DIW BERLIN

1168

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2011

Can Internet Ads Serve as an Indicator of Homeownership Rates?

Konstantin A. Kholodilin and Andreas Mense

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Can Internet ads serve as an indicator of homeownership rates?[¶]

Konstantin A. Kholodilin[*] Andreas Mense[§]

October 31, 2011

Abstract

In this paper, we propose an indicator of the homeownership rate based on Internet ads offering the housing for rent and sale. We constructed the HOR estimate using the number of ads in four different markets (flats for rent, flats for sale, houses for rent, and houses for sale). Our HOR indicator was tested using data of German NUTS1 and planning (ROR) regions. The correlation between our estimate of the HOR and the alternative HOR figures varies between 0.834 and 0.874 at NUTS1 level and is 0.761 at the ROR level. All correlation coefficients are statistically significant. Our HOR estimate is particularly highly correlated with the official HOR figures. Thus, it is shown that our Internet-based indices could serve as a good indicator of the homeownership rate in German regions.

Keywords: Internet ads; homeownership rate; German regions; NUTS1; planning regions.

JEL classification: C21; O47; R11.

[*]DIW Berlin, Mohrenstraße 58, 10117 Berlin, Germany, e-mail: kkholodilin@diw.de

[§]Friedrich-Alexander-Universität Erlangen-Nürnberg, Kochstraße 4, 91054 Erlangen, Germany, e-mail: Andreas.Mense@wiwi.phil.uni-erlangen.de

[¶]The authors acknowledge the very helpful comments of Oliver Lerbs from the Institut für Siedlungs- und Wohnungswesen (Münster, Germany).

Contents

List of Tables

List of Figures

In unserem Koalitionsvertrag haben wir die Bedeutung des Wohneigentums unterstrichen. Unser Ziel ist es, die Wohneigentumsquote zu erhöhen.

Peter Ramsauer,
German Federal Minister of Transport, Building, and Urban Development

1 Introduction

Internet provides a lot of information that can be easily and quickly accessed. In particular, it facilitates real-time monitoring of tendencies in the housing market both at the national and regional levels.

In this paper, we investigate the extent to what the Internet ads on housing offered for rent or for sale can approximate the homeownership rate (HOR). In Germany, the homeownership rate is defined as the ratio of the households living in their own homes to the total number of households. The homeownership rate shows the ratio of the owner-occupied housing units to the total stock of occupied housing units or households.

This indicator plays an important role in political discussions. In some countries, e.g., in the USA, it "is often seen as an integral part of the American dream, and encouraging homeownership has historically been an important feature of U.S. public policy", see Haughwout et al. (2010). Glaeser (2011) points to federal policies such as home mortgage deductions as well as local regulations that prohibit tall buildings. The latter boosts the HOR because rental units are much more likely to be in multifamily structures than in single houses. The recent decline of the US homeownership rate to its 1998 level has caused lots of worries in the economic press. In the fourth quarter 2010, it decreased to 66.5% compared to its peak value of almost 70% that was attained in 2004. At this background, the German HOR comprising only 42% is considered by some as a sign of weakness of German housing market, see, for example, Proxenos (2002) and Voigtländer (2006). The current coalition government in Germany has even included in its program the objective

of raising the HOR. It is thought that the growing number of homeowners brings more stability and order for the society.

Whether or not the high HOR *per se* is a desirable goal, is subject to discussion. In any case, it is a key characteristic of the housing market and social conditions of the population. Moreover, it can be an important factor of the price setting, given that the lower HOR, like those observed in Germany and Switzerland, point out to a stronger market power of the tenants, which, in turn, allows explaining at least to some extent the relatively slow growth of house prices in these two countries[1].

The determinants and role of the HOR have been thoroughly examined in numerous publications. For an overview of the HOR studies at the international level see Gwin and Ong (2004) and Forrest (2004), among others. Voigtländer (2006) concentrates upon the German case and investigates both the reasons for a relatively low HOR in Germany compared to other European countries and its socioeconomic importance. Lerbs and Oberst (2011) analyze the spatial variation of the homeownership rates across German planning regions. Hirayama (2010) points to the development of home ownership in Japan where a long period of slow economic growth has made it difficult for younger generations to buy their home. This has also implications for the structure of the welfare system, for which home ownership served as an important pillar. For Canada, Brown et al. (2010) find home ownership to be an important contribution to the implicit household income. This is especially important since explicit household income declines after retirement.

In order to be able to analyze the economic relevance of the HORs, one needs reliable indicators of the homeownership. The situation with data on HORs is very different even across industrialized countries. In the USA, the Census Bureau starting from 1965 releases homeownership rates for the whole country and even the states on quarterly basis. In Germany, the availability of data on homeownership rates is much worse than in the USA. There exist basically two different indicators of the HOR for German regions. One indicator is provided by the German Federal Statistical Office (Destatis). It is based on a supplementary survey of the microcensus, which covers 1% of all German households, i.e., about 400,000 households, and is conducted every 4 years. Another indicator of HOR is computed by the research institute

[1]On the peculiarities of the house price dynamics in Germany and Switzerland see Kholodilin et al. (2010). For a discussion of the risks and returns of home ownership, see Kramer (2010).

empirica. This indicator is based on data from the Income and Consumption Sample, or Einkommens- und Verbrauchsstichprobe (EVS), which covers 60,000 households and is conducted every 5 years. In addition, empirica computes alternative HORs using regional data from the Socio-Economic Panel (SOEP), which starting from 1985 collects the responses of 11,000 households on an annual basis. Thus, for Germany the most representative data that can be used for construction of HORs, are available only every 4 years. The SOEP data are collected on a yearly basis but are not representative enough at the regional level: only for the 3 largest NUTS1 regions, or federal states — Nordrhein-Westfalen, Bayern, and Baden-Württemberg — the number of observations is large enough, see Knies and Spiess (2007). This stands in a striking contrast to the data situation for the USA.

Therefore, given the relevance of the HOR measure, the objective of this paper is to fill this gap and produce the reliable estimates of homeownership rates for German regions, which could be computed and reported at reasonable frequency.

