The Effects of School Proximity and the School Social Index on Housing Prices

4th February 2026

Yanyi Ji Daniel Tobien Luc Wichtmann

Agenda

1. **Motivation, Existing Literature & Research Question**
Why school proximity and social indices matter.

2. **Data & Methodology**
Data used and our empirical model.

3. **Results & Main Findings**
Presentation and interpretation of our findings.

4. **Limitations & Conclusion**
Limitations of our work and concluding thoughts.

Motivation, Existing Literature & Research Question

Motivation

- **Personal Interest**: Analyzing the real estate market's reaction to various factors is particularly interesting from an economic standpoint, as it offers insights into market dynamics. Additionally, the topic is personally intriguing, adding an extra layer of motivation to explore it further.
- **Academic and Practical Relevance**: Housing markets are influenced by various social and economic factors, and school quality is also cited as a underestimated determinant of property values, especially in research (Seo and Simons 2009).

Motivation

Understanding the relationship between school distance, school quality and property prices can provide valuable insights for researchers and individuals interested in the real estate market.

Related Literature Background

Several papers have examined the effect of proximity to schools on housing prices:

- **Empirical Strategies**: While hedonic models are the dominant approach to estimating housing price effects, alternative empirical strategies have also been proposed (see Black and Machin 2011)
- **Findings**: Empirical evidence suggests that proximity to schools positively affects housing prices (Rosiers, Lagana, and Theriault 2001; Huang and Hess 2018)

Related Literature Background

- **Other Factors**: Related work also documents capitalization effects of school reputation (Chin and Foong 2006), differences between public and private schools (Sah, Conroy, and Narwold 2016) and school quality (Edusei, Espey, and Lin 2007; Metz 2015)
- **Environment of studies**: Much of the literature focuses on settings with binding school catchment areas and measures school quality primarily through test scores rather than demographic composition.

Research Question

Out contribution: We study a setting **without binding school catchment areas** (NRW, Germany) and assess whether both school proximity and a **demographically based school social index** are capitalized into housing prices.

Research Question

Research Question: How do school proximity and school social indices affect the prices of detached houses, and do these effects differ between primary and secondary schools in North Rhine–Westphalia?

H1: House prices are higher related to schools
Expected: Negative coefficients on distance variables

H2: Social index moderates the effect of school proximity on house prices
Expected: Better schools leading to stronger price premiums

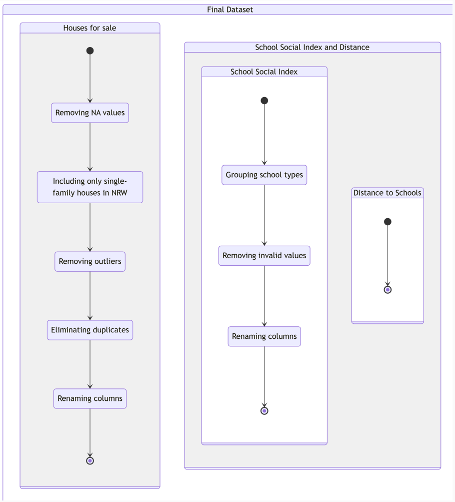
Data & Methodology

Data used

- **Housing Data**: Cross-section real-estate datasets for Germany published by the Research Data Center Ruhr (FDZ Ruhr) (RWI and ImmobilienScout24 2023).
- **School Data**: School social index dataset from North Rhine-Westphalias Ministry of Schools and Education (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen 2024).
- **Geodata**: Geodata for Germany's administrative areas used to create graphics of NRW. Taken from Germany's Federal Agency for Cartography and Geodesy (**BKG_2025?**).

Housing and school data both refer to the year 2022.

