Neural Network adjusted Spatial Dynamic Factor Models for Real Estate Valuation

Koen Zomerdijk

A thesis submitted to the Delft University of Technology in
partial fulfillment of the requirements for the degree of Master
of Science in Applied Mathematics.

Koen Zomerdijk: *Neural Network adjusted Spatial Dynamic Factor Models for Real Estate Valuation* (2023)

The work in this thesis was carried out at:

Faculty EEMCS
Delft University of Technology

Student number:	4712552
Project duration:	November 1, 2022 - July 31, 2023
To be defended on:	September 5, 2023 at 10:00

Daily supervisor:	Dr. Nestor Parolya, TU Delft
Company supervisor:	Prof. dr. Marc Francke, Ortec Finance
Responsible supervisor:	Dr. Dorota Kurowicka, TU Delft

Abstract

This thesis concerns modeling residential real estate selling prices in a hedonic price model framework on a small spatial-temporal granularity. The research addresses the challenge of sparse spatial-temporal real estate data, i.e. many combinations of location and time with few or no transactions, by employing spatial dynamic factor models (SDFMs). Two types of SDFMs are employed: an SDFM with a 1D spatial structure based on the spatial random walk and an SDFM with a 2D spatial structure based on the Gaussian random field. To capture the information on the property characteristics, spatial dynamic factor models are combined with two different data-driven models, namely a neural network (NN) and an interpretable version of an NN, the local generalized linear model network (LGLMN). Both a Bayesian approach and an algorithmic approach are employed to estimate the models on both a PC and a high-performance computer (HPC). A simulation study is conducted to demonstrate the ability of an NN to capture linear and non-linear structures when combined with an SDFM and to show the ability of the LGLMN to replicate a linear structure. Furthermore, the models are evaluated on real transaction data from the municipality of Rotterdam. The findings demonstrate that the algorithmically estimated NN-adjusted SDFM based on the spatial random walk (NN-SRW-DFM) outperforms the other models in terms of accuracy with an out-of-sample MAPE of 0.128. Moreover, the results highlight a trade-off between accuracy, speed, and interpretability.

Preface

This thesis has been submitted for the degree Master of Science in Applied Mathematics at Delft University of Technology. The research has been conducted at Ortec Finance and concerns real estate valuation using dynamic factor models and neural networks on spatial-temporal sparse transaction data. To model real estate selling prices a framework is used where a spatial dynamic factor model is combined with a feed-forward neural network. Results show that the feed-forward neural network is able to replicate linear and non-linear structures in this framework and that a neural network in combination with a spatial dynamic factor model outperforms more structured models in combination with spatial dynamic factor models. This research is conducted under the supervision of Prof. dr. Marc Francke (Ortec Finance) and dr. Nestor Parolya (TU Delft). The responsible supervisor is dr. Dorota Kurowicka.

I want to thank Nestor Parolya for his willingness to always help me during the process, his good suggestions, and for sharing his statistical expertise. Furthermore, I want to thank Marc Francke as he was always willing to help me during the project, for his good ideas, and for sharing his wide knowledge in the field of real estate valuation. I want to thank Ortec Finance for giving me the opportunity and resources to perform my thesis there. Additionally, I want to thank Dorota Kurowicka for taking the time of reading this report and for taking place on my thesis committee.

Contents

1. Introduction **1**
1.1. Thesis Outline 3

2. Literature **5**
2.1. Hedonic Price Model 5
2.2. Artificial Neural Networks in Real Estate Valuation 6
2.3. Spatial-temporal Effects 6

3. Theoretical Background **9**
3.1. Fully Connected Feed-Forward Neural Network (FFNN) 9
3.2. Local Generalized Linear Model Network (LGLMN) 11
3.2.1. Interpretability 11

4. Methodology **15**
4.1. Spatial Dynamic Factor Model 15
4.1.1. Spatial Random Walk 16
4.1.2. Gaussian Random Field 17
4.2. Property Characteristics Component 18
4.2.1. Linear Component 18
4.2.2. Non-linear Component 18
4.2.3. Neural Network Component 19
4.2.4. Local Generalized Linear Model Network Component 19
4.3. Model Representation 20
4.3.1. Model Extensions 20
4.4. Model Estimation 21
4.4.1. Bayesian Estimation 21
4.4.2. Algorithmic Estimation 22
4.5. Assessment of Performance 23
4.5.1. Prediction Accuracy Measures 23
4.5.2. Validation Methods 24
4.5.3. Convergence of Bayesian Estimation 25

5. Data **27**
5.1. Data Preparation 27
5.2. Data Analysis 27

6. Simulation Study: Ability of the Neural Network in Capturing Linear and Non-linear Structures **31**
6.1. Data Generating Process 31
6.2. Assessment 32
6.3. Results 32
6.3.1. Linear Structure 33