The paper is structured as follows: Section 2 describes the data used in the study. In section 3, the necessary data transformations are described and the methodology of computing the HOR is explained. Section 4 discusses results. Finally, section 5 concludes.

2 Data

This study uses data contained in Internet ads on flats and houses offered for rent and for sale in Germany. The data were downloaded in August 2011 from three popular German websites — Immobilienscout24.de, Immonet.de, and Immowelt.de, where housing ads are published.

There are, of course, other sites, where such ads are placed. However, due to their large market shares, these three sites are representative to a high degree. For example, in July 2011 the number of ads offering flats and houses for rent and sale in Immobilienscout24.de (360,900), Immowelt.de (307,700), and Immonet.de (251,900) was 920,500 in total. Their closest competitors are Ebay.de (218,900), Quoka.de (125,500), Immobilien.de (79,800), and Kalaydo.de (53,300). Given these figures, the three websites have a combined market share of approximately 66%.

Table 1 reports the correlation between population in 2009 by German NUTS1 regions and the total number of ads. The representativeness of the

data seems to be lower for one single website and market segment as compared to the correlation obtained when the three websites are taken as a whole, where the correlation exceeds 0.91 for three of the four segments. This also shows the importance of using more than one website as the data source. The total correlation of ads per region in all market segments and the population is as high as 0.96. The low figures for rental flats can be explained by the overproportionate number of ads placed in Sachsen. While only 5% of the total population of Germany live in Sachsen, its Immonet share exeeds 21% for rental flats. To a lesser extent, the same is true for Immobilienscout24 (14%), and Immowelt (9%).

Given their size and representativeness, we decided to use data from the three sources mentioned above: Immobilienscout24, Immowelt and Immonet. The fact that there might be overlapping ads in different websites diminishes the marginal benefit of additional (and considerably smaller) websites. Three different sources thus seem to be a reasonable choice. Notice also that the number of ads placed on the three websites is much greater than the size of the microcensus sample covering 400,000 respondents, even if duplicate ads are removed.

The ads for all types of flats were downloaded. In case of houses, only the ads for the following types of houses were downloaded: farmhouse (Bauerhaus), semi-detached house (Doppelhaushälfte), one-family detached house (Einfamilienhaus), villa (Villa), townhouse (Reihenhaus), bungalow (Bungalow), since each of these houses is usually inhabited by one household. The ads for houses of types apartment building (Mehrfamilienhaus), special housing (Besondere Immobilie), and other houses (Sonstige) were not downloaded. The house of the first type can contain multiple flats, but the number of flats is not reported in the ads. The houses of the latter two types are difficult to identify, moreover, their share is negligible.

The ads published on the three websites contain numerous characteristics of the housing property, which are listed in Table 2.

Since it is very likely that some ads are published on different websites simultaneously, these duplicates may lead to serious distortions of the results. Therefore, we split the data into ZIP code subgroups and scanned each of these groups separately for exact matches. To do the matching, we used a subset of "critical" variables in order to avoid spurious mis-matches due to data incompleteness. The subset contains only variables that are most likely to be reported by the publisher of the ad and that are best suited to distinguish between different flats or houses. Two single ads were identified

as duplicates if at least all but one of the critical variables matched. Numeric variables were allowed to deviate by up to 1% to catch inaccuracies in the ads. The list of critical variables is reported in Table 3.

The ZIP codes contained in the ads were used to find the geographical coordinates (latitude and longitude) of each flat or house. Then, the ads were assigned to the respective NUTS1 regions, given the information on their borders. The shapefile containing the geographical information on the regional borders was taken from the Eurostat.

3 Data transformations and computation of HOR

In Germany, the homeownership rate is defined as the ratio of the households living in their own homes to the total number of households. Our measure thus has to approximate this definition as close as possible using the Internet housing ads.

There exist basically four (sub)markets of housing: 1) flats for rent, 2) flats for sale, 3) houses for rent, and 4) houses for sale. In addition, as can be seen in columns 2 and 4 of Table 2, some of the flats and houses offered for sale may be already occupied by tenant (rented out). Hence, the ads used to approximate the stock of the tenant-occupied housing should comprise the ads of flats and houses offered for rent as well as the ads of the rented out flats and houses offered for sale. The estimation of the stock of the owner-occupied housing should be based only on the number of ads offering for sale the tenant-free flats and houses.

The "raw" number of ads, however, can hardly be used to compute the HOR. The reason is that many of the objects, especially houses, offered in the ads are not constructed yet and those ads are placed by the construction firms in order to attract new customers. Hence, a substantial part of these houses exist only on paper and may never be built. Not accounting for this would lead us to biased results, since non-existing flats or houses are virtually never offered for rent. Therefore, the data must be "cleaned" before we can compute the homeownership rates. We filtered the data by taking advantage of the information contained in the ads. Theoretically, this could be done by eliminating the ads that have the current or future year as a construction year or have the value "new" of variable "Condition" and by keeping only

those ads of houses, which have in the category "Phase of construction" the value "constructed".

However, the data are very incomplete. For instance, most of ads have missing values in the variable "Year of construction". The "Phase of construction" variable is reported in Immoscout24 but not in Immonet and Immowelt[2]. In addition, only about 25% of the Immoscout24 ads contain that information. The "Year of construction" variable also has a significant share of missing values (Immoscout24: 24%, Immonet: 28%, Immowelt: 34% for the "houses for sale" market). The "Condition" of the object is missing for 33% (Immoscout24) to 58% (Immowelt) of the ads.

Therefore, some other device must be used in order to discriminate between the existing and non-existing housing units. We take advantage of the fact that, although some values of the relevant variables might be missing in the ad, the short textual description of the real estate is almost always present. Using text processing techniques, we are able to identify ads of non-existing housing units, which can have some common features like the length of text and frequency of keywords.