Pre-Processing of Data



Empirical Framework

We estimate regression-based hedonic models to

- examine how school proximity and school characteristics are capitalized into housing prices and
- whether the impact of school proximity on housing prices differs by the school's social index

Empirical Framework: Hedonic Price Model

$$\log(P_i) = \beta_0 + \beta_1 \text{Dist}_i + X_i'\gamma + \varepsilon_i$$

- P_i denotes the listing price of house i
- Dist_i measures distance to the nearest school
- X_i is a vector of housing characteristics
- γ is a vector of coefficients corresponding to X_i
- ε_i captures unobserved factors

Empirical Framework: Hedonic Price Model + Interaction

$$\log(P_i) = \beta_0 + \beta_1(\text{Index}_i \times \text{Dist}_i) + X_i'\gamma + \varepsilon_i$$

- P_i denotes the listing price of house i
- Index_i is the school social index
- Dist_i measures distance to the nearest school
- X_i is a vector of housing characteristics
- ε_i captures unobserved factors

Empirical Framework: Final Model

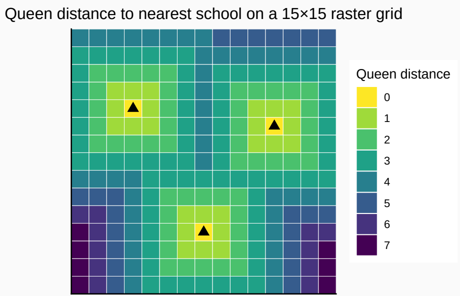
$$\log(P_i) = \beta_0 + \beta_1 \text{Dist}_i + \beta_2 \text{Index}_i \\ + \beta_3 \left(\text{Dist}_i \times \text{Index}_i\right) + X_i' \gamma + \mu_{\ell(i)} + \varepsilon_i$$

- P_i denotes the listing price of house i
- Dist_i measures distance to the nearest school
- X_i is a vector of housing characteristics
- Index_i is the school social index
- $mu_{\ell(i)}$ are municipality fixed effects
- ε_i captures unobserved factors

Empirical Framework: Queen contiguity

$$\text{Dist}_{ij}^{Q} = \max\Big(|x_i - x_j|,\ |y_i - y_j|\Big)$$

- Results for distance and the social index are re-estimated using Queen (Chebyshev) distance to check robustness of our results.



Results and Main Findings

Distance to Schools and House Prices: Hypothesis

H1: House prices are higher related to schools
Expected: Negative coefficients on distance variables

Empirical Framework

- **Outcome**: Log house price
- **Key variables**: Distance to nearest *primary* school, distance to nearest *secondary* school
- **Modeling Strategy**: Continuous, quadratic, and binned specifications

Distance to Schools and House Prices: Regression Strategy

Baseline OLS Controls

Standardizing housing attributes

- Living area
- Plot area
- Number of rooms
- House age

Two Distance Designs

Alternative Specifications

1. Continuous
   - quadratic (km^2)
2. Binned
   - 0–3 km, 3–6 km, 6–9 km, $>$9 km

Regression Specifications

(1) Naive (Unconditional)

$$\ln(P) = \alpha + \beta_1 D^{prim} + \beta_2 D^{sec} + \varepsilon$$

- **Goal:** Establish raw spatial correlation.
- **Risk:** High Omitted Variable Bias (OVB).

(2) Baseline (Preferred)

$$\ln(P) = \alpha + \mathbf{D} + \mathbf{X} + \varepsilon$$

- **Controls:** Area, Plot, Rooms, Age.
- **Hypothesis:** $\beta < 0$ (Negative semi-elasticity).

(3) Non-Linearity (Polynomial)

$$\ln(P) = \cdots + \delta_1 (D^{prim})^2 + \ldots$$

- **Test:** $\delta \neq 0$ implies varying MWTP.
- **Logic:** Test for convex/concave decay.

(4–5) Robustness (Single Type)

$$\ln(P) = \alpha + \beta_k D^k + \mathbf{X} + \varepsilon$$

- **Idea:** Estimate primary / secondary separately.
- **Check:** Stability against multicollinearity.

