within a country.

Figure 4.10: House price indexes (2009=100)



(a) Germany

(b) France

(c) Italy

(d) Spain

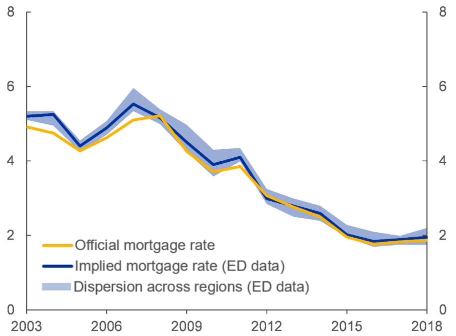
(e) The Netherlands

(f) Belgium

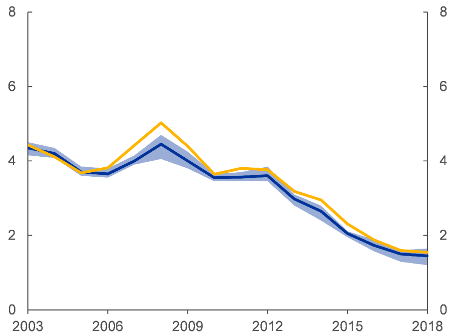
(g) Portugal

(h) Ireland

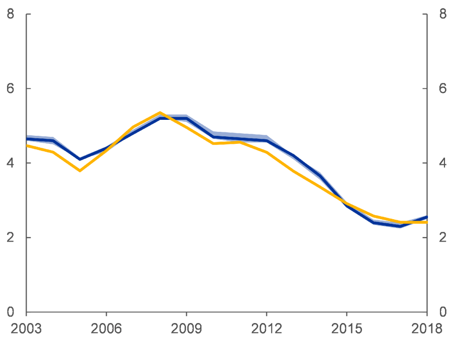
Figure 4.11: Mortgage rates (in percentages per annum)



