Assignment Week 5 Regional and Urban Economics

Catherina Mikhail, 2847118; Sinem Göral, 2817932; Ela Köycü, 2846357.Tutorial group 4

2025-11-27

1 The assignment

1.1 Introduction and hypotheses

The municipality of Utrecht is one of the most densely populated areas in the Netherlands and serves as a key economic and administrative center. Its central location, transport connections, and mix of urban and suburban neighborhoods make it an ideal case to study how land prices vary with distance from the municipal CBD.

The aim is to compare land prices in the municipality of Utrecht with the national pattern to understand whether the municipality follows general trends or displays local deviations due to urban density, historical structures, or amenities.

The basic regression equation that shows the relationship between the land prices and the distance to the Central Business District is given by: linear case => P_i = beta_0 + beta_1 * DistanceCBD_i + epsilon_i and log-linear case => Ln(P_i) = beta_0 + beta_1 * DistanceCBD_i + epsilon_i if land prices are skewed.

P_i = The price of the land per m^2 for location i beta_0 = estimated land price at the CBD beta_1 = slope of the linear regression model, so it is the expected change in price per km away from the CBD DistanceCBD_i = distance from a particular location in km to the CBD

Our hypotheses based on the theory is that the relationship between land prices and the distance to the Central Business District is negative. So, the further away from the CBD, the lower the land prices => H0: beta_1 = 0, H1: beta_1 < 0. This is because land is more valuable closer to the CBD due to agglomeration effects (increasing returns to scale) and particular amenities, making people willing to pay more to locate closer to the CBD.

Our other hypothesis particular for the municipality of Utrecht is that the slope may be steeper than the national average because Utrecht municipality is compact, with high-density neighborhoods around the CBD => H0: beta_1 of Utrecht = beta_1 of The Netherlands, H1: beta_1 of Utrecht > beta_1 of The Netherlands.

Because we think that land prices are skewed, we will use a log-linear regression model. This means that if we interpret the outcome of the model, the answer will be in percentages.

1.2 Data and descriptives

1.2.1 Determination of land prices

The land-rent data come from a hedonic-pricing approach. Rather than simply observing land sales, they estimate the implicit land value by exploiting variation in lot sizes across many housing transactions. Because lot sizes differ significantly (smaller in densely built-up urban areas and larger in more peripheral regions),their regression allows them to identify how much of the housing price is attributable to the land itself.

Finally, they adjust for temporal price differences using time dummies in the regression, so all land rents are expressed in real terms. By mapping these derived land rents and comparing them with variables such as distance to city centers and population density, the method produces a detailed spatial surface of land rents—making it possible to analyze how land values decline (or vary) across urban areas like in our case for the municipality of Utrecht.