Descriptive Evidence: Price-Distance Gradient

Negative Slope (H1)

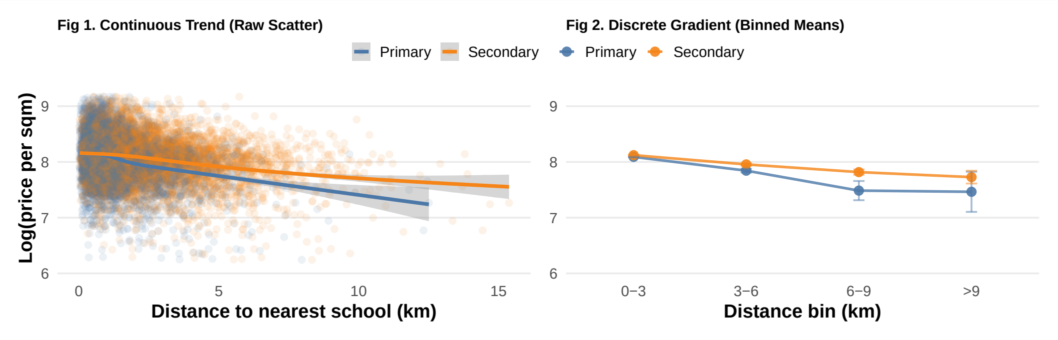
Clear decay in prices as distance increases across specifications.

Non-Linearity

Steeper drop in the first **3–6 km**, flattening afterwards.

Heterogeneity

Primary schools show a visibly steeper gradient than Secondary.



Main Results: Continuous Model

	Naive		Baseline		Polynomial		Primary Only		Secondary Only	
Variable	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**
Distance to primary (km)	-0.07***	(0.007)	-0.071***	(0.005)	-0.107***	(0.011)	-0.103***	(0.005)		
Distance to secondary (km)	-0.031***	(0.004)	-0.034***	(0.003)	-0.023***	(0.008)			-0.057***	(0.003)
Distance to primary (km)^2					0.007***	(0.002)				
Distance to secondary (km)^2					-0.001	(0.001)				
N	5794		5794		5794		5794		5794	
R^2	0.06		0.436		0.438		0.425		0.419	
Adj. R^2	0.06		0.436		0.437		0.425		0.418	

Note:
Dependent variable: log house price.
HC1 robust standard errors in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Main Results: Binned Model

	Naive		Baseline		Primary Only		Secondary Only	
Variable	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Primary: 3-6 km	-0.147***	(0.027)	-0.146***	(0.021)	-0.268***	(0.019)		
Primary: 6-9 km	-0.361***	(0.094)	-0.369***	(0.073)	-0.577***	(0.071)		
Primary: >9 km	-0.833***	(0.255)	-0.538***	(0.199)	-0.793***	(0.192)		
Secondary: 3-6 km	-0.142***	(0.017)	-0.146***	(0.013)			-0.179***	(0.012)
Secondary: 6-9 km	-0.241***	(0.033)	-0.256***	(0.026)			-0.331***	(0.024)
Secondary: >9 km	-0.209***	(0.08)	-0.294***	(0.063)			-0.435***	(0.059)
N	5794		5794		5794		5794	
R^2	0.046		0.42		0.401		0.413	
Adj. R^2	0.045		0.419		0.4		0.412	

Note:
Dependent variable: log house price. Reference: 0-3 km.
HC1 robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Compare: Continuous vs Binned

	Continuous		Binned	
Variable	**Coef.**	**SE**	**Coef.**	**SE**
Distance to primary (km)	-0.071***	(0.005)		
Distance to secondary (km)	-0.034***	(0.003)		
Primary: 3-6 km (ref: 0-3)			-0.146***	(0.021)
Primary: 6-9 km (ref: 0-3)			-0.369***	(0.073)
Primary: >9 km (ref: 0-3)			-0.538***	(0.199)
Secondary: 3-6 km (ref: 0-3)			-0.146***	(0.013)
Secondary: 6-9 km (ref: 0-3)			-0.256***	(0.026)
Secondary: >9 km (ref: 0-3)			-0.294***	(0.063)
N	5794		5794	
R^2	0.436		0.42	
Adj. R^2	0.436		0.419	

Model comparison (main models)

- Continuous fits slightly better.