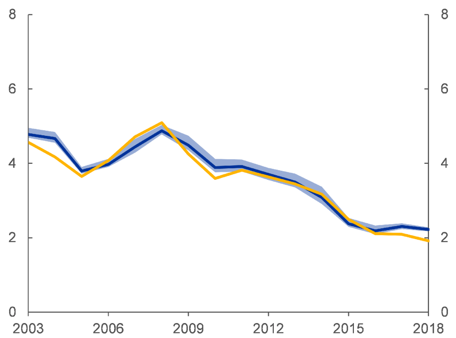
(a) Germany



(b) France



(c) The Netherlands



(d) Belgium

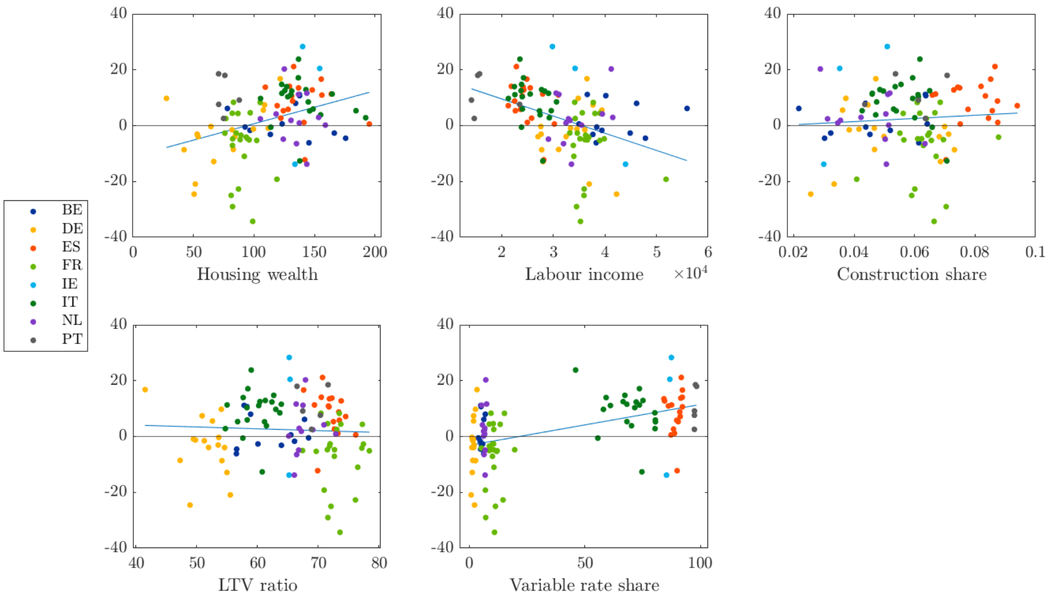
Additional tables and charts

Table 4.4: Summary Statistics over sub-periods

		Mean	Median	Minimum	Maximum	Standard Deviation
GDP	1999-2008	28923	28021	13786	66418	8844
GDP	2009-2012	29438	28133	14070	65112	9233
GDP	2013-2018	30394	28307	14914	65178	10368
Employment	1999-2008	43.86	43.08	31.31	65.43	6.68
Employment	2009-2012	43.48	42.78	31.55	66.83	7.06
Employment	2013-2018	43.33	42.24	30.15	68.23	7.36
House prices	1999-2008	132.89	129.02	93.25	187.27	22.32
House prices	2009-2012	163.03	164.57	93.44	233.93	34.34
House prices	2013-2018	156.51	156.57	104.69	237.29	26.20

Notes: Real GDP and employment are given in per capita terms. National GDP and employment are calculated as cross-regional aggregate of all regions within a country. National house prices are given by GDP-weighted cross-regional means of all regions within a country.

Figure 4.12: Monetary policy impact on house prices and regional factors



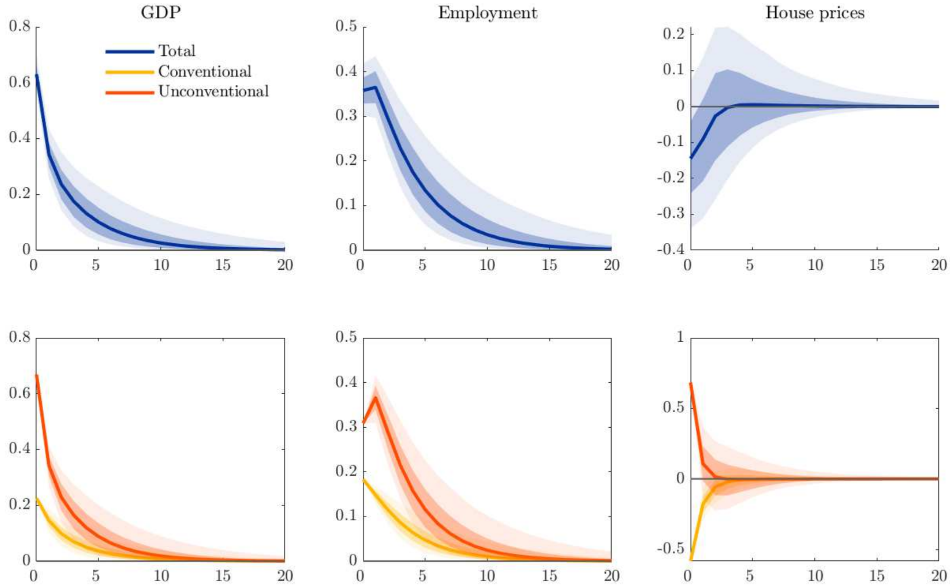
Notes: The y-axis reports the cumulative percentage change in (detrended) levels for house prices 5 years after an accommodative monetary policy shock. The x-axis reports the regional housing wealth (thousand euros per household), labour income (euros per employee, at 2015 prices), construction share (percent of value added), LTV ratio (percent), share of variable-rate loans (percent of total loans). Each dot represents a region.

Table 4.5: Relationship between monetary policy impact on house prices and regional factors

