

Discussion Paper

Deutsche Bundesbank
No 46/2013

**Assessing house prices in Germany:
evidence from an estimated stock-flow
model using regional data**

Florian Kajuth

Thomas A. Knetsch

Nicolas Pinkwart

Editorial Board:

Klaus Düllmann
Heinz Herrmann
Mathias Hoffmann
Christoph Memmel

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet <http://www.bundesbank.de>

Reproduction permitted only if source is stated.

ISBN 978-3-86558-980-4 (Printversion)

ISBN 978-3-86558-981-1 (Internetversion)

Non-technical summary

After several years of subdued growth residential property prices in Germany started to increase markedly in 2010. Urban agglomerations in Germany saw rather strong price increases, while the whole country average indicates relatively moderate house price rises. Against this background it is of interest to investigate whether the recent house price increases are signals of an incipient overheating of the German housing market, in particular in large urban areas, or whether they are attributable to changes in macroeconomic factors.

However, on the basis of aggregate time-series data it is difficult to obtain economically plausible relationships between German house prices and standard explanatory variables such as income or unemployment. The statistical analysis is plagued by insufficient variation in some of the variables and house prices either due to the low frequency of observations or a relatively short time span covered.

The contribution of the paper is, first, to use regional data to estimate an equilibrium house price equation. The dataset exploits the cross-sectional variation in housing market variables, which reflects a larger amount of information than the time-series dimension. We use an instrumental variables estimator based on a random-effects setup to make the most of the cross-section variation while accounting for potential endogeneity of some regressors. The results from our regional panel model point to significant effects of the housing stock per capita, income per capita, the unemployment rate, the size of the middle age cohorts, the population density and growth expectations on house prices.

Second, the proposed approach can serve to assess house prices relative to their explanatory factors for the whole country or other sub-aggregates such as urban areas or large cities. This is particularly relevant at the current juncture given that house price developments have recently differed substantially across regions.

Comparing actual house prices to their fundamental price for a number of different regional sub-aggregates suggests that apartment prices significantly exceed their fundamental price over the past two years. Their deviation is largest in the group of major cities and smaller for the whole-country average. A similar pattern holds for single-family houses with the size of house price misalignments being generally smaller than for apartments and often not significant. For Germany as a whole apartment prices show a moderate over-valuation, whereas single-family house prices appear to be in line with the level suggested by the model.

Nicht-technische Zusammenfassung

Die Preise für Wohnimmobilien in Deutschland sind seit 2010 deutlich gestiegen. In den Ballungsräumen waren die Preissteigerungen am größten, während sie für Deutschland insgesamt bislang moderater ausfielen. Vor diesem Hintergrund stellt sich die Frage, ob die jüngsten Anstiege der Immobilienpreise Anzeichen einer bevorstehenden Überhitzung am deutschen Immobilienmarkt sind, insbesondere in den Ballungsräumen, oder ob sie die Folge von Änderungen in den zugrundeliegenden makroökonomischen Einflussgrößen sind.

Auf Basis aggregierter Zeitreihen lassen sich für Deutschland nur schwer ökonomisch plausible Beziehungen zwischen den Immobilienpreisen und Größen wie Einkommen oder Arbeitslosenquote herstellen. Die begrenzte Variation einiger Erklärungsfaktoren und der Immobilienpreise selbst aufgrund der geringen Häufigkeit der Beobachtungen sowie des relativ kurzen verfügbaren Stichprobenzeitraums erschweren die statistische Analyse.

Dieser Beitrag nutzt zum einen Angaben zu Immobilienpreisen und makroökonomischen Einflussgrößen auf Ebene der Kreise und kreisfreien Städte, um fundamental begründete Preise für Wohnimmobilien zu bestimmen. Der zugrundeliegende Datensatz stellt die Querschnittsdimension in den Vordergrund, die aufgrund stärkerer Variation einen höheren Informationsgehalt als die Zeitreihenbetrachtung aufweist. Möglichen Endogenitätsproblemen bei der Schätzung im Rahmen eines Random-Effects-Modells wird mit Hilfe eines Instrumentenschätzverfahrens begegnet. So soll sichergestellt werden, dass die Information auf Basis des Datenquerschnitts möglichst vollständig verwendet wird, während Korrelationen zwischen den Regressoren und unbeobachteten Effekten Rechnung getragen wird. Die Ergebnisse der Analyse des regionalen Datensatzes deuten auf signifikante Zusammenhänge zwischen Immobilienpreisen und dem Wohnungsbestand pro Einwohner, dem Pro-Kopf-Einkommen, der Arbeitslosenquote, dem Anteil der Bevölkerung in den mittleren Altersklassen, der Bevölkerungsdichte sowie den langfristigen Wachstumserwartungen hin.

Darüber hinaus können die geschätzten Fundamentalpreise mit den tatsächlichen Immobilienpreisen für Gesamtdeutschland, aber auch für Teilmärkte wie die Ballungsräume oder die größeren Städte verglichen werden. Da sich die Entwicklung der Immobilienpreise gegenwärtig regional stark unterscheidet, ist eine Verdichtung der geschätzten Fundamentalpreise nach verschiedenen regionalen Abgrenzungen wünschenswert. Ein Vergleich der tatsächlichen Immobilienpreise mit ihren Fundamentalwerten für verschiedene Regionen und Gesamtdeutschland zeigt, dass die Preise für Eigentumswohnungen in den größeren deutschen Städten in den vergangenen zwei Jahren deutlich von ihren Fundamentalwerten abgewichen sind. Ein ähnliches Bild ergibt sich für die Preise von Einfamilienhäusern, deren Abweichung vom Fundamentalwert jedoch in den betrachteten regionalen Abgrenzungen schwächer ausfällt und in vielen Fällen statistisch nicht signifikant ist. Für Deutschland insgesamt lässt sich eine moderate Überbewertung bei Eigentumswohnungen feststellen, wohingegen die Preise für Einfamilienhäuser im Einklang mit ihrem fundamental gerechtfertigten Preis stehen dürften.

Assessing house prices in Germany: evidence from an estimated stock-flow model using regional data

Florian Kajuth*
Deutsche Bundesbank

Thomas A. Knetsch
Deutsche Bundesbank

Nicolas Pinkwart
Deutsche Bundesbank

October 1, 2013

Abstract

Based on a stock-flow model of the housing market we estimate the relationship of house prices and explanatory macroeconomic variables in Germany using a regional panel dataset for 402 administrative districts. Using regional data exploits the variation across local housing markets and overcomes time-series data limitations. We take the regression residuals as a measure for deviations of actual house prices from their fundamental equilibrium level. The model specification allows to aggregate district-level residuals for various regional subsets. During the past two years for Germany as a whole single-family house prices appeared to be in line with their fundamental equilibrium level, whereas apartment prices significantly exceeded the fundamental price suggested by the model. The overvaluation of apartments is higher in towns and cities and most pronounced in the major seven cities, while single-family houses in cities appear to be only moderately above their fundamental levels.

Keywords: fundamental house prices, regional house prices, housing demand and supply

JEL classification: E32, R21, R31

*Corresponding author. Address: Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main. Email: florian.kajuth@bundesbank.de. We would like to thank Jörg Breitung, Klemens Hauzenberger, Vladimir Kuzin and participants at the Bundesbank research seminar for valuable comments and suggestions. All remaining errors remain our own. This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.

1 Introduction

From a macroeconomic perspective house prices matter for a number of reasons. Via their allocative signaling role for residential investment house prices have an effect on business cycle dynamics and long-term growth. By influencing households' housing wealth property prices might affect private consumption spending, which in turn might lead to changes in output and inflation. Also, since residential property is usually debt financed to a large extent, a substantial decline in house prices may damage macroeconomic and financial stability. These effects were forcefully demonstrated in many countries including the US and some EU member states, where the recent recession was preceded or accompanied by a boom/bust pattern of residential property prices.¹

By contrast, house prices in Germany did not experience a noticeable boom before the economic and financial crisis nor did they fall dramatically in the wake of it. With the onset of the recovery in 2010 residential property prices started to increase. According to selected house price indicators in Fig. 1 urban agglomerations in Germany saw rather strong price increases over recent years, while the whole country average (BulwienGesa 402 districts) indicates relatively moderate house price rises. An index of 125 German cities and the price index for owner-occupied housing by the Association of German Pfandbrief banks range inbetween. Against this background it is of interest to investigate whether the recent house price increases are signals of an incipient overheating of the German housing market, in particular in large urban areas, or whether they are attributable to changes in macroeconomic factors.

To address this question, the paper exploits the cross-sectional variation in German regional housing markets in addition to the time-series dimension. In particular, a panel dataset for 402 administrative districts from 2004 to 2010 is used to estimate an equilibrium house price equation. The estimated economic relationships between house prices and a number of explanatory variables are more closely aligned with the standard macroeconomic correlations than estimates based on time variation. In addition, our approach aims at assessing house prices relative to their explanatory factors for the whole country or other sub-aggregates such as urban areas or large cities. This is particularly relevant at the current juncture given that house price developments have recently differed substantially across regions.

On the basis of aggregate time-series data empirical models of German house prices often yield imprecisely estimated coefficients of explanatory variables, such as income or unemployment, (e.g., Catte et al. 2004; Igan and Loungani, 2012) or effects that are hard

¹A large literature addresses the interactions between these housing markets and the macroeconomy. E.g., Altissimo et al. (2005), Borio and Lowe (2002), Case, Quigley and Shiller (2005), Glaeser, Gyourko and Saiz (2008), Iacoviello (2004), Iacoviello and Minetti (2008).

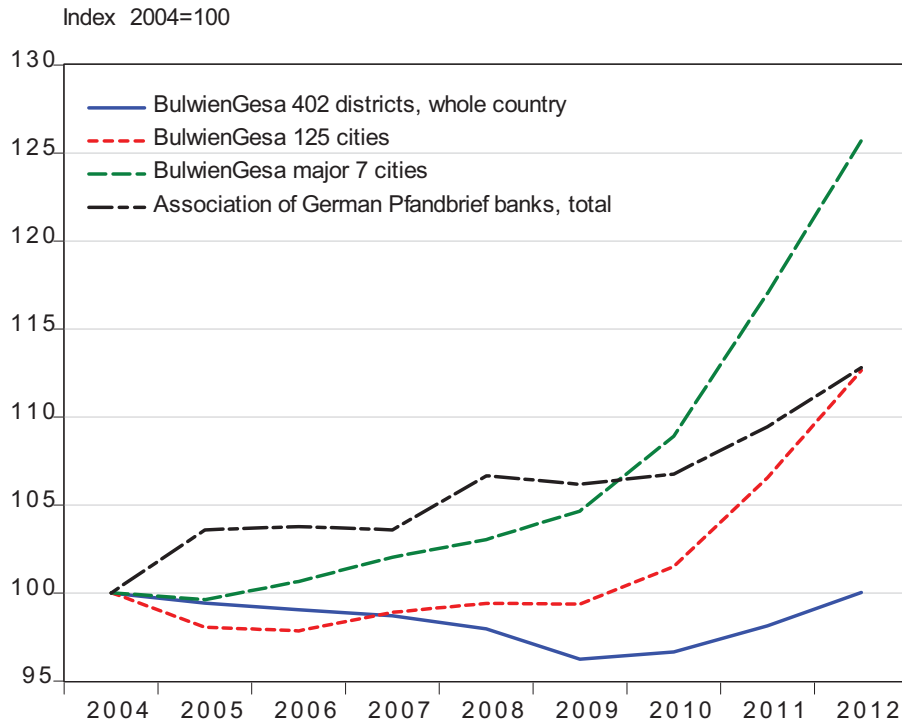


Figure 1: Selected nominal house price indicators. Source: BulwienGesa AG and Association of German Pfandbrief Banks.

to reconcile with standard economic relationships (Dreger and Kholodilin, 2013; Girouard et al., 2004; Hiebert and Sydow, 2011; Hiebert and Vansteenkiste, 2011). This is mainly due to the lack of statistical power using methods based on time series data for the German housing market. Quarterly time series of a sufficiently high quality go back less than ten years. Data that go back over a longer period (back to German re-unification) are available on an annual basis only. In both cases statistical inference is plagued by insufficient variation in some of the explanatory variables and house prices either due to the low frequency of observations or a relatively short time span covered.² Drawing on the information from the cross-sectional variation as in our approach helps to overcome these data limitations.

There is little empirical evidence for German house prices on the basis of regional data, which is in sharp contrast to the availability of studies covering the US and UK real estate markets (e.g., Ashworth and Parker, 1997; Cameron, Muellbauer and Murphy, 2006; Capozza et al., 2002; Hwang and Quigely, 2006). In addition, none of the existing

²These insufficiencies can even lead to the exclusion of German data from multi-country time-series analyses (Adams and Füss, 2010; European Commission, 2012).

studies that look at the German housing market from a regional perspective investigate the implications for aggregate over- or undervaluations at different regional levels. Koetter and Poghosyan (2010) estimate fundamental house prices for the German regions. Their analysis is based on the time-series variation in a panel approach, whereas we focus on the cross-section variation in the data. Moreover, they restrict the set of regressors to GDP per worker and population growth, while our approach incorporates a larger set of economic and demographic determinants. Maennig and Dust (2008) focus on the effects of population growth on house prices. In a cross-section regression they consider the effects of income, population growth and construction costs on the price of single-family houses in German metropolitan areas. However, their dataset is limited to the year 2002, whereas our approach allows to use the estimated effects on the basis of cross-sectional information for the evaluation of house prices over time. Bischoff (2012) also provides estimates of regional single-family house prices in Germany for one year, 2005, and finds income and population growth to be important drivers of house prices.³

To make use of the cross-section information in a panel model while accounting for potential endogeneity of some regressors we use an instrumental variables estimator based on a random-effects setup, which allows to address potential correlation of some of the explanatory variables with the unobserved district-specific effect. Following Hausman and Taylor (1981) we use the district means of variables which are uncorrelated with the unobserved effect as instruments. Assuming that the estimated equation on the district-level represents the true underlying relationship with parameters valid for all regional entities, a model for the fundamental prices of houses at different levels of aggregation can be constructed if the explanatory variables at district level are measured in per-capita terms. In order to gauge the degree of over- or undervaluation in German house prices at different aggregation levels, the deviation of actual house prices from their estimated fundamental price is computed for different regional sub-aggregates.