First, we took two samples of ads concerning the flats and houses for sale using only the ads, which we can unequivocally identify as announcing the existing and non-existing homes. Each of the samples comprised of more than 2000 ads. Second, we constructed a binary dependent variable. It is equal to 1 if the advertized housing unit does not exist yet (construction year is 2011 for flats, 2010 and 2011 for houses, construction phase is "planned" or "under construction" both for flats and houses) and to 0, otherwise. Third, the texts were purged from numbers, non-textual signs (asterisks, stars, and so on), and the non-informative words (articles, prepositions, and so forth). Fourth, the independent variables were constructed on the basis of the resulting text. The first variable is the length of ads measured by the number of characters, whereas the remaining 19 variables are the number of occurrences of the keywords per each ad, that appear to be significant. Fifth, each sample was subdivided into roughly equal estimation and forecasting subsamples. Sixth, a logit model was estimated using the estimation subsample data. For an example of estimation output see Table 4. Seventh, using the estimated regression coefficients the probabilities of advertizing non-existing flats and

[2]Other variables that are missing in one or two sites are "Usable area", "Number of floors" for Immowelt, and "Rental income", "Year of last renovation" for Immowelt and Immonet.

houses were computed. These were compared to the known binary type of each ad and the in- and out-of-sample forecast accuracy measures were computed. As a measure of forecast accuracy the quadratic probability score (QPS) was used:

$$QPS = \frac{1}{T}\sum_{t=1}^{T}(R_t - P_t^j)^2 \tag{1}$$

where R_t is the known binary type of the housing unit; P_t is the model-derived probabilities of non-existing housing units bubbles. QPS varies between 0 and 1. The lower the QPS the more precise are the predictions of the non-existing flats and houses.

Figure 1 depicts the in- and out-of-sample forecasting performance of the logit model based on the ads of houses for sale. The ads were arranged by type; existing and non-existing. Hence, the shaded area corresponds to non-existing houses (binary variable equal to 1). The dashed line represents the average probability by type and the solid lines are the probabilities derived from the logit model. The upper panel shows the in-sample predictive accuracy, that is, the accuracy achieved in the estimation subsample, whereas the lower panel shows the out-of-sample accuracy. It can be seen that the logit model matches the type of housing unit (existing vs. non-existing) quite precisely[3].

In order to exactly identify the type of ad the logit probabilities must be converted into a sequence of 0 and 1. This means that a threshold must be chosen allowing to tell apart the low and high probabilities of advertizing a non-existing housing unit. Several thresholds from 0.1 to 0.9 with a step of 0.1 were tried. It turns out that the threshold 0.5 gives the best results.

We did all the above estimations using the Immobilienscout24 data. However, for the sake of robustness we computed the out-of-sample forecasting performance also for the samples drawn from the Immonet and Immowelt databases. These exercises show that the logit model we have estimated works well also for the ads published by these two providers. We calculated the QPS for those ads from Immonet and Immowelt that had the "year of construction" variable reported, whereby a year of construction of 2011 or 2012 was treated as 1 and all other years as 0. For Immonet, QPS scores were as low as 0.10 (flats for sale) and 0.09 (houses for sale), Immowelt had scores

[3]The forecasting performance for the flat-for-sale ads is somewhat worse but still good enough: in-sample QPS=0.124, and out-of-sample QPS=0.174.

of 0.11 for both market segments. Thus, we can be sure that our method allows us successfully classifying the ads into those announcing the existing and those announcing the non-existing housing units.

In addition, all the ads of flats and houses having the state "new" and "to be occupied for the first time" were excluded. Likewise the ads of flats for sale having "ready to be demolished" as the state were eliminated. Finally, all the outlier ads according to the size of housing and that of the land lot for houses were dropped. An observation is treated here as outlier, when it is greater than 1.5 times the interquartile range.

After all these transformations the homeownership rates can be computed. The homeownership rate for region i in time period t could be computed as follows:

$$HOR_{it} = \frac{N_{it}^{FS.nrented} + N_{it}^{HS.nrented}}{N_{it}^{FR} + N_{it}^{FS.rented} + N_{it}^{HR} + N_{it}^{HS.rented}} \tag{2}$$

where N_{it}^{FR} is the number of ads announcing flats for rent; $N_{it}^{FS.rented}$ is the number of ads announcing flats for sale, which are currently occupied by tenant; $N_{it}^{FS.nrented}$ is the number of ads announcing flats for sale, which are not occupied by tenant; N_{it}^{HR} is the number of ads announcing houses for rent; $N_{it}^{HS.rented}$ is the number of ads announcing houses for sale, which are currently occupied by tenant; $N_{it}^{HS.rented}$ is the number of ads announcing houses for sale, which are not occupied by tenant.

4 Results

The HORs by German NUTS1 regions, which were computed using the methodology described in the previous section, are reported in Table 5. Our homeownership estimates, which are shown in column 10 of the table, were compared to the three alternative estimates contained in columns 11-13. The first alternative measure is that of German Federal Statistical Office (Destatis). We refer to it as the official measure. The Destatis HOR measure is based on microcensus data. The most recent official estimates refer to 2006 and were taken from Statistisches Bundesamt (2010). The second alternative HOR estimate was calculated by the research institute empirica. It uses data from the EVS survey carried out in 2008 and is calculated based on the households. The third alternative HOR measure was also computed by empirica using the same data but persons instead of households. Using

the persons instead of households as calculation unit inflates HOR measures, provided that larger families tend to live in their own housing, whereas singles and couples without children, as a rule, live in the rental housing. The methodologies and data used to compute the alternative HOR measures are described in detail in Braun and Pfeiffer (2004).

Table 6 reports our estimate and official indicator of the homeownership rates for German Raumordnungsregionen (ROR), or planning regions. There is only one official measure of HOR at such disaggregation level, which is produced by German Federal Statistical Office for 2006 and based on the housing data collected through an "Additional census on the living situation of households" conducted in 2006 within the microcensus framework (Zusatzerhebung Wohnen aus dem Mikrozensus). It is worth noticing that this census is carried out only once every 4 years. Since then the ROR division has changed and 2 new planning regions were created: Anhalt-Bitterfeld-Wittenberg and Arnsberg. This was possibly done through rearranging the borders of existing planning regions. Unfortunately, we cannot properly reflect these changes and simply compare the current RORs with 2006 regions bearing the same names. In our sample, there are 96 planning regions. However, due to the lack of observations no HORs could be computed for Altmark and Nordthueringen.