    6.3.2. Nonlinear Structure . . . 34
  6.4. Replication of linear structure by the LGLMN . . . 36

7. Results **41**
  7.1. Performance Results . . . 41
  7.2. Interpretation of the LGLMN Results . . . 43
    7.2.1. LGLMN-SRW-DFM . . . 43
    7.2.2. LGLMN-GRF-DFM . . . 46
    7.2.3. Remarks on Interpretability of LGLMN . . . 49

8. Discussion **51**
  8.1. Conclusion . . . 51
  8.2. Limitations and Future Research . . . 52

A. Appendix **57**
  A.1. Linear Splines . . . 57
  A.2. Building types . . . 58
  A.3. Quantitative variables per building type . . . 59
  A.4. Multimodality of factors and loadings . . . 60
  A.5. Estimated Coefficients Simulation Study . . . 64
  A.6. Experiments on Number of Factors and Building types . . . 67
  A.7. Hyper-parameters . . . 68
  A.8. Result for variations in spatial geographies . . . 69
  A.9. Estimates of property characteristic component parameters . . . 70

List of Figures

3.1. Schematic overview of a fully Connected feed-forward neural network with an input dimension equal to 8, 3 hidden layers with dimension 6, and an output layer with dimension 1. 10
3.2. Schematic overview of a local generalized linear model network with an input dimension equal to 8 and two hidden layers with dimension 6. 13

5.1. Correlation matrix of numerical independent variables. 28
5.2. Observations of the building types. 29
5.3. Observations of the building year classes. 30
5.4. Observations over time in quarters. 30

6.1. Feature contributions of the quantitative variables in the LGLMN-SRW-DFM based on data generated with the L-SRW-DFM. 38
6.2. Attentions of the categorical variables in the LGLMN-SRW-DFM based on data generated with the L-SRW-DFM. 39
6.3. Interaction strength of lot size over different values of house size in the LGLMN-SRW-DFM on data generated with the L-SRW-DFM. 40

7.1. Feature contributions of the quantitative variables in the LGLMN-SRW-DFM. . 44
7.2. Feature contributions of the categorical variables in the LGLMN-SRW-DFM. . 45
7.3. Attentions of the categorical variables in the LGLMN-SRW-DFM. 46
7.4. Feature contributions of the quantitative variables in the LGLMN-GRF-DFM. . 47
7.5. Feature contributions of the categorical variables in the LGLMN-GRF-DFM. . 48

A.1. Factors of the base model: L-SRW-DFM and of the L-SRW-DFM and NN-SRW-DFM estimated on data simulated from the base model. 60
A.2. Loadings of the base model: TSP SDFM with linear components and of the models estimated on data simulated from the base model. 61
A.3. Factors of the base model: TSP SDFM with nonlinear components and of the models estimated on data simulated from the base model. 62
A.4. Loadings of the base model: TSP SDFM with nonlinear components and of the models estimated on data simulated from the base model. 63

List of Tables

4.1. Schematic representation of property characteristics components and the spatial-temporal components of the models. 20

5.1. Summary statistics of quantitative variables. 27
5.2. Summary statistics of the linear splines variables 28
5.3. Observations per neighborhood 29

6.1. Estimated variance of the models simulated by the DGP. 32
6.2. In-sample and out-of-sample performance measures of the models on data simulated from the L-SRW-DFM. 33
6.3. In-sample and out-of-sample distributions of the differences of the true property characteristics components and the property characteristics components estimated from the L-SRW-DFM and NN-SRW-DFM on the artificial data. . . . 33
6.4. In-sample and out-of-sample distributions of the differences of the true spatial-temporal components and the spatial-temporal components estimated from the L-SRW-DFM and NN-SRW-DFM on the artificial data. 34
6.5. In-sample and out-of-sample performance measures of the models on data simulated from the NL-SRW-DFM. 34
6.6. In-sample and out-of-sample distributions of the differences of the true property characteristics components and the property characteristics components estimated from the NL-SRW-DFM and NN-SRW-DFM on the artificial data. . 35
6.7. In-sample and out-of-sample distributions of the differences of the true spatial-temporal components and the spatial-temporal components estimated from the NL-SRW-DFM and NN-SRW-DFM on the artificial data. 35
6.8. In-sample and out-of-sample performance measures of the models on data simulated from the L-SRW-DFM. 36

7.1. In-sample performance measures of the models on data from the municipality of Rotterdam. 42
7.2. Out-of-sample performance measures of the models on data from the municipality of Rotterdam. 42

A.1. Primary and secondary building types in the Rotterdam data set. 58
A.2. Summary statistics of quantitative variables of building type Apartment 59
A.3. Summary statistics of quantitative variables of building type Semi-detached . 59
A.4. Summary statistics of quantitative variables of building type Terraced 59
A.5. Summary statistics of quantitative variables of building type Corner house . . 59
A.6. Summary statistics of quantitative variables of building type Detached 59
A.7. Estimates of the coefficients of the property characteristics component for the true L-SRW-DFM and the L-SRW-DFM estimated on the data generated from the DGP. 64