The results from our regional panel model, which differentiates between single-family houses and apartments, point to significant effects of the housing stock per capita, income per capita, unemployment, a demographic measure, the population density and growth expectations on house prices. Comparing actual house prices to their fundamental prices for a number of different regional sub-aggregates over the past two years suggests that apartment prices significantly exceed their fundamental prices. Their deviation is largest in the group of major cities and smaller for a whole-country average. A similar pattern holds for single-family houses with the size of house price misalignments being generally smaller than for apartments and often not significant. For Germany as a whole apartment

³Kholodilin and Mense (2012) carry out time-series analyses of prices and rents for houses and apartments in the major German cities, without however looking at the cross-section information.

prices show a moderate overvaluation, whereas single-family house prices appear to be in line with the level suggested by the model.

The remainder of the paper is structured as follows. Section 2 outlines the theoretical stock-flow model of the housing market. Section 3 clarifies some aggregation issues, and section 4 presents the data and discusses important specification choices of the panel model. Section 5 contains the estimation results and section 6 offers an assessment of the recent house price increases. Section 7 concludes.

2 The stock-flow model of the housing market

The empirical specification is derived from a theoretical stock-flow model, which accounts for the interaction of house prices with residential investment and allows for sluggish stock and price adjustments (DiPasquale and Wheaton, 1992; McCarthy and Peach, 2002; Steiner, 2010). It is compatible with an error-correction mechanism to describe demand and supply dynamics on the housing market, which is often used in the empirical literature to quantify the misalignment of house prices with their determinants and the subsequent adjustment (e.g., Capozza et al., 2004; Malpezzi, 1999). Also, the stock-flow model nests the asset-pricing approach (Hiebert and Sydow, 2011) and the user-cost approach (Himmelberg et al., 2005; Poterba, 1984) to evaluating house price movements. Beyond that, it provides the flexibility to include additional potential house price determinants.

The stock-flow model describes the housing market using a law of motion for the housing stock, a demand equation for housing and an adjustment equation for house prices and residential investment, whenever housing demand deviates from available supply. The housing stock s_t evolves according to

$$s_t = (1 - \delta) s_{t-1} + b_{t-1} \quad (1)$$

whereby the housing stock at the beginning of period t equals the stock at $t - 1$ adjusted for depreciation at the rate δ plus residential investment b_{t-1} during period $t - 1$.

The demand for housing x_t^d is assumed to depend positively on income y_t , negatively on the housing rent m_t , and on other factors z_t , in particular demographic variables or labour market factors (Cameron, Muellbauer and Murphy, 2006).⁴

$$\ln x_t^d = \alpha_0 + \alpha_1 \ln y_t - \alpha_2 \ln m_t + \alpha_3 \ln z_t \quad (2)$$

Measures of current income could be viewed as capturing business cycle effects or borrowing constraints (McQuinn and O'Reilly, 2008). Current income might also affect affordability and the loan-to-value ratio because, next to the collateral value of the house,

⁴The signs in the coefficient vector α_3 are variable-specific.

banks often use affordability measures to judge the appropriate loan-to-value ratio for a mortgage. Affordability in turn can be computed by comparing the cashflow for loan repayment to a fixed proportion of current income.

The one-period real rent m_t is related to real house prices via the asset pricing condition for houses.

$$r_t^h = \frac{E_t p_{t+1} + m_t - p_t}{p_t} \quad (3)$$

The ex-ante real yield from renting out one unit of housing r_t^h is equal to the percentage difference between the expected house price next period $E_t p_{t+1}$ plus the rent m_t received during period t and the current house price p_t . Re-arranging, iterating forward and imposing the transversality condition yields the familiar equation for the house price in terms of the sum of discounted expected future rent payments.

$$p_t = E_t \sum_{k=0}^{\infty} \frac{m_{t+k}}{\prod_{n=0}^k (1 + r_{t+n}^h)} \quad (4)$$

In order to operationalize the expression two further assumptions are made. First, rents are expected to be a constant fraction of income and to grow at an average long-term rate g_t^e , i.e. $m_{t+k} = (1 + g_t^e)^k m_t$, $k = 1, 2, \dots, \infty$. Second, future rent payments are discounted at an average long-term real interest rate r_t , with $(1 + r_t)^k = \prod_{n=0}^k (1 + r_{t+n}^h)$, $k = 1, 2, \dots, \infty$. These assumptions are in line with the view that it is mainly the persistent component of the growth rate of rents or the interest rate which matters for housing demand. Re-arranging for the price-rent ratio p_t/m_t and assuming $r_t - g_t^e > 0$ results in

$$\frac{p_t}{m_t} = \frac{1}{r_t - g_t^e} \quad (5)$$

Taking logs, substituting out m_t in (2) and imposing the equilibrium condition, $\ln x_t^d \equiv \ln s_t$, yields the inverted demand curve for housing, where the supply of housing is inelastic in the short run (Muellbauer and Murphy, 1994).

$$\ln p_t^* = \frac{\alpha_0}{\alpha_2} - \frac{1}{\alpha_2} \ln s_t + \frac{\alpha_1}{\alpha_2} \ln y_t + \frac{\alpha_3}{\alpha_2} \ln z_t - \ln(r_t - g_t^e) \quad (6)$$

While p_t^* is fully described by observable economic factors, actual house prices may deviate from p_t^* , e.g. due to shocks. Assuming that actual house prices converge back to their level determined by economic factors only, p_t^* can be regarded as the fundamental equilibrium price (DiPasquale and Wheaton, 1992).

3 Aggregation

In order to apply the empirical model at district level for house price analysis at higher levels of aggregation the specification of the panel model and the definition of the explana-

tory variables have to fulfil some conditions. An economically consistent aggregation of the regional house price equation is obtained when the regional variables can be replaced by their aggregate counterparts without violating the equality between the left-hand and the right-hand side of the equation. Given that regional house price indices under consideration are usually *arithmetically* aggregated using population weights, this is achieved if (i) the postulated relationship is linear in all explanatory variables and (ii) all explanatory variables measured in volumes (e.g. income or the housing stock) are measured in per-capita terms.⁵

Let us first write the equation for the house price in district i as

$$p_{it} = \alpha + \beta v_{it}^1 + \gamma v_{it}^2 + \theta w_t \quad (7)$$

where v_{it}^1 and v_{it}^2 are district-specific explanatory variables and w_t comprises the variables that do not vary across the regional units. An aggregate population-weighted house price index can be constructed as

$$p_t^{lin} = \sum_{i=1}^I \frac{n_{it}}{n_t} p_{it} \quad (8)$$

where n_{it} is the number of residents in district i and $n_t = \sum_{i=1}^I n_{it}$. Substituting (7) yields

$$p_t^{lin} = \alpha + \beta \sum_{i=1}^I \frac{n_{it}}{n_t} v_{it}^1 + \gamma \sum_{i=1}^I \frac{n_{it}}{n_t} v_{it}^2 + \theta w_t \quad (9)$$

Under condition (ii) v_{it}^1 and v_{it}^2 are defined in per-capita terms, $v_{it}^1 = \frac{\tilde{v}_{it}^1}{n_t}$ and $v_{it}^2 = \frac{\tilde{v}_{it}^2}{n_t}$. This yields

$$\begin{aligned} p_t^{lin} &= \alpha + \beta \frac{1}{n_t} \sum_{i=1}^I \tilde{v}_{it}^1 + \gamma \frac{1}{n_t} \sum_{i=1}^I \tilde{v}_{it}^2 + \theta w_t \\ &= \alpha + \beta \frac{\tilde{v}_t^1}{n_t} + \gamma \frac{\tilde{v}_t^2}{n_t} + \theta w_t \end{aligned} \quad (10)$$

given the aggregation of volumes $\tilde{v}_t^1 = \sum_{i=1}^I \tilde{v}_{it}^1$ and $\tilde{v}_t^2 = \sum_{i=1}^I \tilde{v}_{it}^2$. As $\frac{\tilde{v}_t^1}{n_t}$ and $\frac{\tilde{v}_t^2}{n_t}$ are the counterparts of v_{it}^1 and v_{it}^2 at the aggregate level the district-level specification (7) is also valid at the aggregate level. In reconciling the theoretical house price equation (6) with the requirements for aggregation, the explanatory variables measured as volumes (e.g. income or the housing stock) have to be expressed in per-capita units, while other variables like the interest rate and growth expectations can be included without adjustment.

The linearity condition (i), however, is at odds with the log-linearised form of (6), which results from a first-order approximation to the non-linear theoretical model. As we

⁵Population weights are supposed to proxy for the expenditure shares of housing, which are not available at district level.

actually do not know whether this approximation is close to the unknown true structure, we cannot reject the linear functional form by theoretical arguments. In addition, we formally investigate the potential aggregation errors stemming from two sources. The first error arises from estimating and consistently aggregating a linear model, when in fact the logarithmic formulation is true at the district level. This error relates to the aggregation of misspecification error. The second error is due to aggregation of the logarithmic model under the null hypothesis that it is the true specification and requiring that the regional equations be economically consistent with their aggregate counterpart. An economically sensible (but inconsistent) aggregation of the logarithmic district-level model

$$\ln p_{it} = \ln \alpha + \beta \ln v_{it}^1 + \gamma \ln v_{it}^2 + \theta \ln w_t \quad (11)$$

might be

$$\ln \sum_{i=1}^I \frac{n_{it}}{n_t} p_{it} \equiv \ln \alpha + \beta \ln \sum_{i=1}^I \frac{n_{it}}{n_t} v_{it}^1 + \gamma \ln \sum_{i=1}^I \frac{n_{it}}{n_t} v_{it}^2 + \theta \ln w_t \quad (12)$$

Of course, this definition differs from the formally consistent aggregation equation given by

$$\ln p_t^{\log} = \ln \sum_{i=1}^I \frac{n_{it}}{n_t} p_{it} = \ln \sum_{i=1}^I \frac{n_{it}}{n_t} \alpha (v_{it}^1)^\beta (v_{it}^2)^\gamma w_t^\theta \quad (13)$$

but allows to use the district-level specification at the aggregate level.

To inform a decision between the two aggregation procedures (9) and (12) a simulation study is carried out comparing the mean errors (MEs) of estimating and aggregating the linear versus the logarithmic model under the null that, at the district level, the logarithmic specification is correct. Appendix C contains the details of the simulation study. The results suggest that the error of an economically consistent aggregation of the potentially wrong linear model is much smaller than the bias from non-exact aggregation of the true logarithmic equation under the null. The numerical evidence supports the theoretical argument above that a linear specification is at least as good an approximation as the logarithmic model. In the following we therefore use a linear specification of the house price equation.

4 Empirical model

4.1 Panel model specification

We specify a linear regional panel model for I administrative districts denoted by $i = 1, 2, \dots, I$ and $t = 1, 2, \dots, T$ periods. It is assumed that the model (1) to (6) is a valid description of the housing market in the individual districts.

$$p_{it} = \beta_0 + \beta_1 s_{it} + \beta_2 y_{it} + \beta_3 z_{it} + \beta_4 r_t + \beta_5 g_t^e + c_i + \epsilon_{it} \quad (14)$$

where variables at district level are defined as above. We allow for district-specific effects c_i , which are unobserved. The interest rate and growth expectations are assumed to be the same for each district. District-specific growth expectations over and above those for the whole economy are unobserved and left to enter the equation via the district-specific unobserved effects c_i . This implies that district-specific growth expectations are assumed time-invariant over the estimation sample and are contained in c_i . Strict exogeneity with respect to the error terms ϵ_{it} is required to hold.

The panel dataset covers house prices (in euros) for existing apartments and existing single-family houses provided by a private firm, BulwienGesa, in all 402 administrative districts of Germany (*Kreise und kreisfreie Städte*) at annual frequency for the years 2004 to 2012, and for a subset of 93 towns and cities for the years 1996 to 2012. They are deflated by regional consumer price indices constructed on the basis of price level information by Kawka (2009), which we backcast using annual CPI data for the state (*Länder*) to which a particular district belongs.