Model	N	R^2	Adj. R^2	AIC	BIC	CV RMSE	CV MAE	CV R^2
Baseline (Continuous)	5794	0.436	0.436	4973.8	5027.1	0.371	0.288	0.434
Baseline (Binned)	5794	0.420	0.419	5143.6	5223.6	0.377	0.293	0.416

Note:
In-sample: higher R^2/Adj. R^2 is better; lower AIC/BIC is better. Out-of-sample (5-fold CV): lower RMSE/MAE is better; higher R^2 is better.

Robustness checks (main models)

- Binned is slightly more robust to influential points.

Model	BP p-value	Max VIF	Influential ($D > 4/n$)	Max Cook's D
Good Reference	$<$0.001	27.57	315	0.033
Average Reference	$<$0.001	109.13	315	0.033
Bad Reference	$<$0.001	578.70	315	0.033

Note:

Inference uses textbfHC1 robust SE. Lower BP p-values indicate heteroskedasticity; higher VIF indicates collinearity; larger Cook's D indicates influential points.

Robustness check: Queen contiguity (Distance)

	Primary Schools		Secondary Schools	
Variable	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Queen distance	-0.077***	(0.006)	-0.026***	(0.003)

Note:

Dependent variable: log house price, control variables included but not shown.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Economic Mechanisms & Interpretations

1. Capitalization

Why closer = expensive?

- **Convenience Premium:** Time savings → Capitalized into property values.
- **Sorting:** Families cluster → Higher local demand.

2. Primary vs Secondary

Why primary is stronger?

- **Age Constraints:** Young kids need escorting (Inelastic demand).
- **Micro-Neighborhoods:** Primary schools anchor local community ties.

3. Non-Linearity

Why gradient flattens?

- **Walkability Limit:** Premium vanishes after **3km**.
- **Substitution:** Driving → Public transit at longer distances.

Exploring the Social Index Impact on House Prices

H2: Social index moderates the effect of school proximity on house prices
Expected: Better schools leading to stronger price premiums

- **Key Finding from Distance Analysis:** We've already established that *households are willing to pay a premium* for proximity to schools
- **Next Question:** Now, we investigate *whether this price premium is driven by school quality*, measured through the *Social Index*

What is the Social Index?

Definition: The Social Index is a measure that shows the socio-economic and demographic background of a schools students.

Components of the social index are:

- Parental Education
- Socio-economic Status of the Family
- Migration Background
- Single-Parent Household
- Unemployment in the Family
- Housing Situation

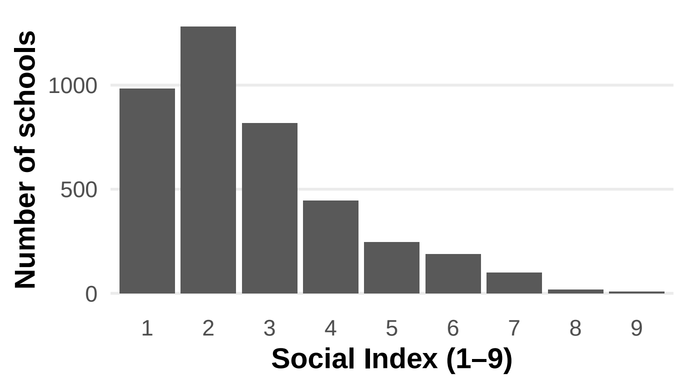
–> It provides insights into the **social environment** that might influence housing prices in that area.