A.8. Estimates of the coefficients of the property characteristics component for the true NL-SRW-DFM and the NL-SRW-DFM estimated on the data generated from the DGP. 65
A.9. Estimate of the coefficients of the true L-SRW-DFM and the L-SRW-DFM estimated on the data generated from the true L-SRW-DFM used for the experiment of the LGLMN-SRW-DFM. 66
A.10.CV values for the SRW-DFlin when using the 5 primary building types and 16 secondary building types. 67
A.11.CV values on an experiment with 5-fold CV to see how many dynamic factors to use for the SRW-DFlin. The model is estimated for a different number of factors $K = 1,...,6$. 67
A.12.Hyper-parameters of the Bayesian estimation procedure. 68
A.13.Settings of the No-U-Turn Sampler (NUTS)-algorithm. 68
A.14.Hyper-parameters of the iterative estimation procedure. 68
A.15.In-sample and out-of-sample performance measures of the models with different spatial geographies on data from the municipality of Rotterdam. 69
A.16.Estimation of the time-invariant parameters of the L-SRW-DFM 70
A.17.Estimation of the time-invariant parameters of the NL-SRW-DFM 71
A.18.Estimation of the time-invariant parameters of the -LGRF-DFM 71
A.19.Estimation of the time-invariant parameters of the NL-GRF-DFM 72

List of Algorithms

1.
Iterative algorithm to estimate models with an FFNN or LGLMN23

Acronyms

ANN	artificial neural network	2
CV	cross-validation	24
DFM	dynamic factor model	1
DGP	data generating process	31
FFNN	fully connected feed-forward neural network	3
GLM	generalized linear model	11
GRF	Gaussian random field	15
HPC	high performance computer	21
HPM	hedonic price model	5
HTM	hierarchical trend model	5
LGLMN	local generalized linear model network	2
MAPE	mean absolute percentage error	23
MCMC	Markov chain Monte Carlo	6
ML	machine learning	6
MSE	mean squared error	24
NUTS	No-U-Turn Sampler	xiv
RMSE	root mean squared error	23
RSM	repeat sales model	5
SAR	spatial auto-regression	6
SDFM	spatial dynamic factor model	7
SGD	stochastic gradient descent	22
SRW	spatial random walk	15
TSP	travelling sales person	16

Acronyms of the Models

L-GRF-DFM	Linear adjusted GRF-DFM	20
LGLMN-GRF-DFM	LGLMN adjusted GRF-DFM	20
LGLMN-SRW-DFM	LGLMN adjusted SRW-DFM	20
L-SRW-DFM	Linear adjusted SRW-DFM	20
NL-GRF-DFM	Non-linear adjusted GRF-DFM	20
NL-SRW-DFM	Nonlinear adjusted SRW-DFM	20
NN-GRF-DFM	NN adjusted GRF-DFM	20
NN-SRW-DFM	NN adjusted SRW-DFM	20

1. Introduction

This thesis concerns the modeling of residential real estate selling prices. A good prediction of a house price is highly relevant, as the value of a real estate property can be used for taxation (wealth tax depends on the value of your house), finance (obtaining mortgages), and investment decisions (e.g. whether to rent out or sell a house).

Various characteristics affect the value of residential real estate, e.g. house size, lot size, building type, and building year (Firstenberg et al. [1988] ; Sirmans et al. [2006] ; Bourassa et al. [2009]). In addition to these property characteristics, the value of a house depends on location and time. Therefore, incorporating spatial-temporal effects in the real estate valuation model is essential.

Assume we model the spatial-temporal effect $g_{m,t}$ of a property on location m and at time t. One way to model this effect is by the sum of a random effect for location ψ_m and a random effect for time χ_t, i.e.

$$g_{m,t} = \psi_m + \chi_t. \tag{1.1}$$

However, using this approach, the time and location of observations are incorporated separately, and not a combination of the two is incorporated in the model. Therefore, no information on price indexes is taken into account. This is undesirable as the price dynamics of real estate differ across locations (Geltner et al. [2001]). Hence, modeling the spatial-temporal effect by (1.1) is unsuitable.

To incorporate price dynamics across locations, a trend effect can be used for every location involved in the model, i.e. having an effect, $\omega_{m,t}$, for every combination of location m and time t:

$$g_{m,t} = \omega_{m,t}. \tag{1.2}$$

When considering a larger number of M locations and T time steps, $M \times T$ parameters need to be estimated only for the spatial-temporal components. Besides computational issues of estimating a large number of parameters, there arises a problem with very sparse spatial-temporal data of real estate transactions when using an effect for every location-time combination. In general, there is a lot of data available on real estate transactions. However, when considering a small spatial-temporal granularity, e.g. transactions in a specific month in a specific neighborhood, the available data is very limited, as there might be a lot of location-time combinations with no or just a few observations. This concept of spatial-temporal sparsity of the data is a common point of attention in research in the field of real estate valuation (Francke et al. [2022]; Ren et al. [2017]; Silver and Graf [2014]; Peng et al. [2021]; Geltner [2015]; Schwann [1998]). Due to the spatial-temporal sparsity of the data, it is complex to include spatial-temporal information in a real estate valuation model.