(a) Dependent variable: *Impact of TMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-12.237***	-10.965***				-11.144***	-10.637***	-17.825**
Housing wealth	7.749***					7.839**	7.189*	8.024
Homeownership rate		0.158***						
House price level		7.093***						
Share of construction in GVA			0.406			-0.297	-0.265	-0.433
Share of manufacturing in GVA			0.080			0.020	0.019	-0.004
Share of variable-rate mortgages				0.079***		0.015	-0.011	-0.057
Lending activity					4.500***	-1.112	-0.687	-5.339
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.331	0.332	0.013	0.243	0.066	0.339	0.343	0.407
(b) Dependent variable: *Impact of CMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-5.918**	-1.797				0.284	0.752	-21.486**
Housing wealth	-1.780					-5.839	-6.439	9.553
Homeownership rate		0.080						
House price level		-7.738**						
Share of construction in GVA			0.943*			0.325	0.354	-1.407
Share of manufacturing in GVA			0.101			0.080	0.078	0.046
Share of variable-rate mortgages				0.039**		0.046	0.022	-0.044
Lending activity					-0.633	1.396	1.789	-8.752
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.067	0.109	0.052	0.062	0.023	0.098	0.100	0.187
(c) Dependent variable: *Impact of UMP shock*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Compensation per employee	-5.259**	-10.905***				-19.487***	-19.834***	-10.680
Housing wealth	6.996***					10.942***	11.387***	0.549
Homeownership rate		-0.037						
House price level		15.781***						
Share of construction in GVA			-0.747*			-0.823*	-0.845*	-0.403
Share of manufacturing in GVA			0.063			0.077	0.078	0.084
Share of variable-rate mortgages				0.011		-0.104***	-0.086*	-0.073
Lending activity					3.001*	1.364	1.072	3.703
Demographics controls	✓	✓	✓	✓	✓	✓	✓	✓
Vulnerable dummy	-	-	-	-	-	-	✓	-
Country dummies	-	-	-	-	-	-	-	✓
Observations	105	105	105	105	105	105	105	105
R-squared	0.171	0.301	0.071	0.036	0.061	0.342	0.344	0.418

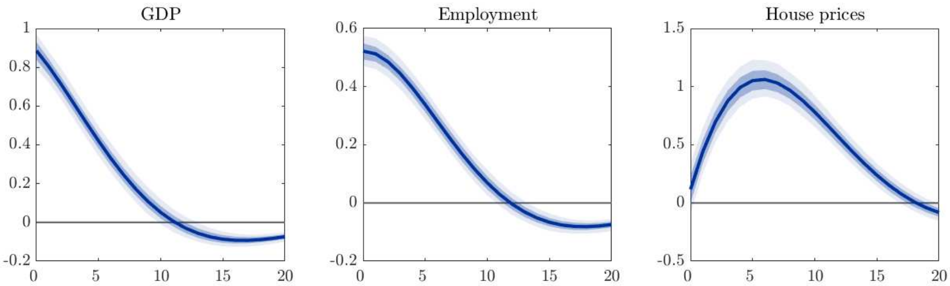
Notes: The table present regressions of the cumulative monetary policy impact on house prices at the regional level (as estimated in section 4.4) on regional factors (compensation per employee in logs, housing wealth in logs, homeownership rate in percent, the average house price level in logs, the share of construction and manufacturing in GVA, the share of variable-rate mortgages in percent, and a proxy for lending activity). Housing wealth is computed as the product of the homeownership rate and the average house price level. The proxy for lending activity is computed as the product of housing wealth and the LTV ratio. Demographics controls include total employment and population density at the regional level. The Vulnerable dummy is a binary variable that takes value one for regions of Italy, Spain, Portugal and Ireland, and zero for regions of Germany, France, the Netherlands and Belgium. A constant is included. An outlier is excluded. *** p <0.01, ** p<0.05, * p<0.1

Figure 4.13: Impulse response functions to an expansionary monetary policy shock - common components



Notes: The y-axis reports the percentage change in (detrended) levels of each variable over the considered horizon. The x-axis reports the years. This specification includes non-monetary policy common components. Solid lines denote point estimates and light (dark) shaded areas 95 percent (68 percent) confidence bands.

Figure 4.14: Impulse response functions to an expansionary monetary policy shock - pooled fixed-effect estimator



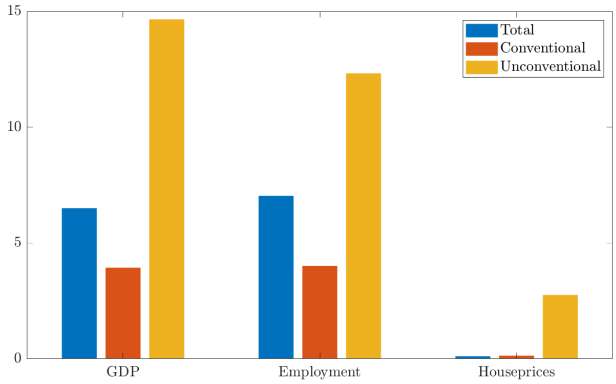
Notes: The y-axis reports the percentage change in (detrended) levels of each variable over the considered horizon. The x-axis reports the years. These are the results of a fixed-effects regression. Solid lines denote point estimates and light (dark) shaded areas 95 percent (68 percent) confidence bands.