The correlations between our HOR estimate and alternative HOR measures both at NUTS1 and ROR levels are reported in Table 7. The correlation between our HOR indicator and the alternative ones is examined using three correlation coefficients: Pearson's, Spearman's, and Kendall's coefficients. The latter two tests are nonparametric and thus free of any assumptions about the distribution of the correlations in question. As one can see, the correlation coefficients computed for 16 German NUTS1 regions are almost always higher than 0.8. The Spearman coefficients are the highest, achieving approximately 0.9. Our HOR estimate is particularly highly correlated with the official HOR figures. All these coefficients are statistically significant. The correlation computed for 92 RORs is significant, but lower in magnitude. This may have to do with smaller amount of observations per planning region and border changes that have taken place since 2006. It is worth noticing that the Pearson's correlations between the alternative HOR measures are all higher than 0.9.

Figure 2 compares our estimate of the HOR to the official one. It can be seen that, despite high concordance between these two measures, in all but one (Sachsen) cases our estimate lie above the 45^0 dashed line. The official

HOR figure is 31%, whereas our estimate is about 23%. This can be related to the fact that Sachsen is producing overproportionately large number of flat rental ads: while for the whole country 1.8 ads per 1000 households[4] are placed monthly on the three largest sites under inspection, in Sachsen the figure is 3.8. This possibly reflects the excess supply of flats for rent in this region, especially in such cities as Dresden and Leipzig.

Similar picture can be observed in Figure 3, which compares our estimate and the official indicator of HOR. Given the large cross-section size we do not display the names of the ROR regions, but their codes. The correspondence between the numerical codes and the region names can be found in Table 6. Again, as in the case of NUTS1 regions, our HOR estimates are systematically higher than the official ones, since most of the points lie above the 45^0 line. One possible explanation is that our filtering technique, which eliminates the non-existing homes, is too liberal and does not exclude all such cases. We tried different thresholds lower than 0.5 applying them to the logit probabilities, but this does not lead to a noticeable improvement of results in terms of both higher correlation between our and official HOR measures and decreased level of our HOR measure. Probably, a more important explanation is that the homes for sale remain longer time in the internet, since making a decision on purchasing home takes more time than that of renting it. In the former case, the transaction costs are substantially higher. For example, in addition to dealer's fee, in case of purchasing real estate in Germany one must pay an immobile property transfer tax, which in Germany amounts, as a rule, to 3.5% of the property value. It should be also noticed that the ads represent the flows, whereas the homeownership is a stock variable. Therefore, there are fewer (more) ads announcing homes for rent (sale) than implied by the actual stock, given their higher (lower) "velocity of circulation". We acknowledge the problem but, unfortunately, until now were not able to solve it properly. A third explanation, deals with the fact that our estimate refer to 2011, whereas the official ones to 2008 and even 2006. We can argue that during these 3 years the HOR in Germany might have increased. The upward trend in the German HOR during the recent decades is mentioned, for instance, in Voigtländer (2006).

A deeper insight into the distribution properties of our HOR estimate compared to the alternative HOR measures can be gained from the boxplots

[4]The number of ads refers to the August 2011, whereas the most recent data on the number of households published by Destatis refer to 2010.

displayed in Figure 4. As seen, our estimates both at NUTS1 and ROR levels are higher and somewhat more dispersed than the alternative ones. Our HOR estimates are closer to those constructed by empirica, which are based on persons.

5 Conclusion

In this paper, we proposed an indicator of the homeownership rate based on Internet ads that are announcing homes for rent and sale. We constructed the HOR estimate using the number of ads in four different markets (flats for rent, flats for sale, houses for rent, and houses for sale). The raw data were processed to make sure that they contain no homes, which do not exist yet, and no outliers.

Our HOR indicator was tested using data on 16 NUTS1 regions and 92 planning regions of Germany. The correlation between our estimate of the HOR and the alternative HOR figures varies between 0.7 and 0.9 at NUTS1 level and between 0.5 and 0.8 at ROR level and is always statistically significant. Our HOR estimate is particularly highly correlated with the official HOR figures.

The great advantage of our HOR estimate is that it can be computed relatively inexpensive and quickly, unlike the alternative measures that are based on either the microcensus conducted once a year or Income and Consumption Sample collected once every 5 years. Thus, our Internet-based HOR measure can serve as a good indicator of the homeownership rate in German regions. Our HOR is readily available and allows closely monitoring the dynamics of the homeownership in time.

References

Braun, R. (2009a). Vermögensbildung in Deutschland. Teil 2: Haushaltsbezogene Wohneigentumsquoten. empirica.

Braun, R. (2009b). Vermögensbildung in Deutschland. Teil 3: Personenbezogene Wohneigentumsquoten, Gebäudeart und Zweitwohnungen. empirica.

Braun, R. and U. Pfeiffer (2004). Haushalts- und personenbezogene Wohneigentumsquoten in Deutschland. empirica.

Brown, W. M., F. Hou, and A. Lafrance (2010). Income of Retirement-age and Working-age Canadians: Accounting for Home Ownership. Economic Analysis Research Paper Series.

Forrest, R. (2004). Economic downturns and the prospects for home ownership. University of Bristol.

Glaeser, E. L. (2011). Rethinking the Federal Bias Toward Homeownership. *Cityscape: A Journal of Policy Development and Research 13*, 5–37.

Gwin, C. R. and S.-E. Ong (2004). Do we really understand homeownership rates? An international study. Baylor University Working Paper Series 2004-053-ECO.

Haughwout, A., R. Peach, and J. Tracy (2010). The homeownership gap. *Current Issues in Economics and Finance 16*(5), 1–11. Federal Reserve Bank of New York.

Hirayama, Y. (2010). The role of home ownership in Japan's aged society. *Journal of Housing and the Built Environment 25*, 175–191.