Inspired by Francke et al. [2022], this thesis uses a dynamic factor model (DFM) to capture the spatial-temporal information of a real estate property. DFMs are introduced by Geweke

[1977]. The idea behind a DFM is that a few factors can explain a large fraction of the variance of many macroeconomic variables. Consider $K << M$ dynamic factors. Then $g_{m,t}$ can be modeled using a dynamic factor model by

$$g_{m,t} = \mathbf{\Gamma}_m \boldsymbol{F}_t, \tag{1.3}$$

where Γ $(M \times K)$ are the loadings and F $(K \times T)$ are the factors. In this case, only $M \times K + K \times T$ parameters need to be estimated. Since $K << M$, this number of parameters is considerably less than the $M \times T$ parameters in (1.2).

Usually, DFMs are used to explain a large fraction of time series into a small number of factors, i.e. latent trends. That being the case, the objective is to determine the latent trends F. In this research, the goal of the DFM differs from the typical situation. That is, DFMs are used to summarize neighborhood sub-trends in a small number of dynamic factors. However, we are not necessarily interested in the latent trends F, but in the combinations of the latent trends and the loadings, which constitute the sub-trends of neighborhoods. Furthermore, in this thesis, two types of DFMs are used that incorporate spatial correlation between the sub-trends of the neighborhoods.

Apart from the spatial-temporal information, property characteristics affect the price of a residential real estate property. The effect of property characteristics on the value of the property might not be straightforward. In advance, it is sometimes unknown in what structure these characteristics affect the house value and whether these characteristics have an interaction effect on the value of the house. Furthermore, the effects of the property characteristics on the property value might depend highly on the used case. Therefore, a model with a data-driven structure is preferred over a model representation with a fixed structure. This research incorporates this need for flexibility by using an artificial neural network (ANN). A drawback of an ANN is that it lacks interpretability since it acts like a black box on the in and outflow of information. This research uses the local generalized linear model network (LGLMN) introduced by Richman and Wüthrich [2023] to incorporate the need for interpretability in the structure of the ANN.

This thesis provides a solution to the problem of spatial-temporal sparse real estate data by incorporating spatial-temporal dependencies and including property characteristics in a data-driven way. This, while keeping mind in can be generalized for different applications and that categorical effects of the neighborhood can be captured in a data-driven way. This is done by introducing models that combine an ANN with for each a different type of DFM to predict selling prices of real estate.

This research has several contributions to the existing research on modeling real estate prices on spatial-temporal sparse data. A simulation study is performed in which data is generated from a model with a linear structure and a model with a nonlinear structure for property characteristics components. It is demonstrated that a neural network is capable of mimicking linear and non-linear structures when combining it with a spatial dynamic factor model. Results of the simulation study show that the property characteristics components and the dynamic factors are replicated when using a neural network on the generated data. Furthermore, the simulation study shows that the LGLMN is capable of replicating the regression coefficients with its ANN structure. Another contribution of this research is that spatial dynamic factor models are used in a hedonic price model framework to predict real estate transaction prices. Additionally, a spatial dynamic factor model is used in combination with a neural network, to obtain a data-driven model for the real estate properties while using a spatial-temporal functionality from an econometric model. Lastly, it uses a 2-dimensional spatial dynamic factor model while past research on real estate prices only considered a 1-dimensional spatial dynamic factor model (Francke et al. [2022]).

1.1. Thesis Outline

This thesis starts with a review of the literature regarding real estate valuation in Chapter 2. Subsequently, a background on the fully connected feed-forward neural network (FFNN) and the local generalized linear model network (LGLMN) is given in Chapter 3. Chapter 4 elucidates the models used in this research. Furthermore, a description is given of the estimation methods and assessment measures of the models. Chapter 6 covers an experiment about the ability of an ANN to capture linear and non-linear structures on simulated data and an experiment on the LGLMN to analyze whether it can replicate regression coefficients from a linear model. A case study on data from the municipality of Rotterdam is performed for the assessment of the models. Chapter 5 gives a description of the data preparation and gives an analysis of this data. The results of the models on the data from the municipality of Rotterdam are outlined in Chapter 7. Finally, the conclusions are discussed in Chapter 8. This chapter also highlights the limitations of the research and gives recommendations for future research.

2. Literature

This chapter reviews some of the important literature about real estate valuation methods. This thesis uses a hedonic price model (HPM) framework, which will be described in Section 2.1, on cross-sectional data including information on space and time. Section 2.2 covers the literature regarding the use of ANNs in real estate valuation. In Section 2.3 the literature on spatial-temporal effects in real estate valuation will be analyzed.

