Answer to the referee report 2 to the paper

Urban House Prices: A Tale of 48 Cities

March 20, 2015

1. I have a bit of an issue with what exactly is measured. It is not really "house prices", because you do not observe the prices at which the houses are sold. It is the value of the first offer by the seller. Usually, if I am not mistaken, this is more of an upper bound, because the seller will decrease the price gradually if the appartment is not sold at the current price, and because the final price is a matter of negotiation between the parties. In some cases, the final price might be higher than the first offer, but I would argue this is usually not the case.

   Our answer: This is an important remark. What we use here are offer prices and not the final transaction prices. There are several studies comparing both prices: e.g., Faller et al. (2009) and Henger and Voigtländer (2014) for Germany. The findings of these studies indicate that on average the offer prices are 6-8% above the real transaction prices. Significantly smaller gaps are found for urban locations. The differences may also systematically change across the phases of business cycle. That said, we still have to make use of the offer prices as proxies for the transaction prices, which are simply not available for all the cities in question.

2. Moreover, I wonder why in some countries, there are so many, and in others so few appartments on offer online. Might this be because the proportion of flats being sold via internet differs between countries? I believe that many flats do not even appear online because they are directly sold through an agency, especially high-price flats. I was wondering if the appartments on offer online are somehow selective. The analysis is still absolutely worthwhile, it should just be made clear what the implications and limitations may be of the way that you measure house prices, and what kind of bias(es) might have been introduced by it.

Our answer: The number of advertisements online can depend on various factors: 1) size of the city, 2) internet accessibility (which must be nowadays more or less universal in European countries), 3) exclusive placement agreements between the landlords and internet portals, if there are many of such portals (for example, the number of internet ads for Leipzig is disproportionately large, given its population, because of such agreements between the housing cooperatives and the internet portal Immobilienscout24). This number *per se* should not pose any problem as soon as the structure of offers is representative. This is not the case when for some reason entire segments of dwellings are not sold over the internet. This can lead to a selection bias. In that case, however, it can become a problem when the coverage is systematically different in different cities: in some cities these segments are present in the internet ads, while in other cities they are not. Why should this be a case? The segment to be more likely underrepresented is that —as the referee correctly points out— of the expensive luxury apartments. These apartments represent rather a marginal part of the market and should be in a sense regarded as "outliers". So, their omission, if it takes place in all the cities, should not lead to a serious

bias.

3. I was wondering how you define "city". Does it include suburbs?

Our answer: Typically, in the internet real estate ad portals, the administrative borders are delineated. So, all the advertisements for Berlin should belong to the official administrative region of Berlin. Indeed, the suburbs even if they formally are not inside the official city boundaries may be closely related to the metropolis. Here, we confined ourselves to the official city boundaries, without considering the suburbs. The task of identifying "metropolitan areas" as opposed to the "administrative city" as defined by official documents and statistics is a paramount one and goes well beyond the limits of the paper. We are aware of the attempts of the OECD to define the metropolitan areas in the sense of the referee and to collect the corresponding statistics, see `http://www.oecd.org/regional/redefiningurbananewwaytomeasuremetropolitanareas.htm`. However, the database does not include all the cities from our sample: In particular, the cities in the former USSR countries are missing. Furthermore, to assign single observations (dwellings) to these areas would require the knowledge of their coordinates. Tackling this challenge could be a nice extension for this paper in some future research.

4. I was wondering why you do not control for housing stock, presumably because you do not have the data. Population size per se is not problematic if there is a large stock of maybe even empty housing, like it was the case in Berlin for a long time.

Our answer: Indeed, the relation between the supply (housing stock) and demand (number of households) is an important factor determining the house prices. The data availability is surely an issue. However, we can try to collect the corresponding data and use them

5. Also, and even more importantly, instead of using population and population density, I would rather use change in population as a determinant. When the population grows faster than the housing stock, that pushes up prices. And when the population shrinks or grows more slowly than in the past, housing will be widely available which will presumably cause prices to fall. It should be possible to get that data at the city level.

   Our answer: In this paper, we conduct a cross-section analysis concentrating on price levels. The population change definitely affects the price dynamics (changes) but not the level observed in a single point of time.

6. I thought it was unclear how and why you exclude outliers. E.g. what does it mean to be an outlier for "area"? Do you have any idea how the outliers might have been caused? Does it change the results to leave them in?

   Our answer: In order to recognize outliers a standard statistical criterion was used: If an observation is higher (lower) than the median by 1.5 times interquartile range, it is dropped from the sample. This was done for each city separately. In case of area it can mean an apartment, which is, say, 2000 m^2 large. The outliers are mostly caused by typos or announcements placed in a wrong section (for example, an entire multi-storey building place in the apartments section.).