Real house prices per square metre are modelled as in (6) using the number of housing units per resident (s_{it}), real GDP per capita (y_{it}), the real long-term interest rate (r_t) and long-term real growth expectations (g_t^e) as explanatory variables. The vector z_{it} includes the population aged between 30 and 55 as a ratio of total population (a_{it}), a measure of the population density (d_{it}) (area per resident) and the unemployment rate (u_{it}) (in % of population). Note that these three variables are defined such that they fulfil the aggregation condition (ii). The age-cohort variable is supposed to capture life-cycle motives for purchasing a house, which are not covered by the remaining variables. The population density variable serves to account for a scale effect, namely that districts with a higher population density are usually characterised by higher house prices *ceteris paribus*. The unemployment rate could affect house prices through its role as an indicator for district-specific income perspectives. The real 10-year mortgage rate proxies for the required return in the asset-pricing equation (4) and is obtained by subtracting expected inflation over the ten years ahead from 10-year nominal mortgage rates (see appendix A for the list of data sources and definitions).

For the whole set of districts the estimation period is 2004 to 2010 because the full list of explanatory variables is only available up to 2010. In case of the 93 towns and cities, estimation is carried out over the sample starting in 1996. We estimate separate equations for apartments and single-family houses.

4.2 Cross-section vs. time-series dimension

A crucial question is whether to treat the unobserved district-specific effects c_i as random or fixed effects. From a theoretical perspective the answer would depend on whether

Table 1: Summary statistics

Variable	Definition	Mean	Std. Dev.			Min.	Max.
			between	within	overall		
p_{it}	Prices in 2009 euros per square meter						
	Single-family houses	1839	507.0	115.5	519.5	761	3917
	Apartments	1460	329.9	111.2	347.8	601	3578
s_{it}	Housing units per capita						
	Single-family houses	0.202	0.067	0.005	0.067	0.044	0.364
	Apartments	0.278	0.104	0.004	0.104	0.101	0.584
y_{it}	GDP per capita in 2009 euros	30730	10854	1545	10952	13742	98843
a_{it}	Population aged 30 to 55 as share of total	0.37	0.014	0.005	0.015	0.32	0.41
d_{it}	Population density (square kilometre per capita)	0.006	0.005	<0.001	0.005	<0.001	0.026
u_{it}	Unemployment (as share of population)	0.046	0.021	0.009	0.022	0.009	0.131
r_t	10-year real mortgage rates	0.030	-	0.004	0.004	0.022	0.035
g_t^e	Real growth expectations	0.015	-	0.001	0.001	0.012	0.017

the relationship between house prices and their determinants are governed by the same underlying distribution regardless of an individual district, or whether this relationship is to be viewed as conditional on being in a specific district. From the point of view of econometric inference, it must be taken into account whether the unobserved effect is correlated with any of the regressors in the house price equation. If so, an omitted variable that is correlated with the regressors gives rise to an endogeneity problem and the random-effects estimator produces inconsistent coefficient estimates, whereas the fixed-effects estimator is consistent. Usually, a Hausman test is carried out to test for correlation between the unobserved effect and the regressors (Hausman, 1978). The Hausman test statistic ξ_H compares the coefficient estimates from the random-effects and the fixed-effects estimator relative to the difference in their covariance matrices.

$$\xi_H = \left(\hat{\beta}_{FE} - \hat{\beta}_{RE} \right)' \left[\hat{V}\{\hat{\beta}_{FE}\} - \hat{V}\{\hat{\beta}_{RE}\} \right]^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE} \right) \quad (15)$$

where \hat{V} are estimates of the covariance matrices. A significant difference between the two coefficient estimates usually indicates that a fixed-effects estimator is required for consistent estimates.

However, in our case, the results of a Hausman test in its standard form (15) are to be treated with caution because over the available sample the empirical correlations between house prices and regressors differ strongly depending on whether they are estimated along the time-series or cross-section dimension. The summary statistics in Table 1 show that most of the variance in the individual regressors comes from the variance *between* district means, while the *within* variances are relatively small. Additionally, Figs. 2 to 5 plot the unconditional correlations between the prices for apartments and single-family houses with each of the district-specific regressors.

The first column presents their overall correlation, while the second column relates the deviations of prices and regressors from their district means, and the third column shows the correlation of the district means of variables. Using the *within* variance (second column) is likely to yield counterintuitive coefficients attached to some key regressors, e.g. a negative effect of GDP per capita on house prices and a positive effect of the unemployment rate. We consider the correlations based on deviations from the district mean implausible because the *within* variance of these variables is small relative to the *between* variance. This is due to little time variation of the regressors over the sample period of $T = 7$ relative to the variation across 402 districts in the sample. The *between* variance is more likely to capture the true underlying structural relationships between variables than time variation in the sample.⁶

⁶This issue is related to the problem of finding sensible coefficients in an aggregate time-series estimation of house prices for Germany, which provided the motivation for including the cross-sectional variation in a panel estimation.

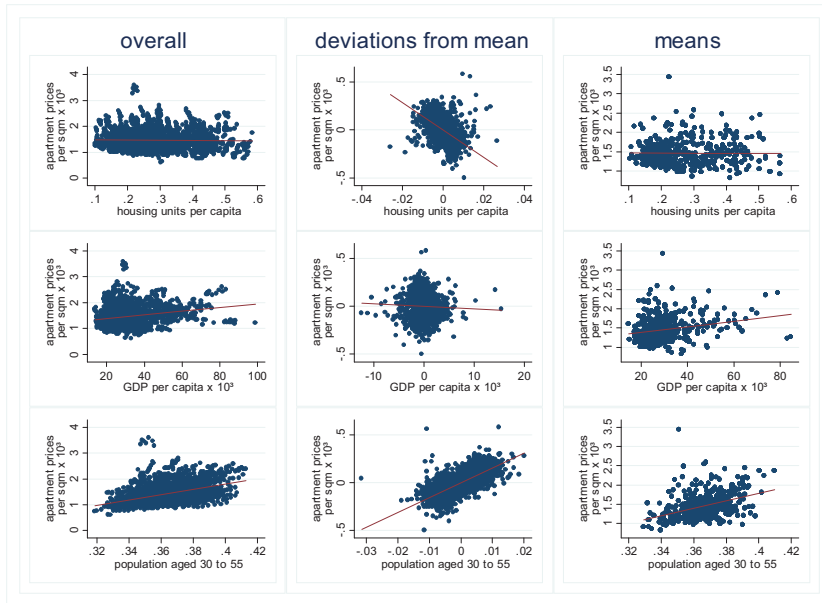


Figure 2: Unconditional correlations of regressors with apartment prices (overall, in deviations from district mean and in district means).

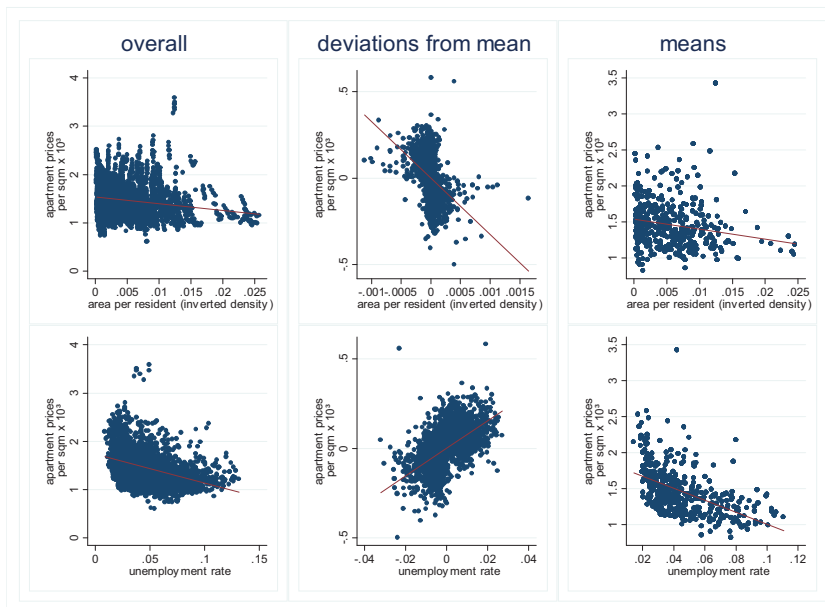


Figure 3: Unconditional correlations of regressors with apartment prices (overall, in deviations from district mean and in district means).

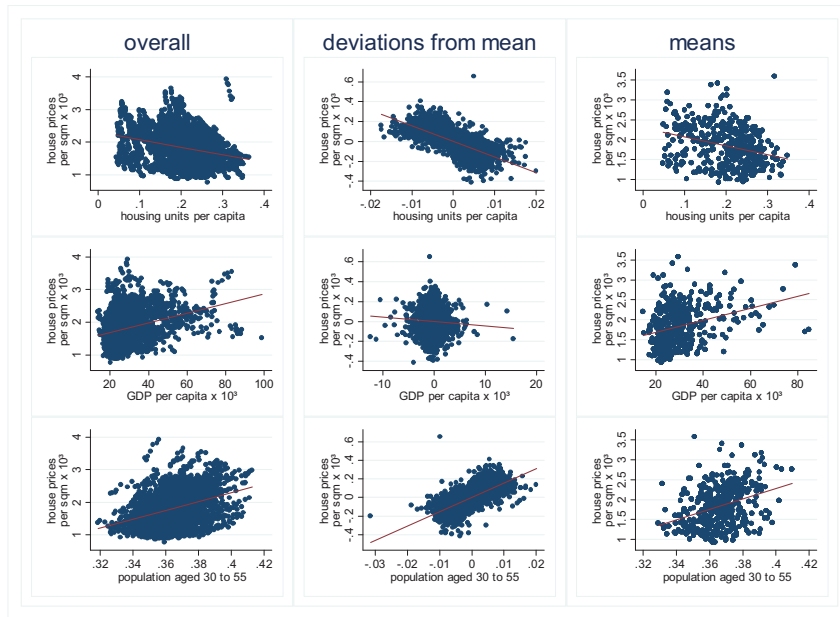


Figure 4: Unconditional correlations of regressors with single-family house prices (overall, in deviations from district mean and in district means).

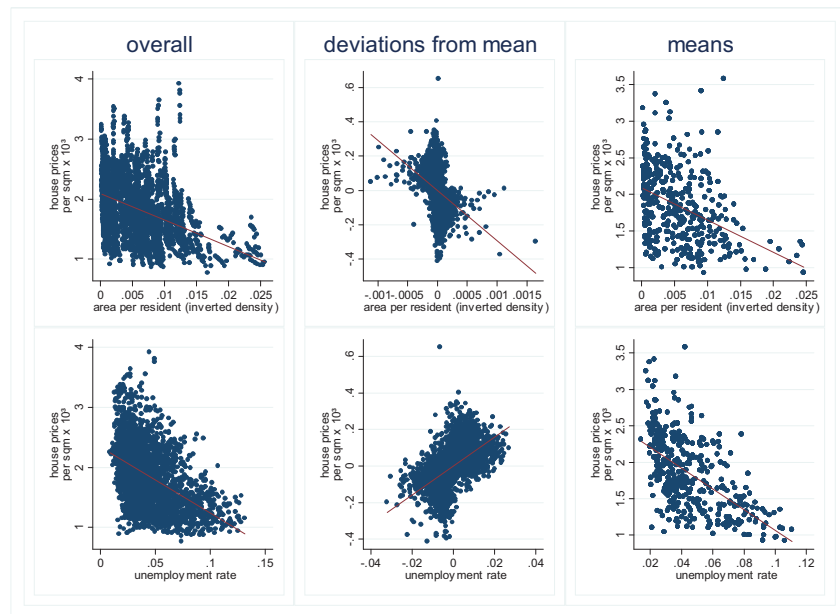


Figure 5: Unconditional correlations of regressors with single-family house prices (overall, in deviations from district mean and in district means).

The small *within* variance also complicates the interpretation of the Hausman test statistic. Suppose the Hausman test rejects the null of no correlation between the unobserved effect and the regressors, which would be reflected in large differences of the coefficient estimates. It would be unclear whether the differences are truly due to an endogeneity problem, or whether they are due to untypically signed fixed-effects coefficient estimates over the sample period.

In contrast to the district-specific explanatory variables, data on growth expectations and mortgage rates do not exhibit cross-sectional variation.⁷ However, both are important explanatory factors according to the theoretical model. In addition, growth expectations are likely to capture the downward trend in real house prices since the mid-1990s. Fig. 6 plots the long-run growth expectations (left panel) and 10-year mortgage rates (right panel) along with real house prices in 93 towns and cities, which are available over the longer sample starting in 1996. Long-run growth expectations are highly positively correlated with house and apartment prices over the sample, which is in line with the theoretical model.

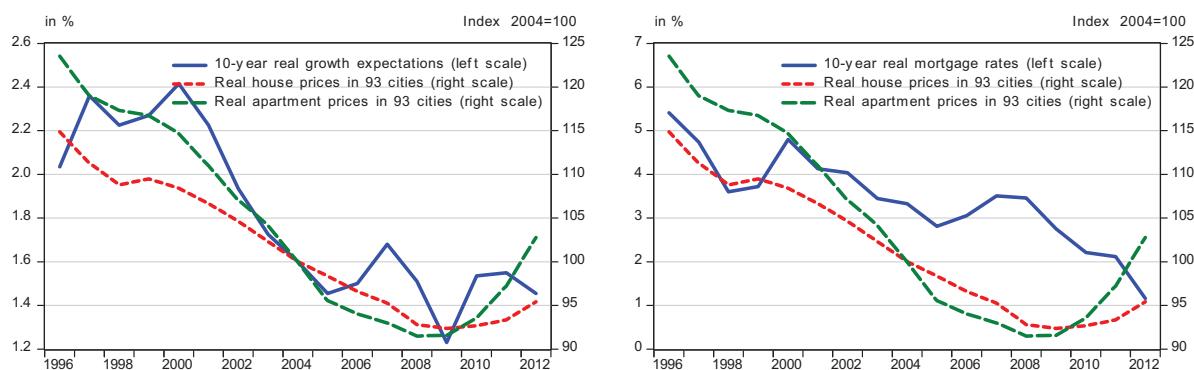


Figure 6: 10-year real growth expectations (left panel) and 10-year mortgage rates (right panel) along with real house and apartment prices for 93 cities.