Table 4.6: Granger causality test results

	GDP	Employment	House Prices
GDP	/	0.000	0.005
Employment	0.269	/	0.010
House prices	0.143	0.587	/

Notes: The table shows the p-values of a Granger causality test. If the value in row i and column j is smaller than 0.01 (0.05), then the null hypothesis that variable i does not Granger cause variable j has to be rejected at the 1% (5%) significance level.

Figure 4.15: Variance decomposition of key variables - alternative ordering



Notes: The y-axis reports the contribution of a total, conventional and unconventional monetary policy shock to variations in GDP, employment and house prices at the 5-year horizon.

References

Adalid, Ramón, and Matteo Falagiarda. 2020. "How Repayments Manipulate Our Perceptions about Loan Dynamics after a Boom." *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 240(6): 697–742.

Adam, Klaus, and Michael Woodford. 2021. "Robustly optimal monetary policy in a new Keynesian model with housing." *Journal of Economic Theory*, 198: 105352.

Aladangady, Aditya. 2017. "Housing Wealth and Consumption: Evidence from Geographically-Linked Microdata." *American Economic Review*, 107(11): 3415–46.

Altavilla, Carlo, Desislava C. Andreeva, Miguel Boucinha, and Sarah Holton. 2019*a*. "Monetary policy, credit institutions and the bank lending channel in the euro area." European Central Bank Occasional Paper Series 222.

Altavilla, Carlo, Luca Brugnolini, Refet S. Gürkaynak, Roberto Motto, and Giuseppe Ragusa. 2019*b*. "Measuring euro area monetary policy." *Journal of Monetary Economics*, 108(C): 162–179.

Altavilla, Carlo, Wolfgang Lemke, Tobias Linzert, Jens Tapking, and Julian von Landesberger. 2021. "Assessing the efficacy, efficiency and potential side effects of the ECB's monetary policy instruments since 2014." European Central Bank Occasional Paper Series 278.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. 2021. "Five Facts about the Distributional Income Effects of Monetary Policy." CESifo Working Paper Series 9062.

Amisano, Gianni, and Carlo Giannini. 1997. *Topics in Structural VAR Econometrics. MIT Press Books.* 2nd ed., Springer-Verlag, Heidelberg.

Amzallag, Adrien, Alessandro Calza, Dimitris Georgarakos, and João Sousa. 2019. "Monetary policy transmission to mortgages in a negative interest rate environment." European Central Bank Working Paper Series 2243.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydró. 2020. "Monetary policy and inequality." Department of Economics and Business, Universitat Pompeu Fabra Economics Working Papers 1761.

Battistini, Niccolò, Matteo Falagiarda, Johannes Gareis, Angelina Hackmann, and Moreno Roma. 2021. "The euro area housing market during the COVID-19 pandemic." *Economic Bulletin Articles*, 7.

Bech, Stine Ludvig, and Jakob Guldbæk Mikkelsen. 2021. "Measuring household interest-rate sensitivity in Denmark." Danmarks Nationalbank Working Papers 183, Copenhagen:Danmarks Nationalbank.

Beetsma, Roel, Jacopo Cimadomo, and Josha Van Spronsen. 2021. "One Scheme Fits All: A Central Fiscal Capacity for the EMU Targeting Eurozone, National and Regional Shocks." C.E.P.R. Discussion Papers 16829.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *The Quarterly Journal of Economics*, 134(1): 109–183.

Calza, Alessandro, Tommaso Monacelli, and Livio Stracca. 2013. "Housing Finance And Monetary Policy." *Journal of the European Economic Association*, 11: 101–122.

Campos, Nauro F., Jarko Fidrmuc, and Iikka Korhonen. 2019. "Business cycle synchronisation and currency unions: A review of the econometric evidence using meta-analysis." *International Review of Financial Analysis*, 61(C): 274–283.

Casiraghi, Marco, Eugenio Gaiotti, Lisa Rodano, and Alessandro Secchi. 2018. "A "reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households." *Journal of International Money and Finance*, 85(C): 215–235.

