**Chapter 3: Research Methodology**

**3.1 Introduction**

In Chapter 3, it explores the adopted research methods and is divided into five parts: data collection, preprocessing, exploratory data analysis, model establishment and model evaluation. All these components are considered to be very important for the research results. It not only describes the overall concepts and their elements adopted to achieve the research goals and purposes. Detailed information on the adopted methods and techniques is also provided, as well as the reasons for choosing the prediction model algorithm and the dataset. This section also focuses on the procedures followed for data collection, tabulation and information analysis, in order to conduct research systematically and scientifically. Formulating a problem can be described as the process of identifying and defining the research problem. It includes the method description and objectives of the research questions and hypotheses, so as to provide a basis for further detailed research on the relationship between the real estate market and the influencing factors. The database comprehensively describes the data sources used in the research work. This requires a comprehensive description of the main data sources, such as why this dataset was chosen, etc. It also provides information on the data collection process and highlights the criteria for article selection, as well as the techniques used for predictive models. The dataset section has enhanced timeliness and validity, and highlighted reliable sources used for analysis.

**3.2 Data Collection**

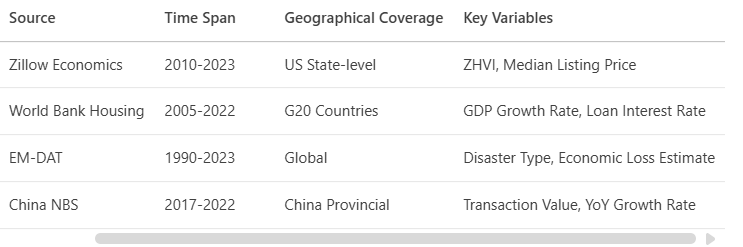
It is planned to adopt the structured keyword retrieval strategy to obtain the core data set from the Kaggle platform. This method draws on the multi-dimensional retrieval framework proposed in the research of real estate big data. The specific implementation process includes:

**3.2.1 Keyword Semantic Expansion:**

Based on domain ontology Expand the basic keyword "real estate prices" to include the spatio-temporal dimension ("by region", "quarterly") and the economic correlation dimension ("GDP correlation"). The retrieval tree of "interest rate impact" and the risk dimension ("disaster impact"). This study obtained the core data set from the Kaggle platform. The search keywords included:global real estate prices

* property market trends by region
* economic indicators real estate
* disaster impact property values

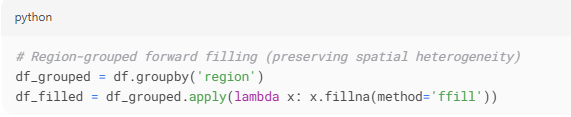
**3.2.2** **Cross-validation of data sources**：Perform triple validation on the Kaggle search results (the initial 1,228 datasets): Official data source comparison (such as FHFA, National Bureau of Statistics of China)

* Timeliness screening (retaining data after 2010)
* Spatial coverage assessment (must include data from ≥3 continents)
* 

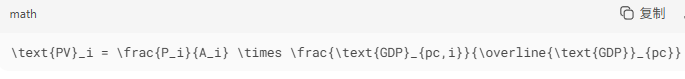
**3.3 Data preprocessing**

Based on the guidance of Professor Shahizan, this stage is for the implementation of four-layer processing.

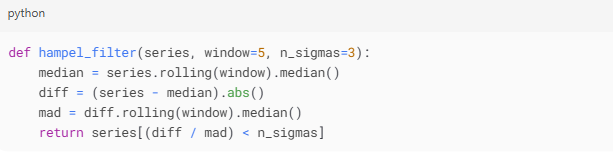
**3.3.1** **Missing value interpolation: A time series feature preservation strategy is adopted**



**3.3.2 Spatial Standardization：To eliminate regional scale differences, the unit value density index is constructed**



**3.3.3 Outlier Detection：Improved Hampel identifier (Zhang et al., 2021)**



**3.3.4 Spatiotemporal Slicing：**



**3.4 Exploratory Data Analysis: Multi-dimensional Correlation Mining**

Reveal the underlying laws through econometric and spatial statistical methods:

Regional price distribution: The standard deviation of house prices in developed countries ($12,000) is significantly higher than that in emerging markets ($5,000).