2.1. Hedonic Price Model

The hedonic price model (HPM) is a major approach in modeling real estate transaction prices. The HPM is derived from the consumer theory of Lancaster [1966]. The theoretical framework of this model is provided by Rosen [1974]. The HPM is a valuation model that quantifies the contribution of factors to the value of a property (Malpezzi et al. [2003]). The HPM is a frequently used methodology in modeling house selling prices (Herath and Maier [2010]). In real estate valuation literature, multiple versions of the HPM are used to investigate the way how certain attributes (e.g. location, structural attributes, and neighborhood attributes) affect real estate selling prices (Chau and Chin [2003]). A disadvantage of the usage of a HPM is that without prior knowledge about the effect of variables on house selling prices, it is difficult to determine a functional form.

Another approach in real estate valuation is the repeat sales model (RSM) (Bailey et al. [1963]). In an RSM the house transaction price is not modeled by its characteristics but only by using transactions that occurred before. An advantage of using a repeat sales framework is that there is no need to have data about property characteristics. A drawback of this methodology is that only transactions of properties that are sold more than once can be used. Another disadvantage is that changes in property characteristics are unaccounted for in the model. Furthermore, a repeat sales model can not accommodate changing attribute prices over time. This research considers data with property characteristics and therefore, this research is based on the methodology of an HPM.

An expansion of a HPM is the hierarchical trend model (HTM), introduced by Francke and Vos [2004]. The HTM is a state-space model that is estimated using a Kalman filter. The HTM contains spatial and temporal dependencies of the selling prices and the housing characteristics are incorporated in the model with a nonlinear structure. The HTM contains a general trend and cluster trend for the districts and the house types. The model's statistical framework allows for the testing of model assumptions and the comparison of competing models through likelihood analysis. However, a drawback is the need to make distributional assumptions on both the parameters and structure.

2.2. Artificial Neural Networks in Real Estate Valuation

An artificial neural network (ANN) is a machine learning (ML) technique constructed to function as a neural network in a human's brain. ANNs have the capacity to learn complex functions from input to output. An ANN is able to detect nonlinear relationships and interactions among variables (Goh [1995]). A drawback of using ANNs is the interpretability, i.e. the capacity of a human to understand and reason the model (Fan et al. [2020]).

There are various applications of ANNs for the appraisal of real estate. Peterson and Flanagan [2009] demonstrated that the ANN has a greater out-of-sample pricing precision than linear HPMs. Likewise, Tay and Ho [1992] found that the mean absolute error of the ANN model was lower than the mean absolute error of a multiple regression model when predicting residential apartment prices in Singapore. In contrast, Worzala et al. [1995] did not find results that show that ANNs perform better than multiple regression models.

Using dummy variables for the spatial information of the properties may result in many dummies. Therefore, spatial information of the property is typically not explicitly (i.e. using the coordinates of the property) used in studies that use an ANN for the valuation of real estate. Chiarazzo et al. [2014] implicitly uses spatial information of the property by using statistics of the neighborhood, including environmental quality.

2.3. Spatial-temporal Effects

Literature on real estate valuation accounts for location and time effects on the real estate property in various ways.

Ripley [2005] introduced spatial auto-regression (SAR) models which incorporated spatial information in an auto-regressive model. This is used in various applications on real estate valuation (Haider and Miller [2000] ; Fan et al. [2005] ; Bottero et al. [2017] ; Bidanset and Lombard [2014]).

Hui et al. [2010] models the log selling price of properties in Hong Kong by including distinct model components for the time, the block unit, and the floor. The time effect is modeled with a random walk with drift. The block-unit effect is driven by intrinsic features of the property unit and the floor effect is modeled as a linear function of the floor level. The hierarchical Bayesian specification of the model is sampled using a Gibbs sampler (Casella and George [1992]) with weakly informative normal priors.

Brunauer et al. [2013] uses a multi-level version of structured additive regression models to regress house prices on individual attributes and locational characteristics in a four-level hierarchical model. Due to its level structure, nonlinear covariate effects can be incorporated at each level of the hierarchy. The hierarchical structure of the spatial effect of this model allows for improved predictive quality when spatial units are missing because the model borrows information from other spatial levels when information from one spatial level is missing. The model is estimated using a Markov chain Monte Carlo (MCMC) simulation technique that takes advantage of its hierarchical structure.

To account for the scarcity of transactions at a fine spatial-temporal granularity, Ren et al. [2017] introduced a factor model based on a Bayesian non-parametric approach to cluster correlated sub-regions. This approach has a data-driven structure that learns to correlate sub-regions that share similar underlying price dynamics. This method provided reliable monthly housing indices at the census tract (a sub-region smaller than a neighborhood) level.