Chudik, Alexander, and M. Hashem Pesaran. 2015. "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors." *Journal of Econometrics*, 188(2): 393–420.

Cœuré, Benoît. 2018. "The local impact of the ECB's monetary policy." European Central Bank Speech, 4 October 2018.

Cœuré, Benoît. 2019. "Heterogeneity and the ECB's monetary policy." European Central Bank Speech, 29 March 2019.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary policy and inequality." *Journal of Monetary Economics*, 88(C): 70–89.

Corsetti, Giancarlo, Joao B. Duarte, and Samuel Mann. 2020. "One Money, Many Markets: Monetary Transmission and Housing Financing in the Euro Area." International Monetary Fund IMF Working Papers 2020/108.

Dedola, Luca, and Francesco Lippi. 2005. "The monetary transmission mechanism: Evidence from the industries of five OECD countries." *European Economic Review*, 49(6): 1543–1569.

Elbourne, Adam. 2008. "The UK housing market and the monetary policy transmission mechanism: An SVAR approach." *Journal of Housing Economics*, 17(1): 65–87.

Enders, Zeno, Franziska Hünnekes, and Gernot J. Müller. 2019. "Monetary policy announcements and expectations: Evidence from german firms." *Journal of Monetary Economics*, 108(C): 45–63.

Ertan, Aytekin, Maria Loumioti, and Regina Wittenberg-Moerman. 2017. "Enhancing Loan Quality Through Transparency: Evidence from the European Central Bank Loan Level Reporting Initiative." *Journal of Accounting Research*, 55(4): 877–918.

European Central Bank. 2022. "Financial Integration and Structure in the Euro Area." European Central Bank Occasional Reports.

European Commission. 2021. "Euro Area Housing Markets: Trends, Challenges and Policy Responses." European Commission Discussion Paper 147.

Falagiarda, Matteo, and Stefan Reitz. 2015. "Announcements of ECB unconventional programs: Implications for the sovereign spreads of stressed euro area countries." *Journal of International Money and Finance*, 53: 276–295.

Gambacorta, Leonardo, Boris Hofmann, and Gert Peersman. 2014. "The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis." *Journal of Money, Credit and Banking*, 46(4): 615–642.

Garbinti, Bertrand, Pierre Lamarche, Frédérique Savignac, and Charlélie Lecanu. 2020. "Wealth effect on consumption during the sovereign debt crisis: households heterogeneity in the euro area." European Central Bank Working Paper Series 2357.

Gaudêncio, João, Agnieszka Mazany, and Claudia Schwarz. 2019. "The impact of lending standards on default rates of residential real estate loans." European Central Bank Occasional Paper Series 220.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics*, 7(1): 44–76.

Gianinazzi, Virginia, Loriana Pelizzon, and Alberto Plazzi. 2018. "Residential mortgage defaults and positive equity: Lessons from europe." USI Working Paper.

Giuliodori, Massimo. 2005. "The Role Of House Prices In The Monetary Transmission Mechanism Across European Countries." *Scottish Journal of Political Economy*, 52(4): 519–543.

Granger, Clive. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica*, 37(3): 424–38.

Guerrieri, Luca, and Matteo Iacoviello. 2017. "Collateral constraints and macroeconomic asymmetries." *Journal of Monetary Economics*, 90(C): 28–49.

Hauptmeier, Sebastian, Fédéric Holm-Hadulla, and Katerina Nikalexi. 2020. "Monetary policy and regional inequality." European Central Bank Working Paper Series 2385.

Holm-Hadulla, Fédéric, and Claire Thürwächter. 2021. "Heterogeneity in corporate debt structures and the transmission of monetary policy." *European Economic Review*, 136(C).

Huber, Florian, and Maria Teresa Punzi. 2020. "International Housing Markets, Unconventional Monetary Policy, And The Zero Lower Bound." *Macroeconomic Dynamics*, 24(4): 774–806.

Hülsewig, Oliver, and Horst Rottmann. 2021. "Euro area house prices and unconventional monetary policy surprises." *Economics Letters*, 205(C).

Iacoviello, Matteo. 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review*, 95(3): 739–764.

