Project Outline Presentation
Housing Prices & School Quality in NRW

Daniel Jeffrey Tobien, Yanyi Ji, Luc Wichtmann

Invalid Date

Agenda

Agenda

- Motivation
- Research Question
- Data
- Methods
- Next Steps

Motivation & Research Question

- *Why is this interesting to us?* - We believe that understanding this relationship can offer valuable insights for individuals making decisions about where to live based on local education quality. Die Qualität von Schulen hat einen entscheidenden Einfluss auf den Wert von Immobilien in der Region. In Nordrhein-Westfalen (NRW) untersuchen wir die folgenden Fragen:

- *Why is this important in a broader context?* - *Housing markets* are influenced by various social and economic factors, and school quality is often cited as a key determinant of property values, also in research.[1]

Research Question:
How does the social index of schools influence housing prices, and does the proximity to schools modulate this effect in urban areas?
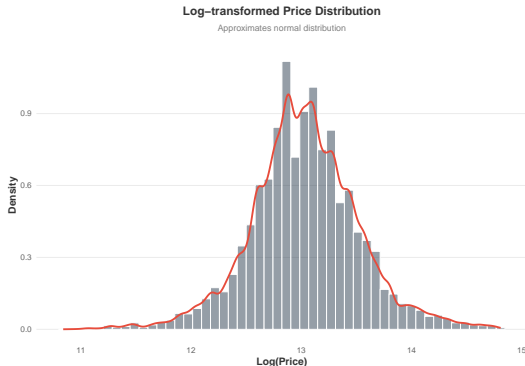
---

[1]Youngme Seo, Robert A. Simons, 2009

Data

- *Housing Data*(Campus File HK 2022): - Property prices (purchase price, price per sqm) - Characteristics: size, rooms, floor space, year built - Location (urban vs rural)

- *School Data*: - Social Index & quality of education - School types (public, private) - Proximity to schools (distance)

- *Regional Data*: - Demographic: income distribution, unemployment rates - Geographic classification (urban vs rural)

Data : Housing Price Transformation

The raw price distribution (left) is highly **right-skewed**, violating the normality assumption required for OLS regression. By applying a **log-transformation** (right), the data approximates a normal distribution, making it suitable for our later analysis model.

Distribution of Purchase Prices
Zoomed in on main market (Top 1% outliers hidden)

Log–transformed Price Distribution
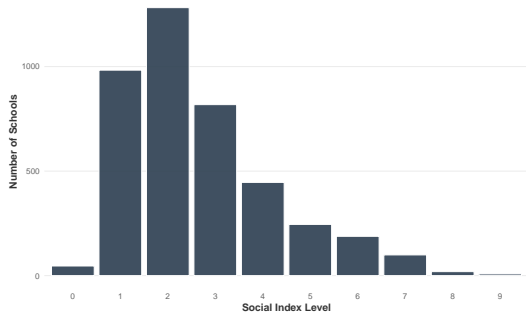Approximates normal distribution

Data : School Quality & Proximity

These two variables directly address our research question regarding the impact of school quality and distance on housing prices.
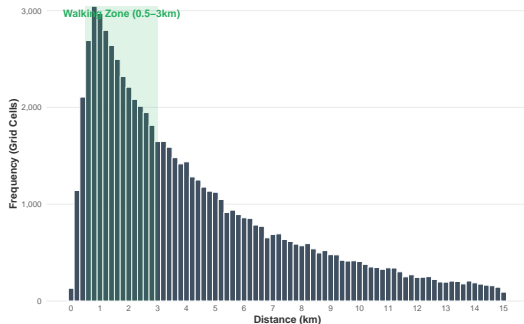