Categorization of the Social Index

Social Index Classification:

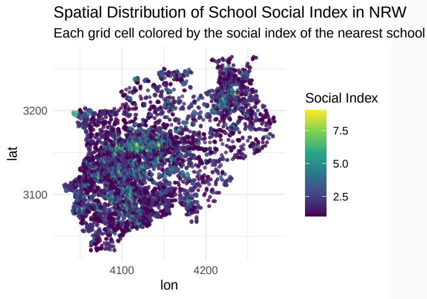
- **Good Schools**: Social Index 1-2
- **Average Schools**: Social Index 3-4
- **Bad Schools**: Social Index 5-8
- **Skewed Distribution**: The Social Index classes are heavily skewed, with many schools falling into classes 1, 2 & 3. We excluded class 9 for our analysis because of too less observations

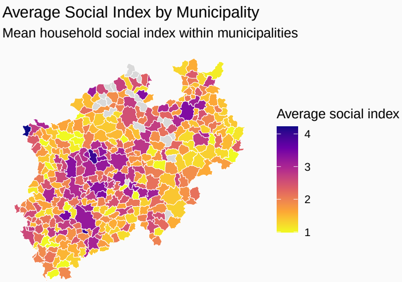
Categorization of the Social Index: Bar Chart



Graph: Distribution of Social Index values across schools showing the skewed nature of the data.

Spatial Distribution of School Social Index in NRW





Hedonic Regression Model for Social Index

$$\begin{aligned}\log(\text{price}_i) &= \beta_0 + \beta_1\,\text{dist\_primary}_i + \beta_2\,\text{dist\_secondary}_i \\ &\quad + \beta_3\,\text{school\_quality}_i + \beta_4\,(\text{dist\_primary}_i \times \text{school\_quality}_i) \\ &\quad + \beta_5\,(\text{dist\_secondary}_i \times \text{school\_quality}_i) + X_i'\gamma + \varepsilon_i\end{aligned}$$

Where:

- $\log(P_i)$ denotes the natural logarithm of the housing price of house i
- $\text{dist}_i^{\text{primary}}$ measures distance to the nearest primary school (km)
- $\text{dist}_i^{\text{secondary}}$ measures distance to the nearest secondary school (km)
- school_quality_i is a categorical indicator of school quality
- X_i is a vector of housing characteristics
- ε_i captures unobserved factors

Social Index: Regression Results

	Reference: Good School Quality
Distance to primary school (km)	-0.080***
Distance to secondary school (km)	-0.042***
Log living area	0.873***
Log plot area	0.054***
Rooms	-0.012***
House age	-0.002***
School Quality: Average	-0.015
School Quality: Bad	-0.027
Interaction: Distance to primary school * Average School Quality	-0.036**
Interaction: Distance to primary school * Bad School Quality	0.057
Interaction: Distance to secondary school * Average School Quality	0.035***
Interaction: Distance to secondary school * Bad School Quality	-0.006
Num Obs	5478

Robustness check: Queen contiguity (Social Index)

Variable	No interaction (Primary)	No interaction (Secondary)	With interaction (Primary)	With interaction (Secondary)
Social index	0.001	0.003	0	0.011*
Social index $\times$ Queen distance	NA	NA	0.001	-0.005*

Note: Dependent variable: log house price. Control variables included but not shown. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Social Index: Interpretaition of the Results

Primary School Proximity

- *Distance to primary school* is still *significant* and negatively affects house prices.
- Interaction Effect: The effect is *stronger for average schools* (Estimate = -0.041, p = 0.009), suggesting that the price drop is greater when the school quality is average.
- *No significant effect* for bad schools (p = 0.2996).

Social Index: Interpretaition of the Results

Secondary School Proximity

- *Distance to secondary school* is also still *significant* and negatively impacts house prices.
- The effect is *stronger for average schools* (Estimate = 0.0307, $p = 0.0096$), meaning proximity to secondary schools is more impactful in areas with average school quality.
- *No significant effect* for bad schools ($p = 0.7414$).