Calainho et al. [2022] utilizes different ML methods to produce property price indices. The results indicated that the prediction accuracy of the ML algorithm is higher than for linear models. However, they found that the results of the ML models were unstable for small samples of data and that they may exhibit estimation bias.

Cafarella et al. [2023] uses ML for the estimation of price indices for different products. A neural network is used in combination with a hedonic econometric framework to obtain the price indices of the goods.

This research is closely related to Francke et al. [2022], in which condominium selling prices are modeled in a RSM framework with a spatial dynamic factor model (SDFM). In this SDFM, a spatial structure is appointed to the factor loadings. The SDFM from Francke et al. [2022] will be used in this research as well and will be described more extensively in Section 4.1.1.

Other related research is from Francke and Van de Minne [2023], in which a random effects model is combined with different machine learning approaches via an iterative process to predict selling prices of commercial real estate. The random effects part of the model includes among other things spatial and temporal dependencies, while the ML approach includes information on covariates. A Besag model (Besag [1974]) is used to include spatial property effects. Using this Besag model, the spatial effect for each property depends on its (near) neighbors. The model is estimated using a neural network in a Bayesian framework and in an iterative way using various ML methods. This iterative approach is faster than estimating the ANN with MCMC simulation and is more flexible as multiple ML methods can be used.

3. Theoretical Background

This chapter describes the theory of models that will be the basis for modeling the property characteristics in the real estate valuation models that will be proposed in Chapter 4. First, the FFNN will be elucidated. Subsequently, the LGLMN, which is an interpretable extension of this FFNN, will be presented.

3.1. Fully Connected Feed-Forward Neural Network (FFNN)

A fully connected feed-forward neural network (FFNN) is a type of ANN in which neurons are organized into layers, and each neuron in a given layer is connected to every neuron in the subsequent layer. The FFNN starts with the neurons in an input layer, which contains the independent variables of the regression model. With the variables from this input layer, the neurons in the next layer will be constructed by giving weights to neurons in the input layer and applying a non-linear activation function to the linear combination of the weights and the values of the neurons of the input layers. This process will continue in this way until the output layer, which contains the dependent variable, is reached.

The mathematical framework of an FFNN is defined as follows. Let $\phi_m : \mathbb{R} \to \mathbb{R}$ a nonlinear activation function. In this thesis, the sigmoid function is used as an activation function. For every layer m choose a number q_m to be the number of neurons in that layer.

The m-th layer of an FFNN is defined by the mapping

$$h^{(m)} : \mathbb{R}^{q_{m-1}} \to \mathbb{R}^{q_m}, \tag{3.1}$$

$$\mathbf{x} \mapsto h^{(m)}(\mathbf{x}) = \left(h_1^{(m)}(\mathbf{x}), \ldots, h_{q_m}^{(m)}(\mathbf{x})\right)^{\top}, \tag{3.2}$$

having neurons $h_j^{(m)}(\mathbf{x})$,$1 \leq j \leq q_m$, for $\mathbf{x} = (x_1, \ldots, x_{q_{m-1}})^{\top} \in \mathbb{R}^{q_{m-1}}$,

$$h_j^{(m)}(\mathbf{x}) = \phi_m\left(\omega_{0,j}^{(m)} + \langle \boldsymbol{\omega}_j^{m}, \mathbf{x}\rangle\right) = \phi_m\left(\omega_{0,j}^{(m)} + \sum_{l=1}^{q_{m-1}} \omega_{l,j}^{(m)} x_l\right),$$

for given network weights $\boldsymbol{\omega}_j^{(m)} = (w_{l,j}^{(m)})^{\top}_{1 \leq l \leq q_{m-1}} \in \mathbb{R}^{q_{m-1}}$ and bias $\omega_{0,j}^{(m)} \in \mathbb{R}$.

An FFNN of depth $d \in \mathbb{N}$ is obtained by composing d FFNN layers (3.1) to provide a deep learned representation,

$$\mathbf{h}^{(d:1)} : \mathbb{R}^{q_0} \to \mathbb{R}^{q_d}, \tag{3.3}$$

$$\mathbf{x} \mapsto \mathbf{h}^{(d:1)}(\mathbf{x}) = \left(\mathbf{h}^{(d)} \circ \ldots \circ \mathbf{h}^{(1)}\right)(\mathbf{x}). \tag{3.4}$$

Subsequently, the FFNN regression function $f(\mathbf{x})$ is given by

$$f(\mathbf{x}) = \beta_0 + \langle \boldsymbol{\beta}, \mathbf{h}^{(d:1)}(\mathbf{x}) \rangle, \tag{3.5}$$

with regression output parameter $\boldsymbol{\beta} = (\beta_1, ..., \beta_{q_d})^T \in \mathbb{R}^{q_d}$ and bias $\beta_0 \in \mathbb{R}$.

A schematic overview of an FFNN is shown in Figure 3.1.