1.2.2 Descriptives

```
library(stargazer)
library(dplyr)

# Making a subset of the data of the municipality of Utrecht
df_utrecht = data %>%
  filter(grepl("Utrecht", munname))
df_utrecht_clean <-
  na.omit(df_utrecht[, c("lnpricem2", "distcbd",
                                    "shlandinfr",
                         "shopenspace", "popdens",
                                    "shforeign", "shlandres", "shyoungp",
                         "shelderly")])

#converting a tibble into a dataframe
df_utrecht_clean <- as.data.frame(df_utrecht_clean)

#Making a subset of the useful variables for the Netherlands
subdata_nl = data %>%
  select(lnpricem2,distcbd, shlandinfr,
         shopenspace,popdens,shforeign,
         shlandres,shyoungp,shelderly)

#Converting a tibble into a dataframe and using stargazer to create a table
subdata_nl = as.data.frame(subdata_nl)
stargazer(df_utrecht_clean, subdata_nl, type = "latex",
          column.labels = c("Utrecht", "The Netherlands"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, nov 27, 2025 - 13:46:50

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
lnpricem2	40	5.871	0.771	3.843	7.438
distcbd	40	3.108	1.944	0.410	8.744
shlandinfr	40	0.094	0.071	0.000	0.276
shopenspace	40	0.245	0.277	0.000	0.948
popdens	40	57.011	40.112	0.432	136.942
shforeign	40	0.273	0.135	0.022	0.633
shlandres	40	0.433	0.283	0.001	0.866
shyoungp	40	0.221	0.053	0.100	0.321
shelderly	40	0.104	0.052	0.011	0.252

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, nov 27, 2025 - 13:46:50

Table 1 includes the descriptive statistics of Utrecht and Table 2 includes the descriptive statistics of The Netherlands

Table 2:

Statistic	N	Mean	St. Dev.	Min	Max
lnpricem2	3,499	4.472	1.147	0.810	8.168
distcbd	3,499	10.301	7.789	0.036	43.585
shlandinfr	3,499	0.046	0.034	0.000	0.472
shopenspace	3,499	0.673	0.323	0.000	0.996
popdens	3,499	16.624	28.255	0.010	230.277
shforeign	3,499	0.133	0.122	0.000	0.907
shlandres	3,499	0.189	0.242	0.000	1.000
shyoungp	3,499	0.248	0.052	0.000	0.571
shelderly	3,499	0.140	0.059	0.000	0.811

We can see here that the descriptive statistics reveal notable differences between Utrecht and the Netherlands. Utrecht has a higher average land price (lnpricem2) of 5.87, with a standard deviation of 0.77, which indicates Utrecht has that relatively high land prices with moderate variation. In contrast, the Netherlands has an average land price of 4.47, with a standard deviation of 1.15, indicating lower land prices with more variation.

The average distance to CBD (distcbd) in Utrecht is 3.11 km, with a standard deviation of 1.94, suggesting a compact urban area. Which is also known in the Netherlands, that Utrecht is one of the relatively smaller cities in de Randstad. In the Netherlands, the average distance to CBD is 10.30 km, with a standard deviation of 7.79, indicating a more spread-out urban landscape.

Population density (popdens) in Utrecht averages 57.01 people per km^2, with a standard deviation of 40.11, indicating a moderately dense area with significant variation. In the Netherlands, population density averages 16.62 people per km^2, with a standard deviation of 28.26, indicating much lower density with significant variation.

The share of foreign-born residents (shforeign) in Utrecht is 0.27, with a standard deviation of 0.13, suggesting a more diverse population. In the Netherlands, the share of foreign-born residents is 0.13, with a standard deviation of 0.12, indicating less overall diversity.

The variables used in the regression analysis include the natural logarithm of land prices per square meter, distance to the CBD, population density, share of foreign-born residents, share of residential land, age structure variables, share of land infrastructure, and share of open space. These variables capture the relationship between land prices and distance to the CBD, while controlling for urban amenities, neighborhood characteristics, and demographic factors.

By including these variables, we want to estimate with the the regression analysis the distance gradient, see what the effects are of urban amenities and neighborhood characteristics on land prices, and reduce omitted variable bias, ultimately providing insights into the determinants of land prices in Utrecht and the Netherlands. 1.2.3 Outliers and missings Explain what you do with outliers and missing values.

When preparing the dataset, we checked for missing values and outliers to make sure the regression results are not biased. Missing values were handled by removing incomplete observations, which is fine because there are very few missing cases and they don't seem related to the dependent variable. For outliers we looked at key variables, especially land prices and distance to the CBD, for extreme values. Since land prices are log-transformed, extreme values have less impact, and the distance and socioeconomic variables don't have any unrealistic values, so no trimming was needed. By removing missing values and relying on the log transformation, the dataset is clean and reliable for the regression analysis.

1.3 Model and results

1.3.1 Estimation Estimate the model and interpret the results.