At the same time, over most of the extended estimation period both house prices and mortgage rates have a downward trend yielding a positive empirical relationship between them. While this empirical finding does not contradict the theoretically negative relationship from the theoretical partial equilibrium model, it is less useful for assessing house prices over the recent past when real mortgage rates decreased to exceptionally low levels and house prices were increasing.

We conclude from this discussion that in order to uncover the structural correlations between house prices and district-specific variables the cross-section variation plays an

⁷Lack of data on district-specific growth expectations or mortgage rates is to some extent plausible because they are not necessarily meaningful concepts. Both are probably more strongly related to sector-specific, national or even international factors, rather than location-specific.

important role, which calls for including as many districts as possible in the estimation. However, this comes at the expense of reducing the time dimension making it more difficult to precisely estimate the impact of variables without variation across districts. Thus, there is a trade-off between optimally exploiting the available time-series dimension versus making best use of the existing cross-section variation. In the empirical part we will address this issue by providing estimates for the sample of 402 districts over a shorter sample and for a sample of 93 towns and cities over a longer horizon.

4.3 Hausman-Taylor instrumental variables estimator

The Hausman-Taylor (HT) estimator (Hausman and Taylor, 1981) provides the basis for an instrumental variables (IV) random-effects estimator that accounts for the potential endogeneity of some regressors, while at the same time making best use of the *between* variance in the data.⁸ Intuitively, it provides two useful elements for our purposes. First, while the random-effects estimator assumes that none of the regressors are correlated with the unobserved effect, and the fixed-effects estimator removes the district-specific means from the estimation equation for all regressors, the HT estimator uses IVs on endogenous regressors in a random-effects environment. Second, the HT estimator does not require instruments from outside the model as long as some regressors can be treated as exogenous.

To illustrate that, suppose that in (14) c_i is correlated with z_{it} but not with any of the remaining regressors. The HT/IV estimator starts from the random-effects formulation (14)

$$\tilde{p}_{it} = \beta_0 + \beta_1 \tilde{s}_{it} + \beta_2 \tilde{y}_{it} + \beta_3 \tilde{z}_{it} + \beta_4 \tilde{r}_t + \beta_5 \tilde{g}_t^e + c_i + v_{it} \quad (16)$$

where a tilde denotes the quasi-time-demeaned variables. Endogenous regressors (e.g. z_{it}) are instrumented by variables uncorrelated with the unobserved effect. These can be the time-demeaned regressors (e.g. $z_{it} - \bar{z}_i$) as in the fixed-effects transformation, or also the mean of any exogenous regressor, e.g. \bar{y}_i . Instrumenting by the means of regressors which are uncorrelated with the unobserved effect, instead of $z_{it} - \bar{z}_i$, uses the *between* variance of the exogenous regressors instead of the *within* variance of z_{it} . More generally, we can use the mean of any exogenous regressor as an instrument for the endogenous regressors, as long as there are at least as many exogenous regressors as endogenous ones. Note that the fixed-effects estimator does not distinguish between exogenous and endogenous regressors, but treats all regressors as potentially correlated with c_i . Since in the example only z_{it} is correlated with c_i we would not need to demean the remaining regressors, which would discard the *between* variance.

⁸Originally, the HT estimator was developed to estimate the coefficients of time-invariant variables in a fixed-effects estimation.

A random-effects IV regression à la Hausman and Taylor (1981) requires the following steps. First, regressors to be treated as exogenous need to be determined. This can be done by economic arguments and using Hausman tests on the individual regressors, taking into account possible distortions for those regressors with a relatively small *within* variance. Second, a random-effects IV regression is run and diagnostic tests are used to choose a particular model specification.

5 Estimation results from panel models

5.1 Fixed- and random-effects estimation

We start by running fixed- and random-effects regressions over the estimation period 2004 to 2010 and check whether the results confirm our conjectures. Columns labeled FE in Table 2 present the results from the fixed-effects estimation of (14) for single-family houses and apartments separately. Regressors without significant effects were removed from the specification.⁹ Since variables are not logged the coefficients do not represent elasticities, meaning that their relative magnitude is not easily interpreted. We focus on the signs and significance of coefficients throughout the discussion. The signs of the coefficients on the housing stock, per-capita income, the demographic variable and growth expectations can be given a plausible interpretation in all specifications, while the coefficients on the density variable and the unemployment rate are signed differently across specifications. The mortgage rate does not appear to be significant in any of the specifications, whereas growth expectations have a consistently positive effect on house prices across specifications. As expected, the fixed-effects model estimates a positive effect of the inverted population density and the unemployment rate because these correlations happened to be dominant over time in the estimation period. Note that the *within* R^2 is much higher than the *between* R^2 . The fixed-effects model does not explain much of the variation across districts (except via fixed effects, which have no economic interpretation), while most of its explanatory power is directed at the within variation over time.

⁹Results for the full list of variables are in the appendix.

Table 2

Estimation of (14) 2004 – 2010, 402 districts				
Variable	Apartments		Single-family houses	
	RE	FE	RE	FE
s_{it}	-618.0*** (195.9)	-4068.6*** (1042.7)	-2056.8*** (391.6)	-11902.5*** (1359.1)
y_{it}	0.004*** (0.001)	0.003** (0.001)	0.002* (0.001)	–
a_{it}	14049.4*** (609.5)	11498.7*** (912.0)	14288.8*** (747.9)	4897.0*** (1304.7)
d_{it}	-20446.0*** (4573.8)	–	-22676.6*** (5661.2)	1820916.0*** (38999.0)
u_{it}	–	1213.5*** (443.1)	-994.4** (413.8)	1425.8*** (400.3)
r_t	–	–	–	–
g_t^e	7845.7*** (7845.7)	8119.4*** (1135.5)	4729.9*** (1213.9)	7497.0*** (1098.8)
<i>constant</i>	-3641.7*** (255.5)	-1896.0*** (544.8)	-2952.5*** (292.7)	1244.1* (720.9)
R^2				
<i>within</i>	0.55	0.56	0.50	0.56
<i>between</i>	0.23	0.02	0.34	0.02
<i>overall</i>	0.26	0.03	0.35	0.01
Obs.	2814	2814	2814	2814

Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.

In contrast, the random-effects model reported in columns labeled RE is able to explain much better the cross-sectional variation in the panel. The signs of all variables are intuitive for both the apartments and the single-family houses equations. For any given level of demand the house price level is lower the higher the supply of housing in form of the housing stock at the beginning of the period. Higher per-capita income, a larger share of middle-aged population and more favourable growth expectations put pressure on house prices, while less densely populated areas and those with higher unemployment are associated with lower house prices. According to the *overall* R^2 of the random-effects model, single-family house prices can be explained better than apartment prices.

5.2 Hausman-Taylor instrumental variables estimation

In order to address the endogeneity issue when using a random-effects model we specify a model using the HT estimator. We assume that the real interest rate and the economy-wide growth expectations are uncorrelated with c_i because they are unaffected by factors specific to an individual district. For the remaining variables we we rely on a combination of economic reasoning and mechanical testing to identify exogenous regressors. First, we

run a Hausman test on all district-specific regressors jointly and a Hausman test on each variable individually.¹⁰ Given that the standard Hausman test statistic might be affected by differing coefficient signs due to variation from the time- vs. cross-section dimension, or by large differences in the variances of the estimated coefficients, we additionally present results from another version of the Hausman test (Wooldridge, 2002).¹¹ It is less susceptible to the distortions from which the standard formulation of the Hausman test suffers. The alternative version of the Hausman test (Hausman-Wu test) is based on the idea that the unobserved effects c_i can be decomposed into a part that is correlated with e.g. the district-specific means of the regressors and a remaining random effect μ_i (Mundlak, 1978).

$$c_i = \lambda_1 \bar{s}_i + \lambda_2 \bar{y}_i + \lambda_3 \bar{z}_i + \mu_i \quad (17)$$

where a bar denotes the district-mean of a variable. The test is constructed by augmenting a random-effects estimation equation by e.g. the means of the regressors (e.g. \bar{s}_i).

$$\begin{aligned} \tilde{p}_{it} = & \beta_0 + \beta_1 \tilde{s}_{it} + \beta_2 \tilde{y}_{it} + \beta_3 \tilde{z}_{it} + \beta_4 \tilde{r}_t + \beta_5 \tilde{g}_t^e \\ & + \lambda_1 \bar{s}_i + \lambda_2 \bar{y}_i + \lambda_3 \bar{z}_i + \mu_i + \tau_{it} \end{aligned} \quad (18)$$

where a tilde denotes a quasi-time-demeaned variable. The null hypothesis is $\lambda_j = 0$, $j = 1, 2, 3$ jointly or individually. A significant test statistic denotes rejection of the null of exogeneity.

Table 3 contains the Hausman test statistics for all variables jointly and individually. For both types of residential property the joint standard Hausman tests strongly reject the null that the regressors are jointly uncorrelated with the unobserved effects. The individual test statistics are subject to a high degree of variation, in particular those for single-family houses show a fairly large difference between the test statistics with the chi-squared statistic for income being far smaller than for the remaining variables. The results from the Hausman-Wu test indicate that income in the single-family house price equation (and marginally in the apartments equation), as well as the age-cohort variable and the inverse population density in the apartments equation appear to pass the test for exogeneity. From an economic perspective a variable with a relatively large transitory component such as current income is less likely to be correlated with a time-invariant unobserved characteristic (e.g. quality of infrastructure) than other more structural factors such as the size of an age-cohort or population density. Moreover, we would expect an regressor to be exogenous in both the equation for single-family houses and apartments

¹⁰We use autocorrelation-robust variance estimates in the Hausman tests throughout. In our case the trade-off between the fixed-effects and random-effects estimator is due to the random-effects estimator being better suited to identify the structural relationships from the *between* variance, rather than due to its efficiency properties.

¹¹Ch. 10.7.3

Table 3

Hausman-Taylor estimation of (14) 2004 – 2010, 402 districts								
Variable	Apartments				Single-family houses			
	Hausman test statistic	Hausman-Wu test statistic	HT/IV-1	HT/IV-2	Hausman test statistic	Hausman-Wu test statistic	HT/IV-1	HT/IV-2
jointly	157.6				395.5			
s_{it}	12.8	-8.7	-2648.1*** (373.9)	-3136.1*** (381.4)	51.1	-6.8	-12567.6*** (907.2)	-3661.9*** (960.0)
y_{it}	4.9	2.0	0.004*** (0.001)	0.003*** (0.001)	6.8	1.9	0.003** (0.001)	-
a_{it}	7.8	0.4	11646.1*** (523.6)	12326.0*** (411.1)	81.2	-2.7	4256.6*** (879.4)	10246.7*** (649.8)
d_{it}	4.6	1.9	-17670.0* (10123.7)	-95035.7*** (12202.4)	28.7	7.4	148191.5*** (16679.5)	-119636.6*** (32014.9)
u_{it}	15.0	14.6	1303.7*** (291.1)	-	342.3	19.8	1221.2*** (394.1)	-
r_t	-	-	662.4 (423.9)	-	-	-	246.2 (582.8)	-
g_t^e	-	-	7487.8*** (1267.5)	7259.4*** (1272.2)	-	-	6235.1*** (1733.8)	6399.5*** (1301.2)
<i>constant</i>			-2285.8*** (245.8)	-1849.7*** (274.6)			1740.1*** (403.6)	-598.2 (330.8)
R^2								
<i>within</i>			0.56	0.55			0.56	0.49
<i>between</i>			0.07	0.13			<0.01	0.22
<i>overall</i>			0.09	0.15			<0.01	0.22
Obs.			2814	2814			2814	2814

Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.
Hausman test statistic distributed as chi-squared, Hausman-Wu statistic as t-distribution.
Significant values (on the 5%-level) in bold denote rejection of the null of exogeneity.
Instruments: HT/IV-1: District means of exogenous regressors and demeaned endogenous regressors.
HT/IV-2: as HT/IV-1 less demeaned regressors whose within-correlation with prices differs in sign from their between-correlation.

or none given that the included regressors are not specific to the type of property. The Hausman-Wu test rejects exogeneity of current income only marginally in the apartments equation, whereas it strongly rejects it for the age-cohort and inverse population density variable in the apartments equation. Therefore we consider it reasonable to treat income as an exogenous regressor and all other district-specific variables as endogenous regressors.

The column labeled HT/IV-1 in Table 3 presents the estimation results for the full list of variables using the district means of the exogenous regressors (including r_t and g_t^e) along with the demeaned endogenous regressors as instruments. Columns labeled HT/IV-2 present the estimation results from stripping down the specification by means of dropping insignificant variables and removing those instruments from the first-stage regression whose within-correlation with house and apartment prices differs in sign from their between-correlation. The interpretation of the coefficients for the remaining regressors is the same as in the standard random-effects model. Note that similarly to the standard random-effects estimation the HT/IV-approach has better explanatory power than the fixed-effects estimator. Compared to the RE estimator the HT estimator reduces the *overall* R^2 of both types of property. However, it accounts for potential endogeneity of some of the regressors, which the random-effects estimator fails to do.