Kholodilin, K. A., J.-O. Menz, and B. Siliverstovs (2010). What drives housing prices down? Evidence from an international panel. *Jahrbücher für Nationalökonomie und Statistik 230*(1), 59–76.

Knies, G. and K. Spiess (2007). Regional data in the German Socio-Economic Panel Study (SOEP). DIW Berlin.

Kramer, B. (2010). The risk and return of home ownership. OFRC Working Paper Series. Applied Paper No. 2010-02.

Lerbs, O. and C. Oberst (2011). Explaining spatial variation in homeownership rates: Results for german regions. CESifo Working Paper Series 3377, CESifo Group Munich.

Proxenos, S. (2002). Homeownership rates: A global perspective. *Housing Finance International*, 3–7.

Statistisches Bundesamt (2010). Statistisches Jahrbuch 2010 für die Bundesrepublik Deutschland mit "Internationalen Übersichten".

Voigtländer, M. (2006). Mietwohnungsmarkt und Wohneigentum: Zwei Seiten einer Medaille. Institut der deutschen Wirtschaft Köln.

Appendix

Table 1: Comparison of the three databases: Immobilienscout24, Immonet, and Immowelt. Correlation with population of NUTS1 regions, July 2011

	houses		flats	
	for sale	for rent	for sale	for rent
	(1)	(2)	(3)	(4)
Immonet	0.824	0.913	0.940	0.643
Immowelt	0.407	0.914	0.726	0.692
Immoscout24	0.946	0.948	0.933	0.855
Cumulated	0.910	0.950	0.938	0.811
Whole market[1]	0.959			

[1] Population to total number of ads, all market segments and websites.

Table 2: List of variables from housing ads

Flats for rent	**Flats for sale**	**House for rent**	**House for sale**
ID	ID	ID	ID
Bundesland	Bundesland	Bundesland	Bundesland
City	City	City	City
District	District	District	District
Address	Address	Address	Address
ZIP code	ZIP code	ZIP code	ZIP code
Location	Location	Location	Location
Area	Area	Area	Area
Usable area	Usable area	Usable area	Usable area
		Land lot size	Land lot size
Cold rent		Cold rent	
Warm rent		Warm rent	
Rent for parking box		Rent for parking box	
Additional costs		Additional costs	
Heating costs		Heating costs	
	Purchase price		Purchase price
	Price of parking box		
Type of flat	Type of flat		
		Type of house	Type of house
Floor	Floor		
Number of floors	Number of floors	Number of floors	Number of floors
Number of rooms	Number of rooms	Number of rooms	Number of rooms
Number of bedrooms	Number of bedrooms	Number of bedrooms	Number of bedrooms
Number of bathrooms	Number of bathrooms	Number of bathrooms	Number of bathrooms
Cellar	Cellar	Cellar	Cellar
Guest WC	Guest WC	Guest WC	Guest WC
Access to garden	Access to garden	Access to garden	Access to garden
Balcony or terrace	Balcony or terrace	Balcony or terrace	Balcony or terrace
Parking box available	Parking box available	Parking box available	Parking box available
Number of parking boxes	Number of parking boxes	Number of parking boxes	Number of parking boxes
Fitted kitchen	Fitted kitchen	Fitted kitchen	Fitted kitchen
Elevator	Elevator	Elevator	Elevator
Accessible for handicapped	Accessible for handicapped	Accessible for handicapped	Accessible for handicapped
Suited for senior	Suited for senior	Suited for senior	Suited for senior
Pets allowed	Pets allowed	Pets allowed	Pets allowed
Year of construction	Year of construction	Year of construction	Year of construction
Year of last renovation	Year of last renovation	Year of last renovation	Year of last renovation
	When vacant		When vacant
Condition	Condition	Condition	Condition
	Monument		Monument
			Phase of construction
Equipment	Equipment	Equipment	Equipment
Heating	Heating	Heating	Heating
Type of lighting	Type of lighting	Type of lighting	Type of lighting
Energy efficiency	Energy efficiency	Energy efficiency	Energy efficiency
Social lodging			
Rent deposit		Rent deposit	
	Condo fee		
	Rented out		Rented out
	Rental income		Rental income
Broker's commission	Broker's commission	Broker's commission	Broker's commission

Table 3: Critical variables for the matching routine

Houses		Flats	
for sale	for rent	for sale	for rent
area	area	area	area
	cold rent		cold rent
	warm rent		warm rent
	additional costs		additional costs
purchase price		purchase price	
		floor	floor
		type of flat	type of flat
type of house	type of house		
number of rooms	number of rooms	number of rooms	number of rooms
number of parking boxes	number of parking boxes	number of parking boxes	number of parking boxes
year of construction	year of construction	year of construction	year of construction

Table 4: The estimation output of logit model estimated on the house-for-sale ads data

Coefficient	Estimate	Std. Error	p-value
Intercept	-1.611	0.234	0.000
Length of ad	-0.002	0.000	0.001
enthalten	1.311	0.280	0.000
fussbodenheizung	0.059	0.217	0.785
garten	-0.411	0.106	0.000
dokumente	0.968	0.304	0.001
enev	3.527	0.710	0.000
rolllaeden	1.026	0.306	0.001
zuhause	0.670	0.338	0.047
kfw	2.296	0.465	0.000
garage	-0.491	0.172	0.004
wuensche	0.734	0.278	0.008
schluesselfertig	3.419	0.683	0.000
bau	0.172	0.059	0.003
kosten	0.355	0.187	0.058
t	0.032	0.014	0.021
badewanne	0.555	0.336	0.099
ihrem	1.080	0.378	0.004
gebaut	-0.283	0.181	0.118
maler	0.602	0.348	0.084
vereinbaren	0.737	0.348	0.034
fusslaeufig	0.474	0.280	0.090
entstehen	2.584	0.563	0.000
individuell	0.799	0.272	0.003
kinder	0.315	0.178	0.077
planung	0.974	0.287	0.001