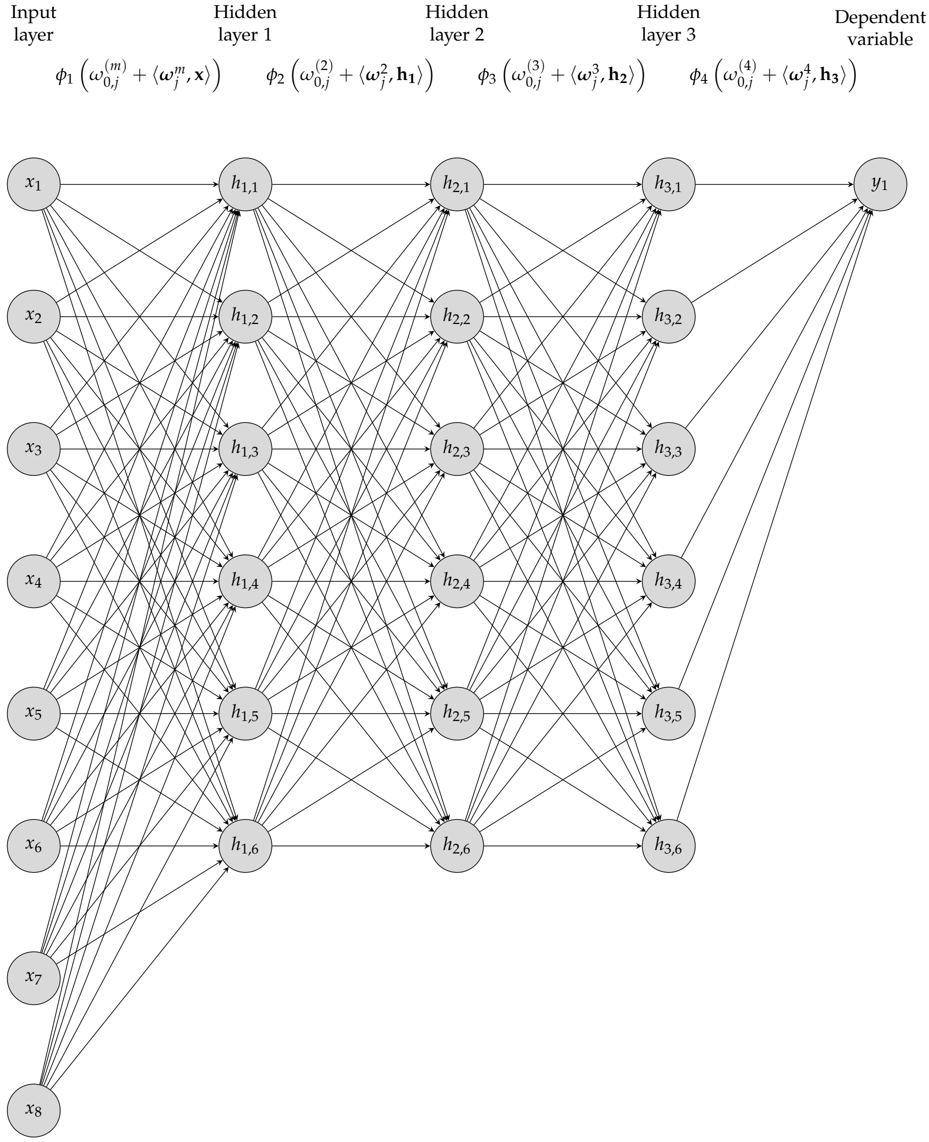

Input layer
Hidden layer 1
Hidden layer 2
Hidden layer 3
Dependent variable
$\phi_1 \left(\omega_{0,j}^{(m)} + \langle \boldsymbol{\omega}_j^m, \mathbf{x} \rangle \right)$
$\phi_2 \left(\omega_{0,j}^{(2)} + \langle \boldsymbol{\omega}_j^2, \mathbf{h_1} \rangle \right)$
$\phi_3 \left(\omega_{0,j}^{(3)} + \langle \boldsymbol{\omega}_j^3, \mathbf{h_2} \rangle \right)$
$\phi_4 \left(\omega_{0,j}^{(4)} + \langle \boldsymbol{\omega}_j^4, \mathbf{h_3} \rangle \right)$
x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
$h_{1,1}$
$h_{1,2}$
$h_{1,3}$
$h_{1,4}$
$h_{1,5}$
$h_{1,6}$
$h_{2,1}$
$h_{2,2}$
$h_{2,3}$
$h_{2,4}$
$h_{2,5}$
$h_{2,6}$
$h_{3,1}$
$h_{3,2}$
$h_{3,3}$
$h_{3,4}$
$h_{3,5}$
$h_{3,6}$
y_1

Figure 3.1.: Schematic overview of a fully Connected feed-forward neural network with an input dimension equal to 8, 3 hidden layers with dimension 6, and an output layer with dimension 1.