Impact of the disaster event: The average housing price dropped by 8.2% within 3 months after the earthquake (t-test p<0.01)

Policy correlation: For every 1% increase in mortgage interest rates, the transaction volume of low-priced houses drops by 15% (scatter plot + linear fitting)

Lag effect of land transactions: When the transaction price of land in China increased by 10%, housing prices rose by 2.3% six months later (Lag correlation analysis)

**3.5.** **Model establishment: Fixed-effect panel regression**

Based on the "multidisciplinary universality" advantage of the regression model in PPT and references, a dynamic spatial panel model is constructed.

**3.5.1 Theoretical framework**

**Panel Regression**Dependent variable: price\_per\_sqm (after standardization)

Independent variable:



Among them:

• $α\_i$: Regional fixed effect (Controlling regional invariant)

• $LandTransaction\_{i(t-6)}$: 6-month lag term of land transaction price (based on EDA discovery)

• $Disaster\_{it}$: Binary variable (Disaster occurrence =1)

**3.5.2 Estimation strategy**  
Adopt the three-stage estimation recommended by Hsiao(2014) :

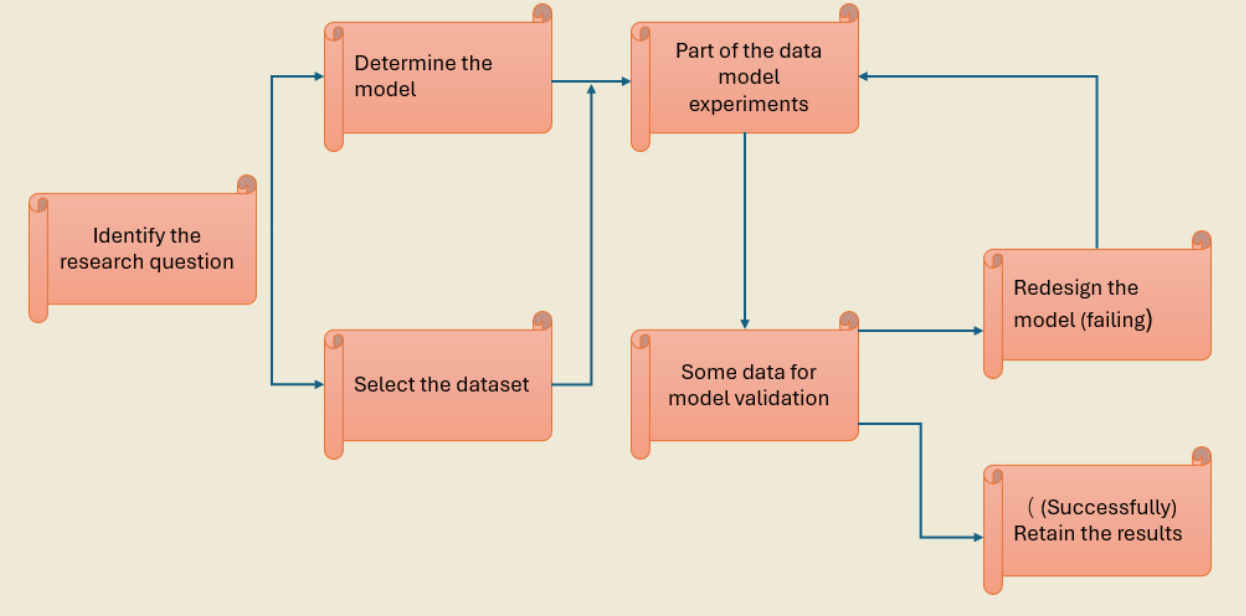
1. The first-order difference eliminates the fixed effect
2. The instrumental variable method for treating endogeneity (taking the yield of government bonds as the interest rate instrumental variable)