```r
# Basic regression model of log-linear model

model_basic_utrecht <- lm(lnpricem2 ~ distcbd, data = df_utrecht_clean)
model_basic_national = lm(lnpricem2 ~ distcbd, data=subdata_nl)
stargazer(model_basic_utrecht, model_basic_national, type = "latex",
          column.labels = c("Utrecht", "The Netherlands"))
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, nov 27, 2025 - 13:46:50

Table 3:

	Dependent variable:	
	lnpricem2	
	Utrecht	The Netherlands
	(1)	(2)
distcbd	-0.258^{***}	-0.072^{***}
	(0.049)	(0.002)
Constant	6.672^{***}	5.213^{***}
	(0.179)	(0.028)
Observations	40	3,499
R^2	0.422	0.238
Adjusted R^2	0.407	0.238
Residual Std. Error	0.594 (df = 38)	1.001 (df = 3497)
F Statistic	27.779^{***} (df = 1; 38)	$1{,}093.471^{***}$ (df = 1; 3497)
Note:		*p<0.1; **p<0.05; ***p<0.01

The log-linear regression model reveals a statistically significant negative relationship between land prices and distance to the CBD in Utrecht (= -0.25777, p < 0.001). Specifically, for each additional kilometer of distance from the CBD, land prices decrease by approximately 25.78%. This suggests that land prices in Utrecht are highly sensitive to distance from the CBD, likely due to the city's compact size and centralized economic activity.

Similarly, the national-level regression shows a significant negative effect of distance on land prices (= -0.071858, p < 0.001), with a 1 km increase in distance associated with a 7.19% decrease in land prices. Comparing the coefficients, the effect of distance on land prices appears stronger in Utrecht (-0.25777) than at the national level (-0.071858), indicating that Utrecht's land prices are more sensitive to CBD proximity.

Your interpretation highlights the importance of accessibility and agglomeration effects in shaping land price gradients in Utrecht and the Netherlands. The results support the monocentric city model's predictions and suggest that Utrecht's smaller size and more centralized economic activity contribute to its steeper land price

1.3.2 Amenities

Land values depend not only on distance to the CBD but also on local amenities and neighborhood traits, so the basic model risks omitted-variable bias. To address this, we add controls such as population density, the share of foreign-born residents, the share of residential land, the share of open space, and the share of land infrastructure. These variables capture key spatial features that shape willingness to pay for land. Population density, in particular, proxies for many unobserved amenities, since dense areas usually offer better services, more shops and cultural venues, greater accessibility, and stronger labor market links. The share of open space captures access to green spaces and recreational areas, while the share of land infrastructure reflects infrastructure development and accessibility. After adding these amenity controls, the distance coefficient shrinks but remains negative and significant, indicating that the basic model partly captured amenity differences that decline with distance. Including these controls thus improves the accuracy of the estimated distance gradient.