Extending the estimation period back to 1996 we follow the same estimation procedure as in the baseline case for 402 districts for selecting the appropriate specification. The empirically positive correlation between mortgage rates and growth expectations (see Fig. 6) might be due to long-run interest rates reflecting expectations about future productivity. Therefore, we consider house prices and 10-year mortgage rates to be partly influenced by long-run growth expectations as a common factor, and do not include mortgage rates in this part of our analysis. Table 4 contains the detailed estimation results. Note that the Hausman/Hausman-Wu test results are in line with our reasoning for the shorter sample period. In particular, current income in the apartments equation, too, is indicated by the tests as exogenous. The coefficients on the basis of the HT/IV estimator are all signed according to economic intuition and in line with the results for the shorter sample period using all 402 districts (columns labeled HT/IV-2 in Table 3). In line with our prior, the coefficient attached to growth expectations increases markedly in size compared to the shorter estimation sample. It is likely to be a better estimate for the structural long-run price because over the shorter sample the correlation between residential property prices and growth expectations appears much looser than over the longer sample due to heightened uncertainty about the longer-term outlook during the economic and financial crisis (see Fig. 6). Therefore, it would not be adequate to expect the coefficient on long-run growth expectations from the shorter sample to help explain the development of house prices over the longer period satisfactorily.

Table 4

Estimation of (14) 1996 – 2010, 93 towns and cities										
Variable	Apartments					Single-family houses				
	Hausman test statistic	Hausman-Wu test statistic	RE	FE	HT/IV	Hausman test statistic	Hausman-Wu test statistic	RE	FE	HT/IV
jointly	301.0					257.5				
s_{it}	55.8	-11.9	-2946.3*** (469.0)	-6288.6*** (566.9)	-5248.4*** (364.2)	24.6	-4.8	-8528.7*** (1097.7)	-15022.2*** (1419.2)	-5320.1*** (1467.3)
y_{it}	2.6	0.6	–	–	–	26.9	1.1	–	–	–
a_{it}	6.0	-3.9	3679.4*** (1163.9)	2428.9* (1290.4)	5124.1*** (834.4)	23.6	-2.9	1813.6** (789.9)	–	1827.0*** (571.4)
d_{it}	8.0	2.6	-40768.5*** (7237.0)	128485.4** (57369.7)	–	11.3	3.6	–	278385.4*** (84695.1)	-249538.8*** (72868.6)
u_{it}	(-42.1)	8.8	2088.0*** (805.6)	3932.2*** (772.6)	-3708.5** (1453.5)	33.6	11.0	–	1361.5*** (623.9)	–
r_t	–	–	–	–	–	–	–	–	–	–
g_t^c	–	–	43573.2*** (2122.2)	39858.2*** (2233.1)	39373.2*** (1424.0)	–	–	28388.0*** (2082.7)	23331.39*** (2209.9)	30439.5*** (1691.5)
<i>constant</i>			696.1*** (423.2)	2280.0*** (501.7)	1485.7*** (290.2)			1807.3*** (342.1)	2856.9*** (128.0)	1770.1*** (256.6)
R^2										
<i>within</i>			0.70	0.73	0.67			0.64	0.67	0.59
<i>between</i>			0.01	0.03	0.03			0.09	0.03	0.16
<i>overall</i>			0.23	<0.01	0.15			0.15	0.01	0.18
Obs.			1395	1395	1395			1395	1395	1395

Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.
Hausman test statistic distributed as chi-squared, Hausman-Wu statistic as t-distribution..
Significant values (on the 5%-level) *in bold* denote rejection of the null of exogeneity.
Instruments: District means of exogenous regressors and demeaned endogenous regressors less demeaned regressors whose within-correlation with prices differs in sign from their between-correlation.

6 Assessing aggregate house prices

6.1 Aggregating district-level regression residuals

Given that the district-level price equation is specified to ensure consistent aggregation the panel estimates can be used to assess residential property prices at aggregate levels. This is done by summing over district-level regression residuals $\hat{\epsilon}_{it}$.

$$\begin{aligned}\hat{\epsilon}_t &= \sum_{i=1}^I \frac{n_{it}}{n_t} \hat{\epsilon}_{it} \\ &= \sum_{i=1}^I \frac{n_{it}}{n_t} p_{it} - \sum_{i=1}^I \frac{n_{it}}{n_t} \hat{p}_{it} \\ &= p_t - \hat{p}_t\end{aligned}\tag{19}$$

Therefore the aggregate residual $\hat{\epsilon}_t$ can be interpreted as the deviation of the actual aggregate house price from its fundamental level.

The estimation of district-specific effects implies $\hat{\epsilon}_i = \sum_{t=1}^T \hat{\epsilon}_{it} = 0$. Hence, the model imposes the (time-series) mean of any residual series, district-specific or aggregate, to be zero in sample.

Based on the variance of the district-specific prediction errors $Var\{\hat{\epsilon}_{it}\} = Var\{p_{it} - \hat{p}_{it}\}$, we can also provide confidence intervals for the estimated aggregate residual according to

$$\begin{aligned}Var\{\hat{\epsilon}_t\} &= Var\left\{\sum_{i=1}^I \frac{n_{it}}{n_t} \hat{\epsilon}_{it}\right\} \\ &= \sum_{i=1}^I \left(\frac{n_{it}}{n_t}\right)^2 Var\{\hat{\epsilon}_{it}\}\end{aligned}\tag{20}$$

given that the ϵ_{it} 's are assumed independent of each other across districts. The confidence bands comprise the sampling variability and parameter uncertainty.

6.2 Deviations from the fundamental equilibrium price

Fig. 7 shows the aggregate regression residuals in percent of the fundamental price estimated on the basis of the whole panel ($I = 402$) over the sample period 2004 to 2010 for single-family house and apartment prices.¹² Before 2011, the in-sample deviations with their associated confidence intervals are plotted. For 2011 and 2012 the out-of-sample prediction error variances are proxied by the average of the in-sample prediction error variance. It has to be borne in mind that the confidence bands over the forecasting period based on the in-sample variances might underestimate the true out-of-sample prediction error variance.

¹²The projections of the regressors for 2011 and 2012 are described in appendix A.

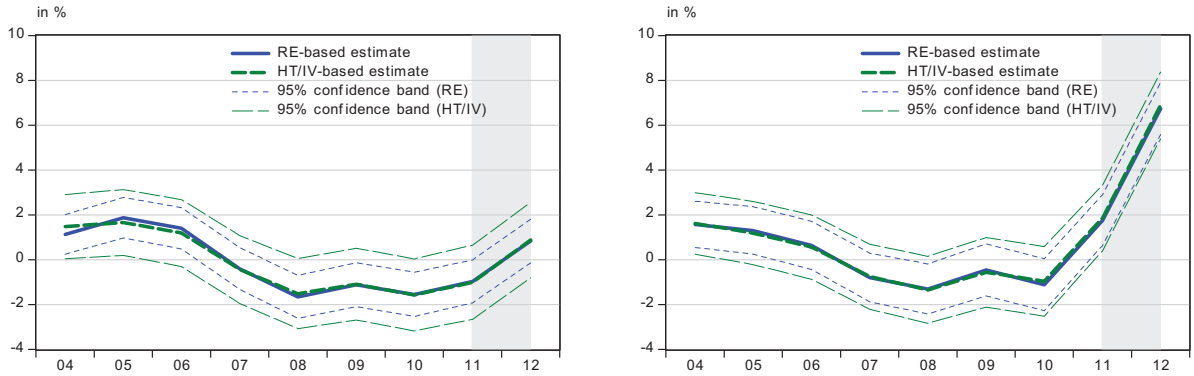


Figure 7: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 402 districts, estimated using data for 402 districts over 2004 to 2010. Shaded areas denote out-of-sample predictions.

On the basis of the two estimators, RE and HT/IV-2, there are two measures of deviations from estimated levels. Over the estimation sample 2004 to 2010 the price misalignments are of modest size and fluctuate around zero. Over the past two years single-family house prices do not show any signs of an over- nor undervaluation on a whole-country average. This holds regardless of the estimator used. In contrast, apartment prices appear to significantly exceed their fundamental levels by around 7%.

Our chosen aggregation procedure is rather flexible in that it allows to vary the scope of regional aggregation. Next to a whole-country aggregate, results for different regional subsets can be obtained, e.g. for 93 towns and cities, which are separately identified in the dataset, or for the major seven cities. Producing different regional composites is convenient because, currently, the degree of house price increases varies across regional subsets. Thus, the relatively strong price increases in the major seven cities might be explained by comparatively favourable developments in the driving factors. Fig. 8 plots the deviations of house and apartment prices for 93 towns and cities. During the past two years single-family house prices in towns and cities appear to exceed their predicted prices by more than on a whole-country average, however by less than 4% and not significantly so. A more pronounced result emerges for apartment prices, whose current deviation from the long-run price is more than 12%.

Finally, we run the same exercise for the subset of the major seven German cities, which is presented in Fig. 9.¹³ The results indicate that single-family house prices in the major cities are above their predicted prices (around 5%), however, subject to high statistical uncertainty. In contrast, the misalignment of apartment prices is highly significant and

¹³The selected cities are Berlin, Düsseldorf, Frankfurt am Main, Hamburg, Cologne, Munich and Stuttgart.

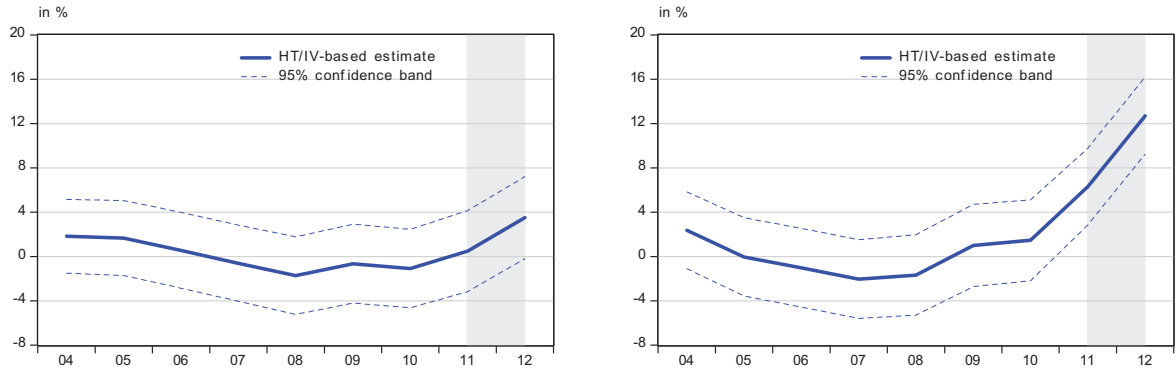


Figure 8: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 93 cities, estimated using data for 402 districts over 2004 to 2010. Shaded areas denote out-of-sample predictions.

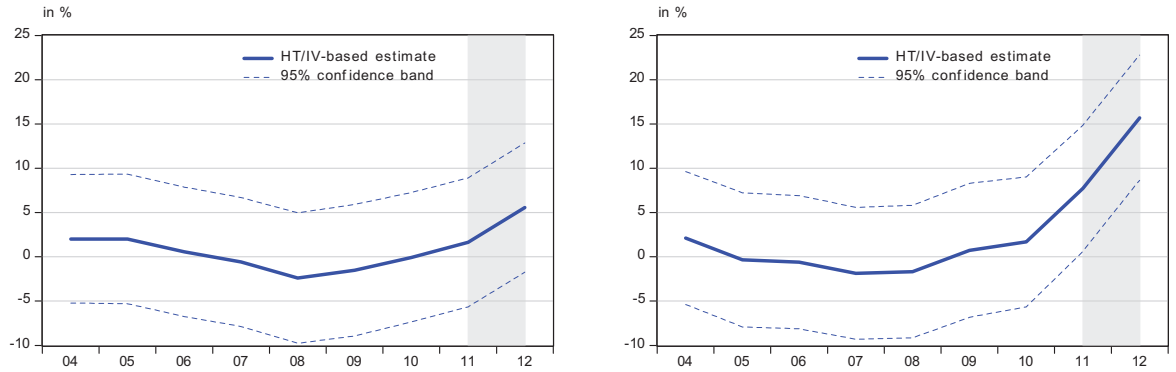


Figure 9: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for major 7 German cities, estimated using data for 402 districts over 2004 to 2010. Shaded areas denote out-of-sample predictions.

reaches more than 15% in 2012.

Extending the sample period backwards to 1996 for 93 towns and cities allows to compare the recent price misalignments in these cities to the situation in the mid-1990s, when residential property prices reached their previous maximum. By integrating the episode in the information set for estimation we make it in principle harder for the model to indicate deviations from the fundamental equilibrium at the current juncture. Letting the model adjust as well as possible to strong house price increases in the past should make the recent price changes look less unusual, provided these episodes are similar in nature. Recall that the longer term perspective comes at the expense of a reduced cross-section sample of 93 towns and cities as opposed to 402 districts covering the whole country.

According to the results for the whole subset of 93 towns and cities presented in Figs. 10 and 11 the qualitative assessment of house prices at the end of the sample does not

change. Apartments in German cities show significant overvaluations, while single-family houses are not subject to any misalignment. Quantitatively, the overvaluation at the sample end is viewed a little less pronounced with around 6% for apartments in 93 towns and cities and around 12% in the major seven cities.