Table 5: Alternative measures of homeownership rates by NUTS1 regions

Region	Number of internet ads published on 3 largest sites								Home ownership rate			
	Flats for rent	Flats for sale rented out	Flats for sale not rented out	Houses for rent	Houses for sale rented out	Houses for sale not rented out	Tenant occupied homes	Owner occupied homes	Authors' estimate, internet ads	Destatis 2006, households	empirica 2008, households	empirica 2008, persons
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Baden-Wuerttemberg	6,655	434	1,832	597	543	9,151	8,229	10,983	57.2	49.3	52.0	60.0
Bayern	5,876	397	1,447	996	539	7,456	7,808	8,903	53.3	48.9	46.0	55.0
Berlin	6,838	189	574	125	35	1,355	7,187	1,929	21.2	14.0	15.0	19.5
Brandenburg	1,770	69	100	211	117	2,718	2,167	2,818	56.5	39.8	41.0	50.0
Bremen	665	9	82	23	12	426	709	508	41.7	35.1	37.0	42.0
Hamburg	1,289	26	124	83	19	774	1,417	898	38.8	21.9	20.0	25.0
Hessen	5,007	324	1,193	522	445	7,026	6,298	8,219	56.6	44.7	47.5	55.5
Mecklenburg-Vorpommern	1,318	30	167	80	101	2,048	1,529	2,215	59.2	35.9	31.0	41.0
Niedersachsen	5,294	152	754	559	590	10,675	6,595	11,429	63.4	51.0	49.0	57.0
Nordrhein-Westfalen	20,623	539	2,160	969	1,126	16,420	23,257	18,580	44.4	39.0	44.0	52.0
Rheinland-Pfalz	2,463	108	430	331	360	7,371	3,262	7,801	70.5	55.7	52.0	61.0
Saarland	422	10	17	31	30	1,976	493	1,993	80.2	56.9	57.0	64.0
Sachsen	8,365	123	182	122	178	2,456	8,788	2,638	23.1	31.0	30.0	38.0
Sachsen-Anhalt	2,841	21	37	51	100	2,137	3,013	2,174	41.9	39.6	32.0	43.0
Schleswig-Holstein	3,500	68	351	239	267	6,086	4,074	6,437	61.2	49.4	51.0	60.5
Thueringen	1,177	25	44	58	65	1,389	1,325	1,433	52.0	41.8	42.0	51.0
Germany	74,103	2,524	9,494	4,997	4,527	79,464	86,151	88,958	52.0	42.6	42.0	53.0

[1] own calculations based on the Internet ads.
[2] Statistisches Bundesamt (2010), based on the microcensus data.
[3] Braun (2009a), based on the EVS data.
[4] Braun (2009b), based on the EVS data.

Table 6: Alternative measures of homeownership rates by ROR

Region	ROR code	Authors' estimate, internet ads	Destatis 2006, house-holds	Region	ROR code	Authors' estimate, internet ads	Destatis 2006, house-holds
Aachen	45	65.4	44.8	Muenster	35	67.8	48.7
Allgaeu	95	61.8	46.0	Neckar-Alb	75	60.4	55.2
Altmark	31	NA	46.4	Nordhessen	48	71.5	48.9
Anhalt-Bitterfeld-Wittenberg	NA	47.1	NA	Nordschwarzwald	71	58.6	52.2
Arnsberg	NA	63.3	NA	Nordthueringen	53	NA	50.0
Augsburg	88	57.9	49.7	Oberes Elbtal/Osterzgebirge	58	22.9	23.5
Bayerischer Untermain	80	67.9	48.6	Oberfranken-Ost	84	68.1	46.8
Berlin	30	21.2	13.6	Oberfranken-West	83	71.2	50.8
Bielefeld	36	60.1	46.3	Oberland	96	45.8	44.0
Bochum/Hagen	43	32.3	32.0	Oberlausitz-Niederschlesien	59	51.6	35.2
Bodensee-Oberschwaben	79	58.8	54.1	Oberpfalz-Nord	85	66.5	54.1
Bonn	46	51.8	45.7	Oderland-Spree	27	66.4	41.5
Braunschweig	22	61.1	42.2	Oldenburg	16	64.3	52.3
Bremen	11	44.6	35.4	Osnabrueck	18	52.3	50.6
Bremen-Umland	15	62.5	58.3	Ost-Friesland	12	74.9	55.0
Bremerhaven	13	60.2	49.2	Osthessen	50	66.3	52.9
Donau-Iller (BW)	74	54.1	52.6	Ostthueringen	56	41.4	35.6
Donau-Iller (BY)	94	65.0	59.6	Ostwuerttemberg	73	65.8	56.5
Donau-Wald	91	72.1	55.2	Paderborn	37	71.4	51.1
Dortmund	39	33.9	30.3	Prignitz-Oberhavel	25	64.1	43.3
Duesseldorf	42	36.4	33.8	Regensburg	90	62.9	49.6
Duisburg/Essen	41	37.4	32.0	Rhein-Main	64	47.5	48.6
Emscher-Lippe	40	27.9	29.8	Rheinhessen-Nahe	51	64.4	36.3
Emsland	17	83.2	63.5	Rheinpfalz	66	68.8	50.8
Franken	69	67.6	54.5	Saar	67	80.2	54.2
Goettingen	24	61.8	41.7	Schleswig-Holstein Mitte	3	45.9	40.5
Halle/S.	34	41.2	35.0	Schleswig-Holstein Nord	1	76.4	53.4
Hamburg	6	38.8	19.5	Schleswig-Holstein Ost	4	60.9	40.3
Hamburg-Umland-Sued	14	65.2	57.9	Schleswig-Holstein Sued	5	61.9	50.0
Hannover	19	52.3	36.6	Schleswig-Holstein Sued-West	2	69.9	51.0
Havelland-Flaeming	29	51.8	36.4	Schwarzwald-Baar-Heuberg	76	62.6	54.7
Hildesheim	23	59.4	44.4	Siegen	47	72.6	54.0
Hochrhein-Bodensee	78	64.5	47.1	Starkenburg	52	57.4	48.8
Industrieregion Mittelfranken	86	47.1	39.1	Stuttgart	72	49.9	46.2
Ingolstadt	89	52.6	56.7	Suedheide	20	69.0	57.0
Koeln	44	44.3	36.4	Suedlicher Oberrhein	77	61.3	44.7
Landshut	92	61.9	55.8	Suedostoberbayern	97	58.9	49.2
Lausitz-Spreewald	28	56.0	38.6	Suedsachsen	55	33.4	46.7
Lueneburg	21	63.4	50.8	Suedthueringen	61	55.5	35.0
Magdeburg	32	36.9	36.1	Trier	63	77.1	54.4
Main-Rhoen	82	57.1	54.0	Uckermark-Barnim	26	52.2	39.7
Mecklenburgische Seenplatte	10	50.9	31.7	Unterer Neckar	68	51.5	39.2
Mittelhessen	49	73.7	52.0	Vorpommern	9	65.3	33.2
Mittelrhein-Westerwald	62	71.9	55.1	Westmecklenburg	7	57.1	36.1
Mittelthueringen	54	51.9	33.6	Westmittelfranken	87	54.5	56.4
Mittlerer Oberrhein	70	60.3	44.3	Westpfalz	65	71.8	53.7
Mittleres Mecklenburg/Rostock	8	60.0	29.2	Westsachsen	57	13.7	25.0
Muenchen	93	35.4	33.0	Wuerzburg	81	59.7	44.0