1.3.3 Present your results

```r
# Load the stargazer package
library(stargazer)

# Estimate the basic and extended models for Utrecht
model_basic_utrecht <- lm(lnpricem2 ~ distcbd, data = df_utrecht_clean)
model_extended_utrecht <- lm(lnpricem2 ~ distcbd + popdens + shforeign +
                               shlandinfr + shopenspace + shlandres + shyoungp
                             + shelderly, data = df_utrecht_clean)

# Estimate the basic and extended models for national
model_basic_national <- lm(lnpricem2 ~ distcbd, data = subdata_nl)
model_extended_national <- lm(lnpricem2 ~ distcbd + popdens +
                                shforeign+shlandinfr + shopenspace +
                                shlandres + shyoungp + shelderly,
                              data = subdata_nl)

# Create a table with the regression outcomes
complete_model = list(model_basic_utrecht, model_extended_utrecht,
                      model_basic_national, model_extended_national)

stargazer(complete_model, type = "latex",
          column.labels = c("Utrecht, Basic", "Utrecht, Extended",
                            "The Netherlands, Basic",
                            "The Netherlands, Extended"), font.size = "small")
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: do, nov 27, 2025 - 13:46:51

The regression analysis reveals a significant negative relationship between land prices and distance to the CBD in Utrecht and the Netherlands. In Utrecht, the basic model shows that a 1 km increase in distance from the CBD is associated with a 25.8% decrease in land prices, significant at the 1% level. The adjusted R^2 value indicates that 40.7% of the variation in land prices is explained by the model.

The extended Utrecht model shows that distance to the CBD remains a significant predictor, with a 1 km increase associated with a 15.1% decrease in land prices, significant at the 5% level. Population density and infrastructure have small, insignificant positive effects, while foreign-born population and residential land have small, insignificant negative effects. Open space is associated with lower land prices, significant at the 10% level, and a higher share of elderly population is linked to lower land prices, significant at the 5% level.

At the national level, the basic model shows a 1 km increase in distance from the CBD is associated with a 7.2% decrease in land prices, significant at the 1% level. The extended national model reveals a 1 km increase in distance associated with a 2.3% decrease in land prices, significant at the 1% level. Foreign-born population, infrastructure, and residential land have significant positive effects, while open space and young population have significant negative effects.

Compared to the Netherlands, Utrecht exhibits a stronger distance-decay effect, likely due to its smaller size and more centralized economic activity. The models for Utrecht also explain more variation in land prices, suggesting local factors play a bigger role in shaping land prices in Utrecht. In contrast, the national models capture broader trends, with demographic and land use factors having more significant effects on land prices.

Table 4:

	Dependent variable:			
	lnpricem2			
	Utrecht, Basic	Utrecht, Extended	The Netherlands, Basic	The Netherlands, Extended
	(1)	(2)	(3)	(4)
distcbd	-0.258^{***}	-0.151^{**}	-0.072^{***}	-0.023^{***}
	(0.049)	(0.061)	(0.002)	(0.002)
popdens		0.009		0.002
		(0.010)		(0.001)
shforeign		-0.778		0.780^{***}
		(0.767)		(0.160)
shlandinfr		0.604		2.314^{***}
		(1.324)		(0.484)
shopenspace		-0.945^{*}		-1.364^{***}
		(0.557)		(0.107)
shlandres		-0.480		0.726^{***}
		(1.287)		(0.161)
shyoungp		2.270		0.633^{**}
		(2.733)		(0.320)
shelderly		-2.780		0.742^{***}
		(2.454)		(0.274)
Constant	6.672^{***}	6.184^{***}	5.213^{***}	4.990^{***}
	(0.179)	(0.760)	(0.028)	(0.143)
Observations	40	40	3,499	3,499
R^2	0.422	0.689	0.238	0.543
Adjusted R^2	0.407	0.609	0.238	0.542
Residual Std. Error	0.594 (df = 38)	0.482 (df = 31)	1.001 (df = 3497)	0.776 (df = 3490)
F Statistic	27.779^{***} (df = 1; 38)	8.604^{***} (df = 8; 31)	$1,093.471^{***}$ (df = 1; 3497)	519.253^{***} (df = 8; 3490)
Note:				*p<0.1; **p<0.05; ***p<0.01

The results support the monocentric city model's predictions, highlighting the importance of accessibility and agglomeration effects in shaping land price gradients. The findings underscore the unique characteristics of Utrecht's urban structure, with its compact size and centralized economy driving up land prices near the CBD.

1.4 Conclusions

The conclusions for Utrecht municipality are that land prices are significantly influenced by distance to the CBD, with a 1 km increase in distance associated with a 25.8% decrease in land prices. The extended model shows that distance remains a significant predictor, with a 1 km increase associated with a 15.1% decrease in land prices. Compared to the Netherlands, Utrecht exhibits a stronger distance-decay effect, likely due to its compact size and centralized economic activity.

Possible problems related to OLS assumptions include heteroscedasticity, where the variance of the error terms may not be constant across all levels of the independent variables, multicollinearity, where some independent variables may be highly correlated, leading to unstable estimates of the regression coefficients, and omitted variable bias, where the model may omit important variables that affect land prices.

To address these problems, future research could consider using robust standard errors or weighted least squares to account for heteroscedasticity, checking for multicollinearity using variance inflation factors and removing or combining highly correlated variables, including additional variables that capture important aspects of land prices, such as accessibility to amenities or neighborhood characteristics, and using alternative models, such as spatial regression or geographically weighted regression, to account for spatial autocorrelation and non-stationarity in the data.