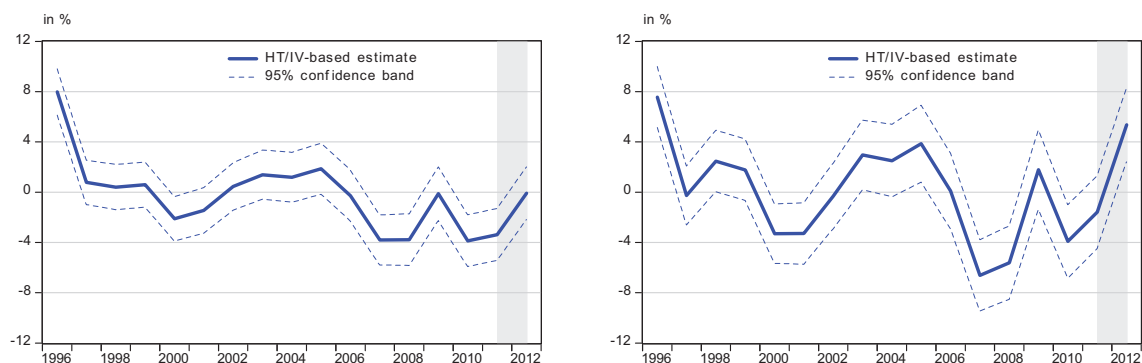


Figure 10: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 93 cities estimated using data for 93 cities from 1996 to 2010. Shaded areas denote out-of-sample predictions.

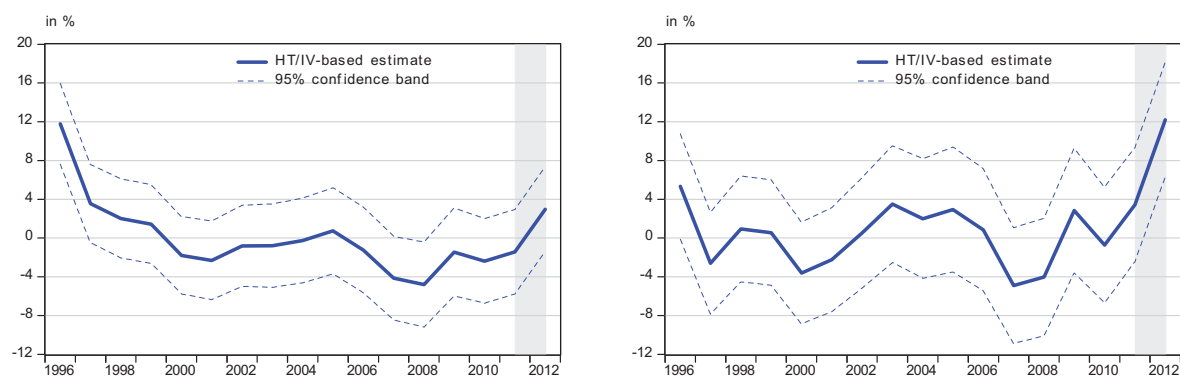


Figure 11: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for major 7 cities estimated using data for 93 cities from 1996 to 2010. Shaded areas denote out-of-sample predictions.

For the mid-1990s a significant overvaluation of both apartments and houses is documented for 93 towns and cities. This is also true of the houses in the major seven cities but interestingly not of the apartments in this subset. The overvaluations observed for apartment prices in 93 towns and cities are of a magnitude comparable to the mid-1990s, while apartment price misalignments in the major seven cities are currently considerably more pronounced than in the past. The period between 1997 and 2011 is not marked as a period of price misalignments. This is mainly due to the role of growth expectations,

which become quite closely aligned with the development of house prices from 1997 onwards (see Fig. 6). Conversely, the end of the sample growth expectations have stabilised on a relatively low level, while house and in particular apartment prices kept increasing from their comparatively low levels.

So far, we have assessed house price increases over the past two years against a fundamental price that is estimated over a period which excludes the years of particular interest (2011 and 2012). The benefit is that the fitted long-run price is not pulled towards actual prices, possibly failing to detect a significant misalignment. After all, the estimator is based on minimizing the deviations of actual from estimated prices. However, we might unduly disregard the information about the recent past, when in fact it should have an impact on the fundamental price. The reported house price deviations might then overstate the true misalignment. In appendix B we repeat the estimation for the extended samples up to 2012 for both sets of districts and cities, and show that, while the estimated overvaluation at the end of the sample is indeed slightly reduced, the results hold up qualitatively, in particular in terms of significance.

6.3 The role of interest rates

Although the interest rate is an essential element in the theoretical model, it is difficult to empirically pin down its effect on house prices. Over the sample period 2004 to 2010 mortgage rates do not have a significant effect on house prices in our final empirical specifications, while over the longer sample 1996 to 2010 they are significantly positively correlated with house prices (see Fig. 6). The latter correlation could be explained by the common effect of growth expectations on house prices and interest rates. However, at the sample end growth expectations and interest rates are no longer positively correlated. Thus, the effect of mortgage rates on house prices on the basis of their historical correlation would lead to counterintuitive results. Therefore, mortgage rates are removed from the set of regressors.

In order to include interest rate effects despite the econometric difficulties, we consider the following solution. Motivated by (6) of the theoretical model we assume that the hypothetical elasticity of house prices with respect to the mortgage rate equals the growth expectations elasticity with a negative sign (without however restricting the elasticity to any particular value).

$$\frac{\partial p_{it}}{\partial r_t} \frac{r_t}{p_{it}} = - \frac{\partial p_{it}}{\partial g_t^e} \frac{g_t^e}{p_{it}} \quad (21)$$

Given the estimated coefficient on growth expectations $\hat{\beta}_5 = \frac{\partial \hat{p}_{it}}{\partial g_t^e}$ we can calculate a value

for $\tilde{\beta}_4 = \frac{\partial \hat{p}_{it}}{\partial r_t}$ according to

$$\tilde{\beta}_4 = - \left. \frac{\partial \hat{p}_{it} g_t^e}{\partial g_t^e r_t} \right|_{\bar{g}^e, \bar{r}} \quad (22)$$

$$= -\hat{\beta}_5 \frac{\bar{g}^e}{\bar{r}} \quad (23)$$

evaluated at the sample average.¹⁴ We plug the additional mortgage rate term $\tilde{\beta}_4 (r_t - \bar{r})$ into the fitted equation (14) giving

$$\hat{p}_{it} = \beta_0 + \hat{\beta}_1 s_{it} + \hat{\beta}_2 y_{it} + \hat{\beta}_3 z_{it} + \tilde{\beta}_4 (r_t - \bar{r}) + \hat{\beta}_5 g_t^e + \hat{c}_i \quad (24)$$

In order to preserve the estimate for the intercept we use the deviation of mortgage rates from their sample mean.

Repeating the assessment exercise on the basis of the baseline estimation sample 2004 to 2010 including this modification reduces the moderate deviation of apartment prices in recent years on a whole-country average to negligible values (Fig. 12).¹⁵ Also, in the subsets with 93 towns and cities and the major seven cities single-family house prices appear to be closer to the fundamental prices. However, although apartment price deviations in cities are reduced by about 5 percentage points when mortgage rates are accounted for they still show signs of sizeable misalignments (Figs. 13 and 14).

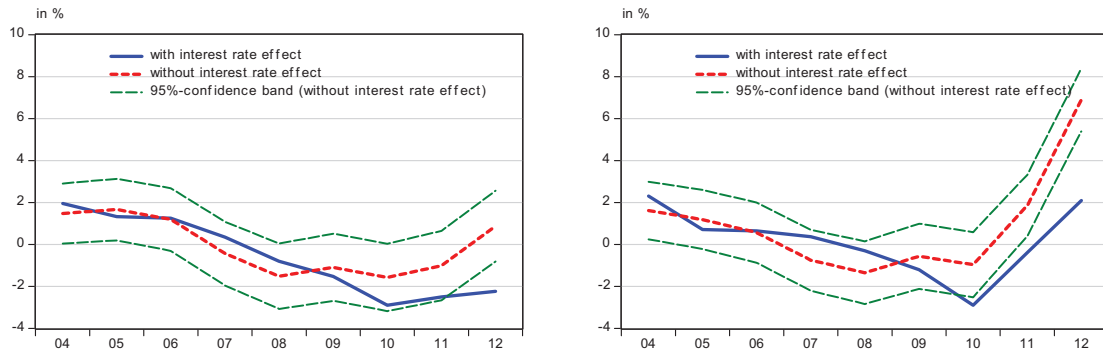


Figure 12: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 402 German districts adjusted for mortgage-rate effect by calibration.

We conclude from this calibration exercise that, for the whole-country average, low mortgage rates appear to be an important factor behind the recent price increases. However, for German towns and cities the low level of mortgage rates cannot fully explain

¹⁴In the special case where $\bar{r} = \bar{g}^e$ the partial effect of growth expectations is equal to the negative partial effect of the mortgage rate.

¹⁵Factoring in mortgage rates by calibration over the longer sample 1996 to 2010 is more difficult due to the empirically strong positive correlation of mortgage rates, growth expectations and house prices over this period.

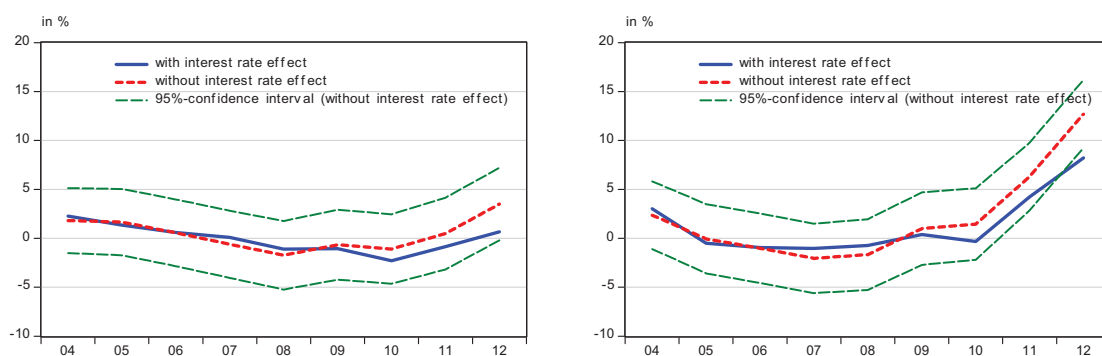


Figure 13: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 93 cities adjusted for mortgage-rate effect by calibration.

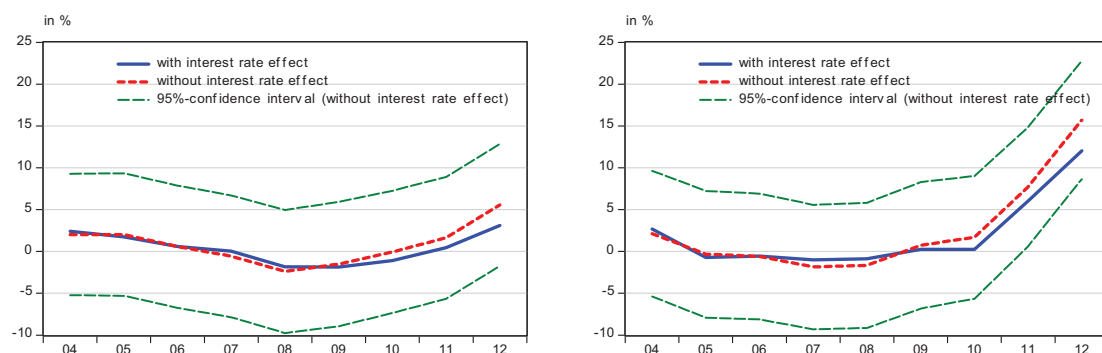


Figure 14: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for the major seven German cities adjusted for mortgage-rate effect by calibration.

the marked increase of apartment prices and to some extent of single-family house prices leaving room for factors, which are not accounted for in this analysis.

7 Conclusion

The main goal of the paper is to examine to what extent the recent price increases on the German housing market are attributable to favourable developments in macroeconomic factors, or whether there is any significant misalignment. Based on a stock-flow model for the housing market we have derived an equation for fundamental equilibrium house prices, which relates residential property prices to explanatory variables. To estimate the equation we make use of a regional panel dataset comprising price data for single-family houses and apartments in all 402 districts in Germany and district-specific explanatory variables. The panel estimation results confirm our expectation that an approach based on the cross-section variance of prices and regressors yields more plausibly signed coefficients

attached to district-specific regressors than a pure time-series framework.

We assess recent and past episodes of strong house price increases by aggregating the district-specific regression residuals into deviations of aggregate house prices from their fundamental levels. We conclude from our estimation results that, for Germany as a whole, there are no signs of a significant or sizeable misalignment of single-family house prices compared to their long-run level. However, apartment prices seem to be overvalued at the current juncture by 5% to 7% on a whole-country average. For the subsets of 93 towns cities and the major seven cities our approach suggests a moderate and statistically insignificant deviation of single-family house prices from their equilibrium levels. In contrast, apartment prices in German cities appear to be markedly overvalued. In the 93 towns and cities, the size of the price deviations at the current juncture are of comparable size to the price misalignments during the mid-1990s, while the overvaluation of apartments in the major seven cities is considerably larger than during the mid-1990s. While the exceptionally low level of mortgage rates technically lower the estimated overvaluation of apartments on a whole-country average close to equilibrium, our assessment of apartment prices in German cities does not change when we factor in the effect of interest rates, when we include the assessment period in the estimation sample, or extend the estimation period back to 1996.