Table 7: Correlation between alternative HOR measures, NUTS1 regions

Regional level	HOR measure	Pearson correlation	p-value	Spearman correlation	p-value	Kendall correlation	p-value	Pearson correlation	Pearson correlation
NUTS1	Destatis 2006[1]	0.874	0.000	0.897	0.000	0.800	0.000		
	empirica 2008,[2] households	0.834	0.000	0.845	0.000	0.711	0.000	0.963	
	empirica 2008,[3] persons	0.835	0.000	0.859	0.000	0.750	0.000	0.976	0.992
ROR	Destatis 2006[4]	0.761	0.000	0.700	0.000	0.519	0.000		

[1] Statistisches Bundesamt (2010), based on the microcensus data.
[2] Braun (2009a), based on the EVS data.
[3] Braun (2009b), based on the EVS data.
[4] DeStatis, based on the microcensus data.

Figure 1: In- and out-of-sample predictive power of the logit model estimated on texts of houses-for-sale ads

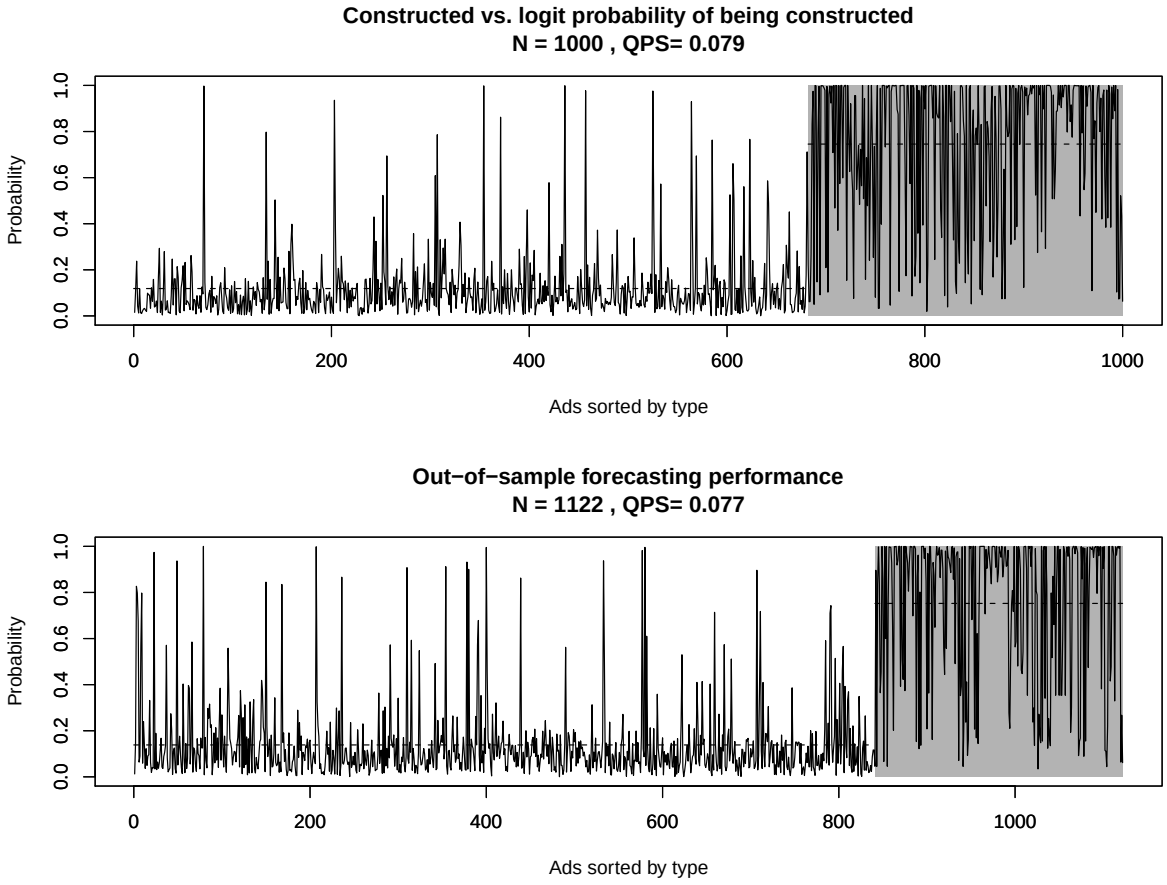

Figure 2: Comparison between our and alternative HOR measures at NUTS1 level

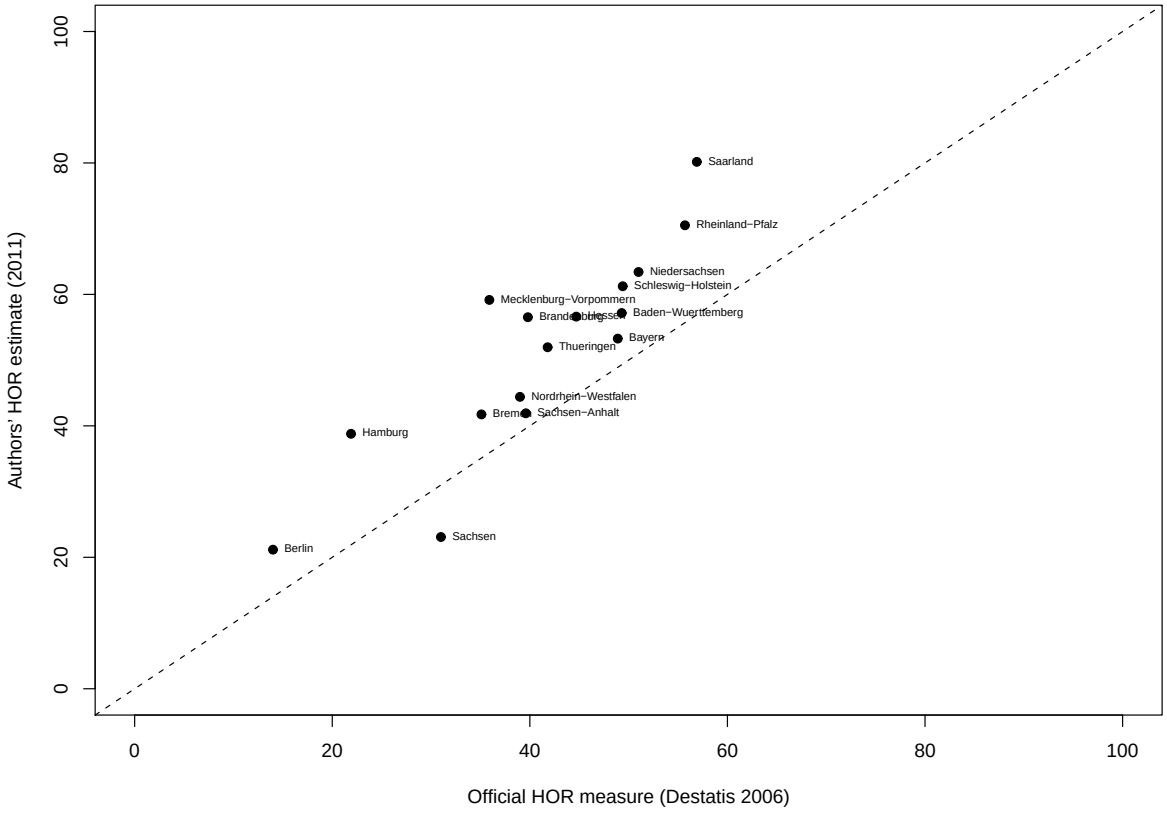

Figure 3: Comparison between our and alternative HOR measures ROR level

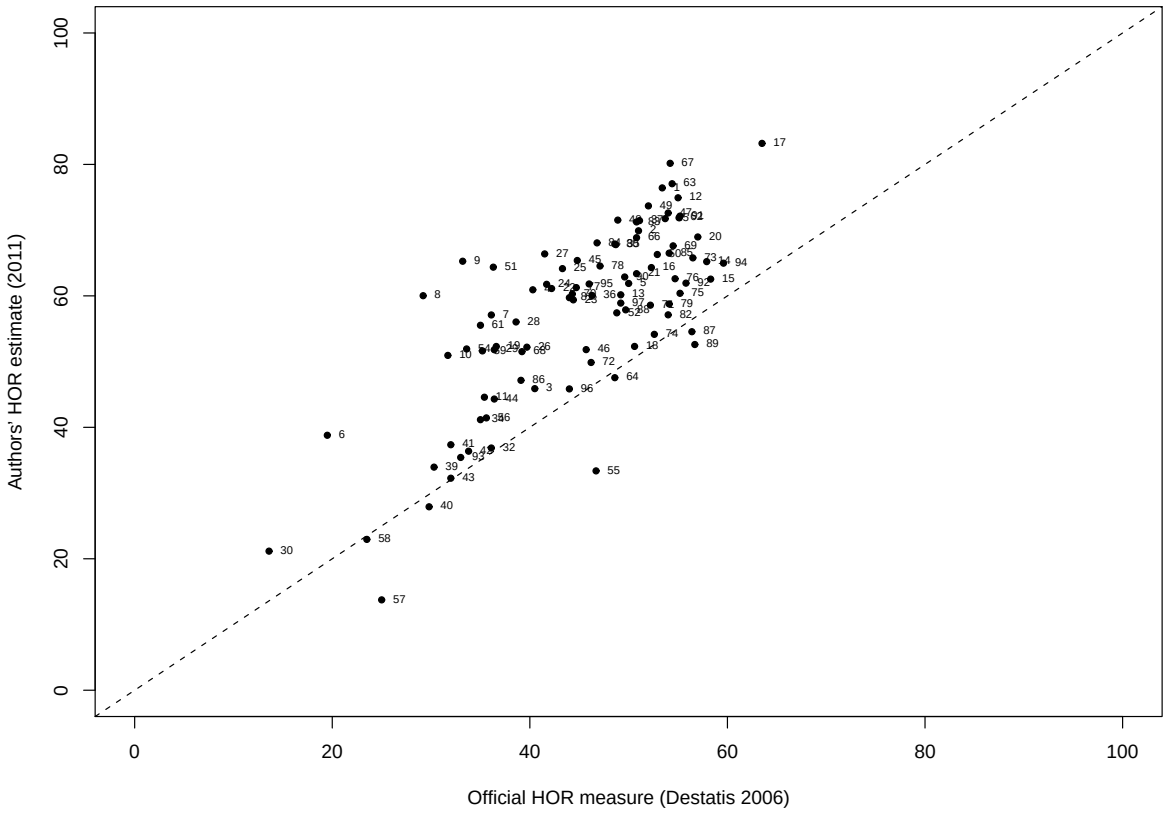

Figure 4: Distribution of our and alternative HOR measures