Distribution of School Social Index
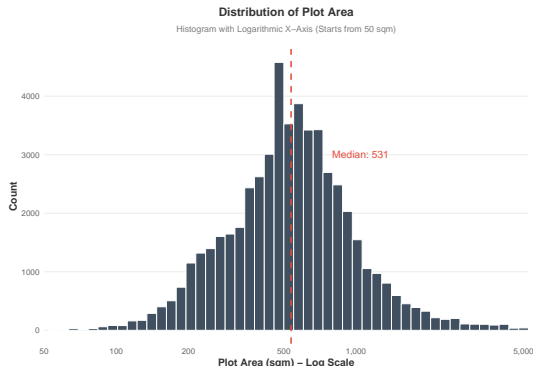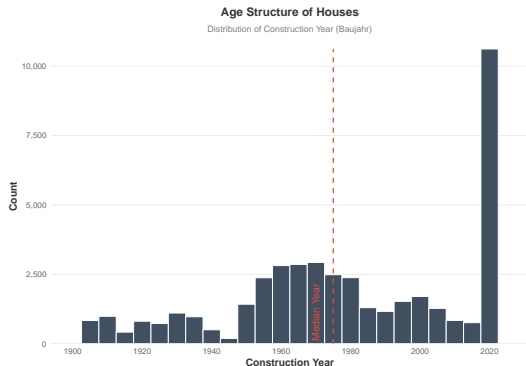1 = Low Challenge, 9 = High Challenge

Distance to Nearest School
Green zone indicates reasonable walking distance
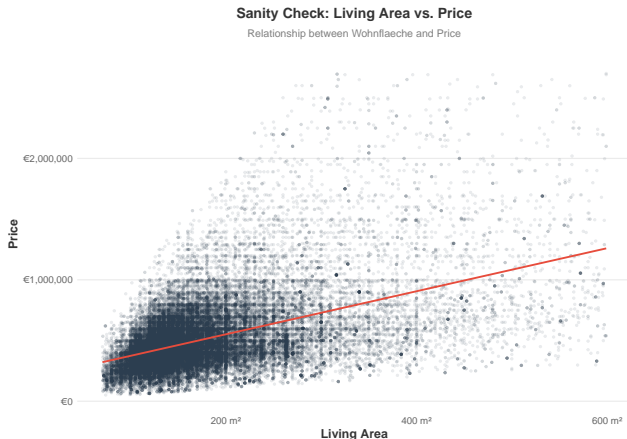
Data : Housing Characteristics

Our sample captures the diversity of the housing stock in NRW. We observe a bimodal
age structure (Post-war reconstruction vs. Modern development) and a log-normally
distributed plot size (Median: 531 sqm).

Age Structure of Houses
Distribution of Construction Year (Baujahr)

Distribution of Plot Area
Histogram with Logarithmic X–Axis (Starts from 50 sqm)

Data Validation

Sanity Check Analysis

- A clear *positive correlation* is observed.
- This confirms economic intuition: *larger houses cost more*.
- The plot helps visually identify *extreme values* for further cleaning.

Sanity Check: Living Area vs. Price
Relationship between Wohnflaeche and Price

Empirical approaches in the literature

- In the paper from Black & Machin, they identified six main approaches to analyze the relationship between school characteristics and housing prices:

1. Regression based estimates

2. Parametric and non-parametric modelling of unobservable factors

3. Instrumental variables approaches

4. Discontinuity methods using administrative boundaries

5. Difference-in-difference, repeat sales and quasi-experimental methods

6. Combined methods

Our Method

- Hedonic Pricing Model: Apply a linear regression model to analyze the impact of the social index of schools on housing prices

- Proximity Effect: Incorporate distance to schools as an additional variable to examine whether proximity strengthens or weakens the effect of the social index on housing prices

- Spatial Model: Utilize a Spatial Lag Model or Spatial Durbin Model to account for spatial dependencies and the influence of neighboring schools and properties on housing prices

- Interaction Effects: Investigate the interaction between the social index and distance to schools to assess how proximity modulates the effect of the social index on housing prices

Shortcomings of our strategy

- A regression-based hedonic model in the housing context might be prone to certain endogeneity problems due to omitted variables.
- A reason for that is the extensive scope of relevant housing characteristics and demographic features

Source: @Machin_2011

Our Next Steps

- Finalize Data Cleaning: Ensure all data is clean, standardized, and ready for analysis.

- Model Application: Apply appropriate regression and spatial models to examine the impact of school quality and proximity on housing prices.

- Result Interpretation: Analyze the model outputs, interpret key findings, and identify trends.

- Prepare Report & Presentation: Summarize results, create visualizations, and prepare findings for presentation to stakeholders.