7. Page 3: I would like you to make clearer that all the factors you include are demand side factors, whereas your dependent variable is the offer made by the supply side. This is not per se problematic but I found it a bit confusing.

   Our answer: We will do this. In addition, we could try to include the supply side factors,

such as housing stock if we find corresponding data.

8. I don't understand why you include a Euro area dummy, but no country dummy. I understand why you would not want to include a country dummy, namely because you don't have more than one observation for some countries and thus the dummy would eat up all of the remaining variance, and also because you are not interested in "cultural" factors, which I like in principle. But I need better reasons for including the Euro dummy.

   Our answer: You are right — the inclusion of 24 dummies in a regression with 48 observations would drastically reduce the number of degrees of freedom. The cultural factors are surely important and interesting, for they may affect the housing preferences. We would be grateful if you could suggest a variable to us capturing the relevant cultural differences. The dummy for the Euro area is included to account for the fact that the EA countries have a common monetary policy. This implies that the refinancing conditions of some countries like Greece and Portugal have largely benefited from the common currency. This not only refers to the government debts but also private loans. The lack of exchange rate risk attracts investors from other EA member states, which increases liquidity and potential investments. In consequence, house prices tend to be lower. Our estimates confirm this hypothesis.

9. This is a small issue: Not all tables and figures are mentioned in the text. I would mention them or leave them out.

   Our answer: We will thoroughly check the paper and drop all irrelevant tables and figures.

10. This one is a bigger issue. I am not so sure that you do not have an endogeneity problem with house prices and population size. I find it very well possible that people are deterred

from moving to a city if the housing prices are crazily high. In fact, I would never move to Paris because housing there is insane, even taking into consideration income levels (as you also find). Well, maybe not "never", but it is and has been a factor in my decision-making. If you proxy the number of housing units demanded in some way by the population, you get a classical demand-supply-simultaneous equations model, which you could solve in principle if you found one factor that is only relevant for supply (e.g. construction cost) and one that is only relevant for demand (e.g. unemployment). I don't know if this is possible to do as a robustness check here, maybe not due to data restrictions and also because distinguishing between demand for rental flats and for property is messy, but then you should argue better why you think that endogeneity is not a problem here.

Our answer: The population of a city in normal times tends to be relatively stable. For example, since 2000, the population of Paris has been staying at the level of 2.2 millions. The fact that some people (especially, foreigners) don not want to migrate to Paris due to its higher prices and lower housing quality apparently does not affect its population. It does not shrink. So, there appears to exist only a one-way causality: population affects prices, but prices have virtually no effect on the population. Hence, we would not be too much preoccupied about the endogeneity problem in this case.

11. P. 13, Table 1: The numbers don't add up correctly in the line "Total Income" (12+3 is not 14).

Our answer: We will thoroughly revise tables for such errors and correct them.

12. P. 20, Table 8: I would have been really interested in the values for the Gini coefficient. Maybe not absolutely necessary, but would be nice.

Our answer: We will make available a table with descriptive statistics for all the explanatory variables. Moreover, you can find the individual values of Gini index in the data set for this paper, which is made available in accordance with the policy of "Economics" at its repository in Dataverse, see `http://thedata.harvard.edu/dvn/dv/economics`.

13. P. 22 Table 10: Please make the table nicer and more readable by adding little stars for significance and by explaining the abbreviations of the regressors in a caption. Also, it confused me that you said that you retained only the variables with at least 5% significance level, but some of them actually had a lower significance level.

Our answer: We will add stars to the table. The somewhat higher p-values of estimated coefficients for some variables can be explained as follows: The selection of "small-model" specification was done in the PcGets that is only available in Ox programming language, while the coefficients for the model were estimated in R. Hence, there might be small differences in estimation results between these two languages. If a variable had a p-value close to 0.05 in PcGets, it could obtain a p-value of 0.057 in R.

14. P. 24. This is a very nice graph.

Our answer: Thank you very much. We like it also.

15. P. 25 This is not a very nice graph, and not very readable (at least in black and white). I would kick it out, since you find the information elsewhere in the paper.

Our answer: OK, this is not a problem (although it is in color), we will drop it from the paper.

16. I also really liked the robustness checks and the graphs on p. 27.

Our answer: Thank you.

17. P. 30 I thought about including a "capital city" dummy. Have you tried that?

Our answer: No, we did not. But we will definitely give it a try.

References

Faller, B., C. Helbach, A. Vater, and R. Braun (2009). Möglichkeiten zur Bildung eines Regionalindex Wohnkosten unter Verwendung von Angebotsdaten. RatSWD Research note 34.

Henger, R. and M. Voigtländer (2014). Transaktions- und Angebotsdaten von Wohnimmobilien — eine Analyse für Hamburg. *IW-Trends* (4).