Jarociński, Marek, and Peter Karadi. 2020. "Deconstructing Monetary Policy Surprises—The Role of Information Shocks." *American Economic Journal: Macroeconomics*, 12(2): 1–43.

Kang, Jung Koo, Maria Loumioti, and Regina Wittenberg-Moerman. 2020. "Lifting the banking veil: credit standards' harmonization through lending transparency." European Central Bank Working Paper Series 2367.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review*, 108(3): 697–743.

Klein, Philipp, Carina Mössinger, and Andreas Pfingsten. 2021. "Transparency as a remedy for agency problems in securitization? The case of ECB's loan-level reporting initiative." *Journal of Financial Intermediation*, 46(C).

Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives*, 15(4): 143–156.

Koenker, Roger, and Zhijie Xiao. 2006. "Quantile Autoregression." *Journal of the American Statistical Association*, 101: 980–990.

Lenza, Michele, and Jirka Slacalek. 2021. "How Does Monetary Policy Affect Income and Wealth Inequality? Evidence from Quantitative Easing in the Euro Area." C.E.P.R. Discussion Papers 16079.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *The Review of Economic Studies*, 61(3): 397–415.

Musso, Alberto, Stefano Neri, and Livio Stracca. 2011. "Housing, consumption and monetary policy: How different are the US and the euro area?" *Journal of Banking & Finance*, 35(11): 3019–3041.

Nocera, Andrea, and Moreno Roma. 2017. "House prices and monetary policy in the euro area: evidence from structural VARs." European Central Bank Working Paper Series 2073.

OECD. 2020. "Housing Amid Covid-19: Policy Responses and Challenges." OECD Institutional Reports.

Ozkan, Serdar, Kurt Mitman, Fatih Karahan, and Aaron Hedlund. 2017. "Monetary Policy, Heterogeneity, and the Housing Channel." Society for Economic Dynamics 2017 Meeting Papers 1610.

Paz-Pardo, Gonzalo. 2021. "Homeownership and portfolio choice over the generations." European Central Bank Working Paper Series 2522.

Peersman, Gert, and Frank Smets. 2005. "The Industry Effects of Monetary Policy in the Euro Area." *Economic Journal*, 115(503): 319–342.

Pesaran, M. Hashem, and Ron Smith. 1995. "Estimating long-run relationships from dynamic heterogeneous panels." *Journal of Econometrics*, 68(1): 79–113.

Piazzesi, M., and M. Schneider. 2016. "Chapter 19 - Housing and Macroeconomics." In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, 1547–1640. Elsevier.

Pica, Stefano. 2022. "Housing Markets and the Heterogeneous Effects of Monetary Policy Across the Euro Area." mimeo.

Ramey, V.A. 2016. "Chapter 2 - Macroeconomic Shocks and Their Propagation." In . Vol. 2 of *Handbook of Macroeconomics*, , ed. John B. Taylor and Harald Uhlig, 71–162. Elsevier.

Rosenberg, Signe. 2020. "Conventional and unconventional monetary policies: effects on the Finnish housing market." *Baltic Journal of Economics*, 20(2): 170–186.

Rostagno, Massimo, Carlo Altavilla, Giacomo Carboni, Wolfgang Lemke, Roberto Motto, and Arthur Saint Guilhem. 2021. "Combining negative rates, forward guidance and asset purchases: identification and impacts of the ECB's unconventional policies." European Central Bank Working Paper Series 2564.

Schnabel, Isabel. 2021. "Monetary policy and inequality." European Central Bank Speech, 9 November 2021.

Slacalek, Jiri, Oreste Tristani, and Giovanni L. Violante. 2020. "Household balance sheet channels of monetary policy: A back of the envelope calculation for the euro area." *Journal of Economic Dynamics and Control*, 115(C).

Stock, J.H., and M.W. Watson. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In *Handbook of Macroeconomics*. Vol. 2 of *Handbook of Macroeconomics*, , ed. J. B. Taylor and Harald Uhlig, Chapter 0, 415–525. Elsevier.

The Economist. 2013. "The rich and QE." 3 June 2013.

The Economist. 2016. "Place-based economic policies as a response to populism." 15 December 2016.

The Financial Times. 2015*a*. "Central banks have made the rich richer." 22 September 2015.

The Financial Times. 2015*b*. "Modest ECB moves dash investor hopes and wrongfoot markets." 3 December 2015.