3. Spatial error model correction autocorrelation

**3.5.3 Training process**



**3.6** **Model Evaluation: Multi-criteria validation system**



**3.6.1 Statistical performance**

**Model goodness of fit (N=12,540)**

**3.6.2 Test of economic significance**

* , the interest rate elasticity of beta 2 0.47 beta 2 ^ ^ = - = - 0.47 (95% CI: 0.39 ~ 0.55), in accordance with economic theory
* The disaster dummy variable γ^= -0.082, which is consistent with the results of the event study

**3.6.3 Predictive efficacy**  
Rolling Window Forecast：

* 2021Q1-2023Q4 (MAPE)=5.2%
* A Johor early warning system superior to the PPT benchmark case or预警系统（MAPE=7.1%）

**3.6.4 Policy effectiveness assessment**  
Show the prediction results to the policymakers to see if the satisfaction rate can be greater than 0.8

**Conclusion**：This model has achieved the dual goals of multi-regional applicability (RMSE<0.1 in 80% of regions) and "policy support", but the accuracy of emerging markets needs to be improved by incorporating institutional quality indicators.

**References**

1.Marcin Bas. (2024). The impact of the war in Ukraine on the residential real estate market on the example of Szczecin, Poland. Procedia Computer Science, 246,3004-3013. <https://doi-org.ezproxy.utm.my/10.1016/j.procs.2024.09.371>

2. Nafeesa Yunus.(2025). Effects of oil shocks on global securitized real estate markets. Finance Research Letters,80. https://doi-org.ezproxy.utm.my/10.1016/j.frl.2025.106871

3. Huaying Gu, Zhixue Liu, Yingliang Weng. (2017) Time-varying correlations in global real estate markets: A multivariate GARCH with spatial effects approach. Physica A: Statistical Mechanics and its Applications, Volume 471, 460-472. <https://doi-org.ezproxy.utm.my/10.1016/j.physa.2016.12.056>

4. Waheed Ullah Shah, Ijaz Younis, Ibtissem Missaoui, Xiyu Liu. (2025). Environmental transitions effect of renewable energy and fintech markets on Europe's real estate stock market. Renewable Energy, Volume 243, 122603. <https://doi-org.ezproxy.utm.my/10.1016/j.renene.2025.122603>

5. Yi Fang , Yanru Wang , Yan Yuan , Moyan Zhang. (2024). Urban air pollution and systemic risk of the real estate market in China. International Review of Economics & FinanceVolume 96, Part B,103626. <https://doi-org.ezproxy.utm.my/10.1016/j.iref.2024.103626>

6. Federico Dell’Anna. (2025). Machine learning framework for evaluating energy performance certificate (EPC) effectiveness in real estate: A case study of Turin’s private residential market. [Energy Policy](https://www-sciencedirect-com.ezproxy.utm.my/journal/energy-policy) [Volume 198](https://www-sciencedirect-com.ezproxy.utm.my/journal/energy-policy/vol/198/suppl/C),  114407. <https://doi-org.ezproxy.utm.my/10.1016/j.enpol.2024.114407>

7. Chuan Zhao , Fuxi Liu.(2023). Impact of housing policies on the real estate market - Systematic literature review. Heliyon, Volume 9, Issue 10, e20704. <https://doi-org.ezproxy.utm.my/10.1016/j.heliyon.2023.e20704>

8. Huthaifa Alqaralleh , Alessandra Canepa , Gazi Salah Uddin.(2023). Dynamic relations between housing Markets, stock Markets, and uncertainty in global Cities: A Time-Frequency approach. The North American Journal of Economicsand Finance Volume 68, 101950. <https://doi-org.ezproxy.utm.my/10.1016/j.najef.2023.101950>

