Assignment 2 — Hedonic pricing

Kieran Marray and Hans Koster

January 2025

Introduction. In this assignment you will estimate the causal effect of being close to a metro on house value. To get a causal estimate, you use the opening of the Noord–Zuid line in Amsterdam in July 2018 as a natural experiment.

Figure 1: The Noord-Zuid metro line in Amsterdam

Your dataset contains estimated market values (WOZ) by house from the municipality of Amsterdam from January 1st 2014 to January 1st 2022. We treat these as a proxy of the house price, and refer to them as price. The dummy variable near_new_metro equals 1 if and only if a houses is in one of the closest 4-digit postcodes, denoted by i, to a new metro station.

Question 1. First, estimate the effect of a metro station opening nearby on house prices by estimating the 'long difference' linear regression

$$\ln \text{price}_{i,2022} - \ln \text{price}_{i,2014} = \alpha + \beta \text{ near\_new\_metro}_i + \epsilon_i$$

using OLS. Include one specification without controls, and another controlling for size, construction year, house type, and district fixed effects. Report coefficients in a journal-style regression table. Comment on the magnitude, sign and significance of the results using using **max. 3 sentences** per coefficient.

Question 2. A colleague is suspicious about the long-difference specification, and suggests that you instead look at the change in prices around the opening of the metro stations. You know that the Noord-Zuid line opened in July 2018. Take only the prices from January 1st 2018 and January 1st 2019. Estimate the effect of a metro station opening nearby (near_new_metro = 1) on house prices **using a difference-in-differences strategy**. To do this, estimate the linear regression

$$\text{price}_i = \alpha + \beta_1 \text{ near\_new\_metro}_i + \beta_2 \mathbf{1}_{2019} + \beta_3 \text{ near\_new\_metro}_i \times \mathbf{1}_{2019} + \epsilon_i$$

where $\mathbf{1}_{2019}$ is 1 if the year is 2019 and 0 otherwise.[1] Include one specification without additional controls. Include another controlling for size, construction year, house type, and district fixed effects. Report coefficients in a journal-style regression table. Comment on the magnitude, sign and significance of the results, using **max. 3 sentences** per coefficient.

Question 3. You often care about the marginal effect of one variable on another – here the marginal effect of being near a new metro station on price

$$\frac{\partial \text{price}_i}{\partial \text{near\_new\_metro}_i}.$$

Based on question 1, what is the marginal effect of being close to a new metro station on price? Using the prices in 2014, compute the marginal effect on total house prices in Amsterdam from opening the Noord-Zuid line in Euros. To do this, use the approximation that $\ln \text{price}_{i,2022} - \ln \text{price}_{i,2014} \approx$ average growth rate of prices over the period 2014–2022.

Question 4. You should find a smaller effect in question 2 than question 1. Why might only looking at changes in prices close to opening dates of stations underestimate the effect of opening the new line on house prices?[2] **max. 3 sentences**

Harder questions

Question 5. Policymakers are (partially) non-random. A colleague suggests that this approach would not work for estimating the causal effect of being close to a metro on house prices. She says that the estimates will be incorrect because policymakers want to build cost-effective infrastructure. So they likely built the Noord-Zuid line in places that would benefit the most from getting a new metro station.

Imagine that this is true. Is she correct that this will bias your estimates? How might you adjust the design to correct for this?

Question 6. Non-parametric regressions. You are worried that the relationship between price and size might not be linear. So you want to estimate non-parametric regression models and compare the fit to the linear model to see if you are correctly controlling for size.

[1] You should have come across difference-in-difference designs in previous econometrics courses. If not, then read Section 9 in Casual Inference: the Mixtape (here `https://mixtape.scunning.com/09-difference_in_differences`) before going on.

[2] If you need a hint, look at the history of the planning of the Noord-Zuid line here `https://nl.wikipedia.org/wiki/Metrolijn_52`

a) Estimate the relationship between price and size using a nearest neighbour local linear regression with $k = 4$ and a random forest. In R, you can estimate a random forest using the KNN.REG function from the FNN package. In Stata, you can use the NPREGRESS command with the correct kernel. In R, you can estimate a random forest using the RANDOMFOREST package. In Stata, you can use the RFOREST command. Plot the predicted price against size from each model for a grid of sizes. Compare this to a fitted linear regression model.

b) Based on the results, do you trust the way that you controlled for size in question 1? Use **max 1 paragraph**.

c) First, what are the marginal effects of size on house prices in our nearest-neighbour local linear regression? You cannot compute marginal effect analytically for our random forest model (if this is not obvious, why?). So approximate the marginal effects of size on price numerically for a grid of sizes using the finite difference method.

$$\frac{\partial y}{\partial x}$$

around $x = x^*$. Because calculating a marginal effect involves looking at the effect of small changes in the variable, you can take a small difference in x before x^* $\partial x \approx \epsilon$ and compute the change $\partial y \approx \hat{y}(x^*) - \hat{y}(x^* - \epsilon)$.)

References

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Koster, H. R. A. and van Ommeren, J. (2019). Place-Based Policies and the Housing Market. *The Review of Economics and Statistics*, 101(3):400–414.

Michaels, G., Nigmatulina, D., Rauch, F., Regan, T., Baruah, N., and Dahlstrand, A. (2021). Planning ahead for better neighborhoods: Long-run evidence from tanzania. *Journal of Political Economy*, 129(7):2112–2156.

Rossi-Hansberg, E., Sarte, P., and Owens, R. (2010). Housing externalities. *Journal of Political Economy*, 118(3):485–535.