The Financial Times. 2021. "A different kind of housing bubble." 8 November 2021.

van Bekkum, Sjoerd, Marc Gabarro, and Rustom Irani. 2018. "Does a Larger Menu Increase Appetite? Collateral Eligibility and Credit Supply." *Review of Financial Studies*, 31: 943–979.

Vansteenkiste, Isabel, and Paul Hiebert. 2011. "Do house price developments spillover across euro area countries? Evidence from a global VAR." *Journal of Housing Economics*, 20(4): 299–314.

Zhu, Bing, Michael Betzinger, and Steffen Sebastian. 2017. "Housing market stability, mortgage market structure, and monetary policy: Evidence from the euro area." *Journal of Housing Economics*, 37(C): 1–21.

APPENDIX: PERSONAL CONTRIBUTIONS TO THE PAPERS OF THE CUMULATIVE DISSERTATION

The evolution of Zipf's Law for U.S. cities

This paper is joint work with Torben Klarl. Data collection and the empirical analysis were conducted by me. The writing of the manuscript was jointly conducted. The paper benefited from advice and comments from colleagues as well as participants of the conferences and workshops, where it had been presented.

House price convergence across German regions

This paper was written by me as a sole author. The paper benefited from advice and comments from colleagues, especially from Philipp Marek and Georgi Kocharkov.

Navigating the housing channel across euro area regions

This paper is joint work with Niccolò Battistini, Matteo Falagiarda and Moreno Roma. It was designed and conducted jointly from the beginning. However, data collection and descriptive statistics as well as the empirical analysis were were mostly conducted by me. The paper benefited from advice and comments from colleagues of the European Central Bank as well as participants of the conferences and workshops, where it had been presented.

ACKNOWLEDGEMENTS

Undertaking this research and writing this dissertation has been an enjoyable and challenging journey. It would not have been possible without the advice, collaboration, support and friendship of many people who I was lucky enough to have beside me, over the past years.

First of all, I want to thank my supervisor, Torben Klarl, for his advice and constant support during these years. Torben enabled me to grow professionally and personally and never failed to provide with motivation and guidance. From the first moment, Torben put trust into me and my capabilities and encouraged me to voice my ideas and develop them. Furthermore, I am very thankful for having been given the chance for various research stays outside the University of Bremen, which meant interruptions and a higher work load for all members of the Working Group Economics, esp. Macroeconomics.

Moreover, I would like to thank Christian Cordes for organizing the ierp seminar at the University of Bremen, which gave me a chance to present my research at a very early stage and also to broaden my research horizon by attending other presentations. My dissertation also benefited from discussions with participants at seminars at the European Central Bank, the University of Bremen and the University of Neuchâtel, at the 8th Ifo Workshop on Regional Economics, at the 6th Household Finance Workshop (Leibniz Institute SAFE) and at the 2022 ECHOPPE Conference on the Economics of Housing and Housing Policies.

I would like to thank David Audretsch for inviting me to a three-month research stay at Indiana University, Bloomington, which provided a great professional and personal experience. This research stay would not have been possible without the financial support of the BremenIDEA scholarship by the University of Bremen, which I am very thankful for. Besides the research stay in the United States, I was also able to do a six-month internship at the research center of the Bundesbank and I am very thankful to Phillip Marek for having supported me in the application process. Furthermore, I would like to thank Phillip Marek and Georgi Kocharkov for their professional advice and the fruitful collaboration. I also appreciate the great team work with Niccolò Battistini, Matteo Falagiarda and Moreno Roma during a one-year traineeship at the European Central Bank. Also after the end of the traineeship, our collaboration continued, which has always been inspiring and supported me in enhancing and broadening my own skills.

During my time at the University of Bremen, I appreciated the great atmosphere of support and team work in the office. I am thankful to my colleagues for always offering a sympathetic ear and sharing ideas, thoughts and worries. This process would not have been this enjoyable were it not for all the other doctoral students, who shared lunch, tea or coffee breaks and great discussions with me.

My last and greatest recognition goes to my family and friends. I would like to thank my family for providing all the encouragement and support over the years I could ask for. I would also like to thank my friends for their company and support during times of intensity and confusion but also during times of joy. Most importantly, I want to thank Alvin, my partner, not only for his linguistic support, but more importantly for going through all ups and downs of writing this dissertation with me and always being understanding.