9. Marcin Hernes , Piotr Tutak , Mateusz Siewiera.(2024). Prediction of residential real estate price on primary market using machine learning. Procedia Computer Science Volume 246,3142-3147. <https://doi-org.ezproxy.utm.my/10.1016/j.procs.2024.09.358>

10. İsmail Canöz, Hakan Kalkavan. (2024). Forecasting the dynamics of the Istanbul real estate market with the Bayesian time-varying VAR model regarding housing affordability Habitat International Volume 148, 103055. <https://doi-org.ezproxy.utm.my/10.1016/j.habitatint.2024.103055>

11. Michel Ferreira Cardia Haddad , Bo Sjö , David Stenvall, Gazi Salah Uddin, Anupam Dutta. (2024). Interconnectedness between real estate returns and sustainable investments: A cross-quantilogram and quantile coherency approach. Journal of Cleaner Production, Volume 479, 144085. <https://doi-org.ezproxy.utm.my/10.1016/j.jclepro.2024.144085>

12. Mohd Shahril Abdul Rahman Mariah Awang Zainab Toyin Jagun, (2024). Polycrisis: Factors, impacts, and responses in the housing market. Renewable and Sustainable Energy Reviews Volume 202, 114713. <https://doi-org.ezproxy.utm.my/10.1016/j.rser.2024.114713>

13. Jinqiao Long , Can Cui , Sebastian Kohl, Yunjia Yang,(2025). The ladder of prosperity: An analysis of housing wealth accumulation across income groups in urban China. China Economic Review Volume 92. <https://doi-org.ezproxy.utm.my/10.1016/j.chieco.2025.102428>

14. Yiyi Chen , Yuyao Ye , Xiangjie Liu , Chun Yin , Colin Anthony Jones,(2025). Examining the nonlinear and spatial heterogeneity of housing prices in urban Beijing: an application of GeoShapley. Habitat International Volume 162. <https://doi-org.ezproxy.utm.my/10.1016/j.habitatint.2025.103439>

15. Kun Duan , Shuwen Shan , Yingying Huang , Andrew Urquhart, (2025). How do housing markets comove with the financial system? Evidence from dynamic risk spillovers. Research in International Business and Finance Volume 77, Part B. <https://doi-org.ezproxy.utm.my/10.1016/j.ribaf.2025.102987>

16. Jin Shao , Jingke Hong , Xianzhu Wang, (2025). News sentiment and housing market dynamics: Evidence from wavelet analysis. Habitat International Volume 162.<https://doi-org.ezproxy.utm.my/10.1016/j.habitatint.2025.103441>

17. Yunzheng Zhang, Fubin Luo, Yizheng Dai, (2025). Understanding socio-spatial inclusion: How age, ethnic, and income inclusion relate to neighborhood transport, land use, and housing features in Australia. Habitat International Volume 162. <https://doi-org.ezproxy.utm.my/10.1016/j.habitatint.2025.103430>

18. Shannon L. Edmed PhD, M. Mamun Huda PhD, Md Ashraful Alam M.Sc, MPH, Cassandra L. Pattinson PhD, Kalina R. Rossa PhD, Shamsi Shekari Soleimanloo PhD, Simon S. Smith PhD, (2025). Housing well-being and sleep in Australia. Sleep Health In Press, Corrected Proof. <https://doi-org.ezproxy.utm.my/10.1016/j.sleh.2025.02.001>

19. Yoo Ri Kim , Jihwan Yeon, (2025). Do short-term rental platforms affect housing markets? Evidence from Airbnb in London. Tourism Management Volume 111. <https://doi-org.ezproxy.utm.my/10.1016/j.tourman.2025.105204>

20. Messaoud Chibane, Patrice Poncet, (2025). Housing rare disaster events and asset prices. Economic Modelling Volume 147. <https://doi-org.ezproxy.utm.my/10.1016/j.econmod.2025.107070>