Concluding Thoughts and Limitations

Concluding thoughts: Hypotheses Assessment

H1: House prices are higher related to schools
Expected: Negative coefficients on distance variables

Supported

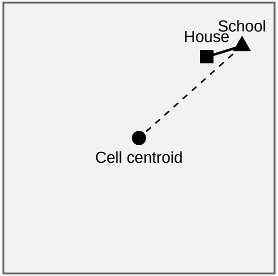
H2: Social index moderates the effect of school proximity on house prices
Expected: Better schools leading to stronger price premiums

Not supported

Limitation: Direct Distance

The distance is not measured from the houses to the nearest school directly but from the school to the center of the raster cell.

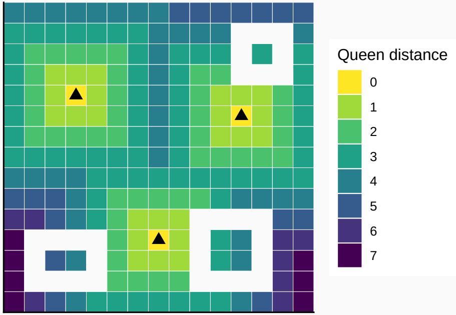
Single raster cell used for distance computation



Limitation: Fragmented Grid

We do not have data for every grid cell of NRW but only a fragmented set of grids containing houses from the available datasets.

Queen distance to nearest school on a 15×15 raster grid



Limitations: Other

- **Listing prices**: The prices we have are not transaction but listing prices.

References I

Black, Sandra E., and Stephen Machin. 2011. "Housing Valuations of School Performance." *Handbook of the Economics of Education* 3: 485–519. https://doi.org/10.1016/B978-0-444-53429-3.00010-7.

Chin, Hoong Chor, and Kok Wai Foong. 2006. "Influence of School Accessibility on Housing Values." *Journal of Urban Planning and Development* 132 (3): 120–29. https://doi.org/10.1061/(ASCE)0733-9488(2006)132:3(120).

Edusei, Kwame Owusu-, Molly Espey, and Huiyan Lin. 2007. "Does Close Count? School Proximity, School Quality, and Residential Property Values." *Journal of Agricultural and Applied Economics* 39 (1): 211–21. https://doi.org/10.1017/S1074070800022859.

References II

Huang, Peng, and Timothy Hess. 2018. "Impact of Distance to School on Housing Price: Evidence from a Quantile Regression." *The Empirical Economics Letters* 17 (April).

Metz, Neil. 2015. "Effect of Distance to Schooling on Home Prices." *Review of Regional Studies* 45: 151–71. https://doi.org/10.52324/001c.8060.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. 2024. "The School Social Index in North Rhine–Westphalia." https://www.schulministerium.nrw/schulsozialindex.

Rosiers, Francois Des, Antonio Lagana, and Marius Theriault. 2001. "Size and Proximity Effects of Primary Schools on Surrounding House Values." *Journal of Property Research* 18 (2): 149–68. https://doi.org/10.1080/09599910110039905.

References III

RWI, and ImmobilienScout24. 2023. "RWI Real Estate Data – Campus File Cross-Section." RWI – Leibniz Institute for Economic Research. https://doi.org/10.7807/immo:red:cross:v4.

Sah, Vivek, Stephen J. Conroy, and Andrew Narwold. 2016. "Estimating School Proximity Effects on Housing Prices: The Importance of Robust Spatial Controls in Hedonic Estimations." *The Journal of Real Estate Finance and Economics* 53 (1): 50–76. https://doi.org/10.1007/s11146-015-9520-5.

Seo, Youngme, and Robert Simons. 2009. "The Effect of School Quality on Residential Sales Price." *Journal of Real Estate Research* 31 (3): 307–28.