Our assessment of residential property prices is subject to an important caveat, which applies to empirical analyses that aim at estimating a reference value which, from a conceptual point of view, can be interpreted as a fundamental equilibrium level. As in a standard time-series approach, "substituting" time-series information by a large regional dataset requires the assumption that the sample mean of the price observations equals the true unconditional average of house prices. Given the relatively short time dimension and the substantive role played by growth expectations without cross-sectional variation, however, the assumption of zero (time-series) residual means, which is imposed by estimation, cannot be evaluated without further information from outside the chosen framework.

Finally, it is beyond the scope of the paper to model regional spillovers in house prices by including spatial effects in a panel setup or even adopting a spatio-temporal approach like the one proposed by Holly et. al (2010).

References

- [1] Adams, Z. and R. Füss (2010), Macroeconomic Determinants of International Housing Markets. *Journal of Housing Economics* 19, 38-50.
- [2] Altissimo, F., E. Georgiou, T. Sastre, M. Valderrama, G. Sterne, M. Stocker, M. Weth, K. Whelan and A. Willman (2005), Wealth and Asset Price Effects on Eco-

- conomic Activity, Occasional Paper Series 29, European Central Bank.
- [3] Bischoff, O. (2012), Explaining Regional Variation in Equilibrium Real Estate Prices and Income. *Journal of Housing Economics* 21 (1), 1-15.
 - [4] Borio, C. and P. Lowe (2002), Asset prices, Financial and Monetary Stability: Exploring the Nexus. Bank of International Settlements Working Paper No. 114.
 - [5] Cameron, G. J. Muellbauer and A. Murphy (2006), Was there a British House Price Bubble? Evidence from a Regional Panel. University of Oxford, Department of Economics Discussion Paper No. 276.
 - [6] Capozza D. P. Hendershott, C. Mack and C. Mayer (2002), Determinants of Real House Price Dynamics. National Bureau of Economic Research Working Paper No. 9262.
 - [7] Capozza D., P. Hendershott and C. Mack (2004), An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets. *Real Estate Economics* 32, 1-32.
 - [8] Case, K., J. Quigley and R. Shiller (2005), Comparing Wealth Effects: The Stock Market versus the Housing Market. National Bureau of Economic Research Working Paper 8606.
 - [9] Catte, P., N. Girouard, R. Price, C. André (2004), Housing markets, Wealth and the Business Cycle, OECD Economics Department Working Paper ECO/WKP(2004)17.
 - [10] DiPasquale, D. and W. Wheaton (1992), The Markets for Real Estate Assets and Space: A Conceptual Framework. *Journal of the American Real Estate and Urban Economics Association* 20, 181-197.
 - [11] Dreger, C. and K. Kholodilin (2012), An Early Warning System to Predict the Speculative House Price Bubbles, *Economics: The Open-Access, Open-Assessment E-Journal* 7, 2013-8.
 - [12] European Commission (2012), Assessing the Dynamics of House Prices in the Euro Area. Quarterly Report on the Euro Area 4.
 - [13] Gerdesmeier, D., Lenarcic, A. and B. Roffia (2012), An Alternative Method for Identifying Booms and Busts in the Euro Area Housing Market. European Central Bank Working Paper No. 1493.

- [14] Girouard, N., M. Kennedy, P. van den Noord, C. André (2006), Recent House Price Developments. The Role of Fundamentals, OECD Economics Department Working Paper 475.
- [15] Glaeser, E., Gyourko, J. and A. Saiz (2008), Housing Supply and Housing Bubbles. *Journal of Urban Economics* 64 (2), 198-217.
- [16] Hausman, J. (1978), Specification Tests in Econometrics, *Econometrica* 46, 1251-1271.
- [17] Hausman, J. and W. Taylor (1981), Panel Data and Unobservable Individual Effects. *Econometrica* 49 (6), 1377-1398.
- [18] Hiebert, P. and M. Sydow (2011), What Drives Returns to Euro Area Housing? Evidence from a Dynamic Dividend-Discount Model. *Journal of Urban Economics* 70 (2-3), 88-98.
- [19] Hiebert, P. and I. Vansteenkiste (2011), Do House Price Developments Spill Over Across Euro Area Countries? Evidence from a global VAR. *Journal of Housing Economics* 20 (4), 299–314.
- [20] Himmelberg, C, Mayer, C. and T. Sinai (2005), Assessing High House Prices: Bubbles, Fundamentals and Misperceptions. *Journal of Economic Perspectives* 19 (4), 67-92.
- [21] Holly, S., Pesaran, H. and T. Yamagata (2011), A Spatio-Temporal Model of House Prices in the USA. *Journal of Econometrics* 158, 160-173.
- [22] Hwang, M. and J.M. Quigley (2006), Economic Fundamentals in Local Housing Markets: Evidence from US Metropolitan Regions. *Journal of Regional Science* 46 (3), 425-453.
- [23] Iacoviello, M. (2004), Consumption, House Prices, and Collateral Constraints: A Structural Econometric Analysis. *Journal of Housing Economics* 13, 304-320.
- [24] Iacoviello, M. and R. Minetti (2008), The Credit Channel of Monetary Policy: Evidence from the Housing Market. *Journal of Macroeconomics* 30, 69-96.
- [25] Igan, D. and P. Loungani (2012), Global Housing Cycles. IMF Working Paper WP/12/217.
- [26] Kawka, R. (2009), Regionaler Preisindex. Bundesinstitut für Bau-, Stadt- und Raumforschung, Berichte Band 30.

- [27] Kholodilin, K. and A. Mense (2012), Forecasting the Prices and Rents for Flats in Large German Cities. German Institute for Economic Research DIW Discussion Paper.
- [28] Koetter, M. and T. Poghosyan (2010), Real Estate Prices and Bank Stability. *Journal of Banking and Finance* 34, 1129-1138.
- [29] Maennig, W. and L. Dust (2008), Shrinking and Growing Metropolitan Areas – Asymmetric Price Reactions? The Case of German Single-family Houses. *Regional Science and Urban Economics* 38 (1), 63-69.
- [30] Malpezzi, S. (1999), A Simple Error Correction Model of House Prices. *Journal of Housing Economics* 8, 27-62.
- [31] McCarthy, J. and R. Peach (2002), Monetary Policy Transmission to Residential Investment. Federal Reserve Bank of New York Economic Policy Review May 2002.
- [32] McQuinn, K. and G. O’Reilly (2008), Assessing the Tole of Income and Interest Rates in Determining House Prices. *Economic Modelling* 25, 377-390.
- [33] Muellbauer, J. and A. Murphy (1994), Explaining Regional House Prices in the UK. Department of Economics Working Paper 94/29, University College Dublin.
- [34] Mundlak, Y. (1978), On the Pooling of Time Series and Cross Section Data, *Econometrica* 46, 69-85.
- [35] Poterba, J. (1984), Tax Subsidies to Owner-occupied Housing: An Asset Market Approach. *Quarterly Journal of Economics* 99 (4), 729-752.
- [36] Steiner, E. (2010), Estimating a Stock-Flow Model for the Swiss Housing Market. *Swiss Journal of Economics and Statistics* 146 (III), 601-627.
- [37] Wooldridge, J. (2002), *Econometric Analysis of Cross Section Panel Data*. 1st Edition, MIT Press.

8 Appendix

8.A Data

In order to assess the size and significance of house price increases from 2004 to 2012 we fit fundamental prices using the actual values of explanatory variables and the parameter values estimated over the period 2004 to 2010. In the out-of-sample period 2011 to 2012 actual data on district level are available for all variables except GDP. For 2012 there are

no actual data for any explanatory variable available yet. Therefore we project the missing district-level variables for 2011 and 2012 by assuming that their shares in each aggregate variable changes at the same rate as in 2010 and 2011, respectively. It is assumed that the process of structural change at district level continues at a constant pace. E.g. some districts' GDP shares keep declining while others' continue to increase. Note that as more years of data at district level become available we can get rid of this assumption. Table 5 presents definitions and sources for the data.

8.B Additional regression results

Table 6 contains the results of the random- and fixed-effects regressions with the full set of variables discussed in section 5.

In addition, we repeat the estimation for the sample 2004 to 2012 using the same specification as for the sample 2004 to 2010 as discussed in section 6.2. Tables 7 and 8 present the estimation results using the HT/IV estimator over the extended sample from 2004 to 2012 for all 402 districts, and from 1996 to 2012 for 93 towns and cities. Note that the imputed values of regressors for which data are not yet available are used. This implies that the uncertainty surrounding the estimated long-run price for 2011 and 2012 might be higher than reported.

Figs. 15 to 17 show the resulting price deviations from the fundamental price estimated using all districts over 2004 to 2012.

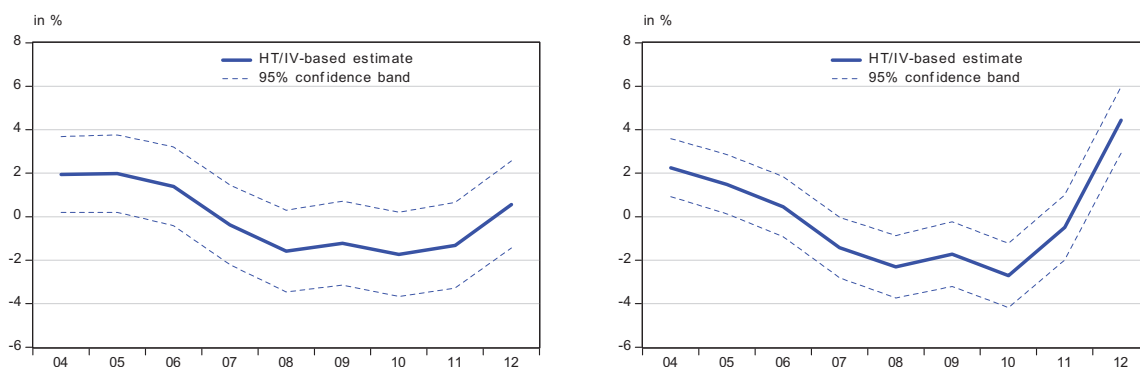


Figure 15: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 402 districts, estimated using data for 402 districts over 2004 to 2012.

Compared to Figs. 7 to 9 there are no qualitative differences in the assessment of residential property prices for the different aggregates. Quantitatively, including the recent past in the estimation sample reduces the estimated overvaluation for apartments by around 2 percentage points. On a whole country average, apartment prices are still

Table 5: Data definitions and sources

Variable	Definition	Source
House prices	Prices per sqm in euros for single-family houses and apartments	BulwienGesa AG for 402 German administrative districts
Housing stock	Housing units of single-family houses and apartments	National Statistical Office (<i>Bauen und Wohnen, Baugenehmigungen/ Baufertigstellungen 2011</i>)
GDP per capita	GDP per capita in euros	National Accounts of the German <i>Länder</i>
Demographic variable	Population aged 30 to 55 in % of total	Regional database at the National Statistical Office
Density measure	Area in square kilometres per inhabitant	Regional database at the National Statistical Office
Unemployment rate	Unemployed in % of total population	Regional database at the National Statistical Office and Federal Employment Agency
Real mortgage rate	Nominal 10-year mortgage rates minus expected inflation over the 10 years ahead	Official banking data statistics for nominal mortgage rates, identical across districts
Inflation expectations	Inflation expectations over the 10 years ahead	Consensus forecast for 1990 to 2012, identical across districts
Real growth expectations	Survey data on expectations for real GDP growth over the 10 years ahead	Consensus forecast for 1990 to 2012, identical across districts
CPI	District-specific CPIs for 2009 based on data provided by Kawka (2009); projected for remaining periods using state-level CPIs	Kawka (2009); regional database at National Statistical Office

Table 6

Variable	Estimation of (14) 2004 – 2010, 402 districts			
	Apartments		Single-family houses	
	RE	FE	RE	FE
s_{it}	-588.2*** (196.0)	-4459.5*** (1100.7)	-2017.0*** (394.4)	-11715.4*** (1413.1)
y_{it}	0.004*** (0.001)	0.003** (0.001)	0.002* (0.001)	0.001 (0.001)
a_{it}	12990.8*** (753.3)	11156.4*** (1000.2)	14107.2*** (800.3)	4754.4*** (1310.7)
d_{it}	-19533.8*** (4277.2)	55016.8 (35176.0)	-23064.1*** (5675.0)	182379.6*** (38779.9)
u_{it}	652.9 (410.6)	1684.8*** (489.2)	-925.1** (441.9)	1597.2*** (462.4)
r_t	758.9* (457.8)	795.5 (515.6)	495.4 (485.2)	635.7 (560.0)
g_t^e	7538.9*** (1038.6)	7663.1*** (1074.9)	4247.9*** (1206.0)	6511.0*** (1204.8)
$constant$	-3328.3*** (294.1)	-1999.6*** (588.3)	-2899.2*** (304.6)	1200.1* (722.7)
R^2				
<i>within</i>	0.55	0.57	0.50	0.56
<i>between</i>	0.21	0.00	0.34	0.02
<i>overall</i>	0.24	0.00	0.35	0.01
Obs.	2814	2814	2814	2814

Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.

Table 7

Estimation of (14)		
2004 – 2012, 402 districts		
Variable	Apartments	Single-family houses
	HT/IV	HT/IV
s_{it}	-3324.3*** (284.4)	-5583.2*** (1159.5)
y_{it}	0.001* (0.001)	-
a_{it}	9105.0*** (344.5)	7524.3*** (570.3)
d_{it}	-106885.5*** (8916.6)	-115774.0** (45334.5)
u_{it}	-	-
r_t	-	-
g_t^e	9375.5*** (1452.6)	6811.5*** (1393.0)
<i>constant</i>	-513.4*** (211.9)	758.4*** (266.3)
R^2		
<i>within</i>	0.46	0.50
<i>between</i>	0.11	0.21
<i>overall</i>	0.12	0.21
Obs.	3618	3618
Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.		

Table 8

Estimation of (14)		
1996 – 2012, 93 towns cities		
Variable	Apartments	Single-family houses
	HT/IV	HT/IV
s_{it}	-4346.2*** (356.7)	–
y_{it}	–	–
a_{it}	6996.0*** (911.1)	3053.1*** (555.4)
d_{it}	–	-446640.8*** (33892.5)
u_{it}	-5808.5*** (1334.0)	–
r_t	–	–
g_t^e	40470.1*** (1493.8)	36436.9*** (1274.9)
<i>constant</i>	523.1* (316.7)	923.7*** (217.4)
R^2		
<i>within</i>	0.63	0.53
<i>between</i>	0.09	0.15
<i>overall</i>	0.22	0.15
Obs.	1581	1581

Robust standard errors in parentheses. *, **, *** denotes significance on the 10%, 5%, 1%-level.

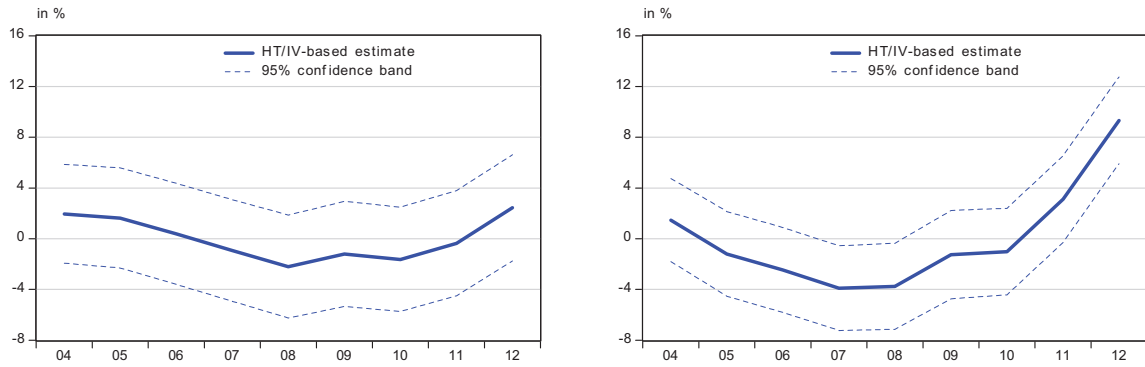


Figure 16: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 93 cities, estimated using data for 402 districts over 2004 to 2012.

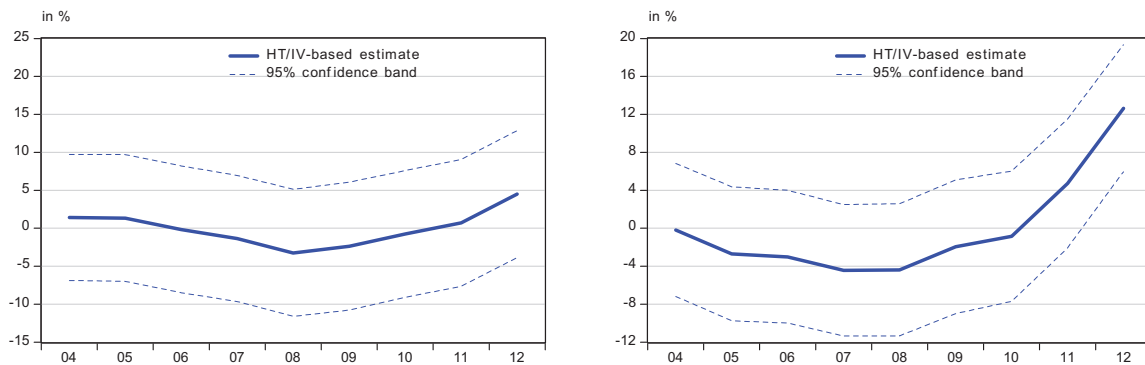


Figure 17: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for major 7 cities, estimated using data for 402 districts over 2004 to 2012.

found to be significantly overvalued. Similarly, for the subset of cities the size of the overvaluation is slightly smaller, but still significant. For the longer sample of 93 cities, the results hold up both qualitatively and quantitatively (see Figs. 18 and 19).

8.C Logs vs. levels: Assessing the aggregation bias

This section describes the simulation exercise used to evaluate the aggregation error from estimating a linear model instead of a log-linear model. Recall that on the basis of the theoretical model we assume that a house price equation in logarithms is the true data generating process. The error is compared to the one that arises from an alternative, economically sensible – although only approximative – aggregation rule based on the logarithmic model. For the purposes of the simulation study we focus on a random-effects

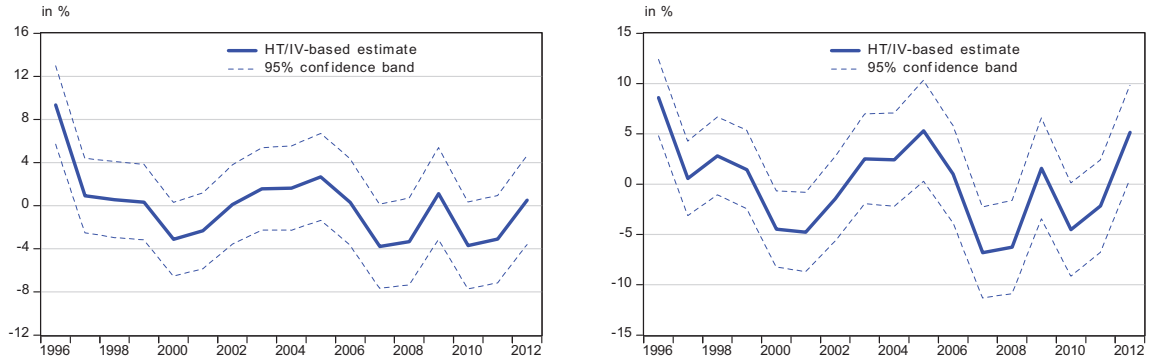


Figure 18: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for 93 cities estimated using data for 93 cities from 1996 to 2012 .

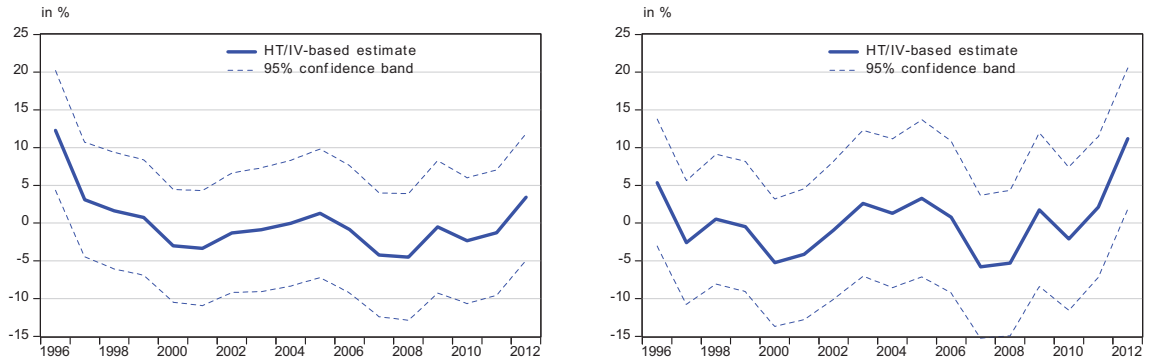


Figure 19: Deviations of single-family house (left) and apartment prices (right) from long-run equilibrium for major 7 German cities estimated using data for 93 cities from 1996 to 2012 .

model for apartment prices of the following form.

$$\ln p_{it}^1 = \phi_0 + \phi_1 \ln s_{it} + \phi_2 \ln y_{it} + \phi_3 \ln a_{it} + \phi_4 \ln d_{it} + \phi_5 \ln g_t + c_i + \epsilon_{it} \quad (25)$$

This corresponds to the chosen final specification for apartment prices using the random-effects estimator in section 5. We proceed as follows.

1. Generate data for n_{it} , s_{it} , y_{it} , a_{it} , d_{it} and g_t using the variance-covariance matrix from their empirical counterparts.
2. Generate $\ln p_{it}^1$ for $i = 1, \dots, N$ and $t = 1, \dots, T$ using (25), coefficient values from the estimated apartments equation (RE) in Table 2, $\epsilon_{it} \sim N(0, \sigma_\epsilon)$ and $c_i \sim N(0, \sigma_c)$.

Compute $P_t^1 = \sum_{i=1}^N \frac{n_{it}}{n_t} p_{it}^1$.

3. Estimate the log-linear model (25) by random-effects. Compute $\hat{P}_t^2 = e^{\ln \hat{P}_t^2}$ from

$$\begin{aligned} \ln \hat{P}_t^2 &= \hat{\phi}_0 + \hat{\phi}_1 \ln \left(\sum_{i=1}^N \frac{n_{it}}{n_t} s_{it} \right) + \hat{\phi}_2 \ln \left(\sum_{i=1}^N \frac{n_{it}}{n_t} y_{it} \right) \\ &+ \hat{\phi}_3 \ln \left(\sum_{i=1}^N \frac{n_{it}}{n_t} a_{it} \right) + \hat{\phi}_4 \ln \left(\sum_{i=1}^N \frac{n_{it}}{n_t} d_{it} \right) \\ &+ \hat{\phi}_5 \ln g_t + \hat{c} \end{aligned} \quad (26)$$

where \hat{c} are the aggregated unobserved effects (weighted by population shares).

4. Estimate the linear model (27) by random-effects

$$p_{it}^3 = \varphi_0 + \varphi_1 s_{it} + \varphi_2 y_{it} + \varphi_3 a_{it} + \varphi_4 d_{it} + \varphi_5 g_t + c_i + \zeta_{it} \quad (27)$$

and compute $\hat{P}_t^3 = \sum_{i=1}^N \frac{n_{it}}{n_t} \hat{p}_{it}^3$.

5. Compute the mean errors $ME^1 = \frac{1}{T} \sum_{t=1}^T (\hat{P}_t^2 - P_t^1)$ and $ME^2 = \frac{1}{T} \sum_{t=1}^T (\hat{P}_t^3 - P_t^1)$.

6. Repeat steps 2 to 5 R times to obtain the means and variances of ME^1 and ME^2 .

We set $T = 10$, $R = 500$ and run the exercise for different values of $N = \{50, 100, 150, 200, \dots, 1000\}$. Fig. 20 plots the means of ME^1 and ME^2 for increasing N . The mean error from estimating and a linear specification and using an exact aggregation rule is about 10 times lower than the one from estimating the true logarithmic specification while using an approximate aggregation rule. In absolute terms the mean error from the linear model, which is mainly due to misspecification, indicates that the aggregate linear model results in an overestimation of apartment prices of around 1%. In contrast, the bias from the aggregate logarithmic model is about 10%.

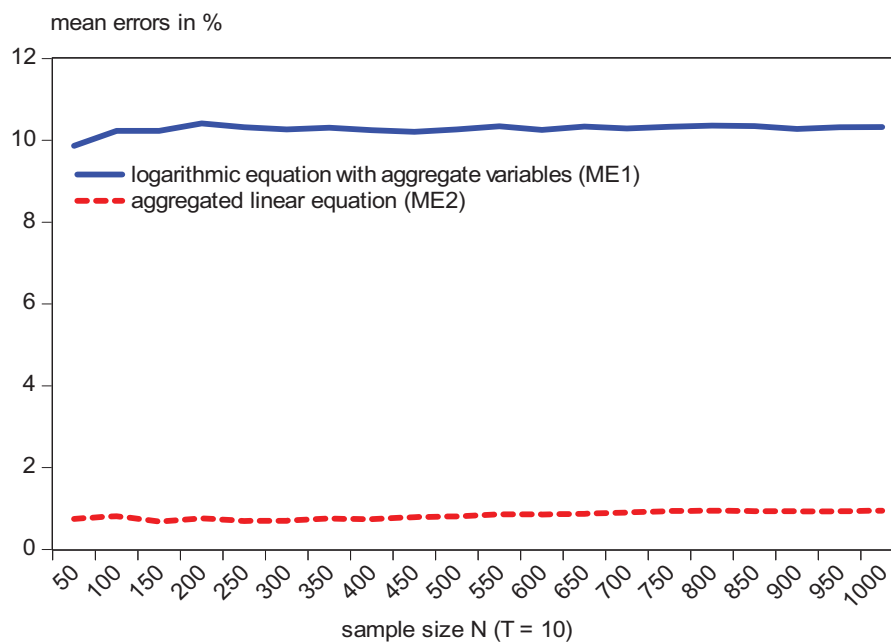


Figure 20: Mean errors from estimating the logarithmic equation and using an approximate aggregating rule vs. estimating a linear equation with consistent aggregation (when the logarithmic equation is true).