3.2. Local Generalized Linear Model Network (LGLMN)

A drawback of using the FFNN from Section 3.1 is its interpretability. Due to the complex structure of an FFNN it is hard to analyze how individual feature components x_j of $\mathbf{x}$ affect the regression function (3.5). Richman and Wüthrich [2023] introduced the local generalized linear model network (LGLMN), which is an extension of a generalized linear model (GLM) with an FFNN, that contains the data-driven structure while keeping the interpretable structure of the GLM.

The idea behind this model is that it uses a FFNN in which is output dimension is equal to the input dimensions (i.e. the number of independent variables used in the regression). Subsequently, a layer is added to the FFNN, and a linear activation function will be applied to the neurons in this added layer and the input variables. Therefore, the last layer consists of a linear function between the neurons of the second-last layer and the input variables. The neurons in this second-last layer can be seen as regression attentions of the input variables. This structure allows for more interpretability than a standard FFNN.

This paper uses the same notation and terminology on the LGLMN as Richman and Wüthrich [2023]. Let there be a mapping

$$\boldsymbol{\beta} : \mathbb{R}^q \to \mathbb{R}^q, \tag{3.6}$$

with equal input and output dimensions q, defined by a neural network

$$\mathbf{x} \mapsto \boldsymbol{\beta}(\mathbf{x}) = \mathbf{z}^{(d:1)}(\mathbf{x}) = \left(\mathbf{z}^{(d)} \circ \ldots \circ \mathbf{z}^{(1)}\right)(\mathbf{x}). \tag{3.7}$$

Then the LGLMN is defined by

$$\mathbf{x} \mapsto g(\mu) = g(\mu(\mathbf{x})) = \beta_0 + \langle \boldsymbol{\beta}(\mathbf{x}), \mathbf{x} \rangle. \tag{3.8}$$

Note that, in the regression structure of (3.8) the $\beta_j(\mathbf{x})$'s are dependent on $\mathbf{x}$. Because of this dependence $\boldsymbol{\beta}(\mathbf{x})$ is called a regression attention, instead of a regression parameter, which it is called when there is no dependence between $\boldsymbol{\beta}$ and $\mathbf{x}$. Due to its structure, the LGLMN allows for variable selection, for the study of interactions, and for variable importance ranking.

A schematic overview of an LGLMN is shown in Figure 3.2.

3.2.1. Interpretability

The structure of (3.8) can be interpreted as an attention mechanism as each $\beta_j(\mathbf{x})$ decides how much attention should be given to feature value x_j. This section provides an explanation of the interpretability of the LGLMN.

Interactions

The structure of the LGLMN gives information about interactions. Condition

$$\beta_j(\mathbf{x}) = \beta_j(x_j), \tag{3.9}$$

indicates that $\beta_j(x_j)x_j$ does not interact with other terms.

The component $\beta_j(\mathbf{x})$ can be analyzed for changes in $\mathbf{x}$. If $\beta_j(\mathbf{x})$ is insensitive in the components different from j, then there are no interactions. This can be analyzed by examining the gradients

$$\nabla\beta_j(\mathbf{x}) = \left(\frac{\partial}{\partial x_1}\beta_j(\mathbf{x}), ..., \frac{\partial}{\partial x_q}\beta_j(\mathbf{x})\right)^{\mathrm{T}} \in \mathbb{R}^q. \tag{3.10}$$

The j-th component of the gradient $\nabla\beta_j(\mathbf{x})$ indicates whether the term in x_j is linear, while the other components quantify the interaction strengths.

Variable importance

Define the variable importance measure of component j by aggregating the absolute values of the attention weights

$$\mathrm{VI}_j = \frac{1}{n}\sum_{i=1}^{n} \left|\hat{\beta}_j(\mathbf{x}_i)\right|, \tag{3.11}$$

for $1 \leq j \leq q$, where the components are averaged over all observations $1 \leq i \leq n$. The components $j \in \{1, ..., q\}$ can be ordered on their importance by their value of VI_j, the higher meaning the more important.

Note that, for a fair comparison of the variable importance VI_j in (3.11), the variables $\mathbf{x}_i$ need to be normalized. The one-hot encoded variables used in this thesis are not normalized. Therefore, the normalized variables will be compared with each other on variable importance, while the one-hot encoded variables will be compared with each other on variable importance. Note that, since the one-hot encoded variables are not normalized, it does not make sense to compare the variable importance between normalized variables and one-hot encoded variables using (3.11) because they have another scale.

Incomplete identifiability

The LGLMN does not have full identifiability in model calibration. Let $j \in \{1, ..., q\}$. Then the following structure could be observed:

$$\beta_j(\mathbf{x})x_j = x_k, \tag{3.12}$$

with $k \in \{1, ..., q\}$ and $k \neq j$. However, Richman and Wüthrich [2023] did not observe any such problems in their experiments.

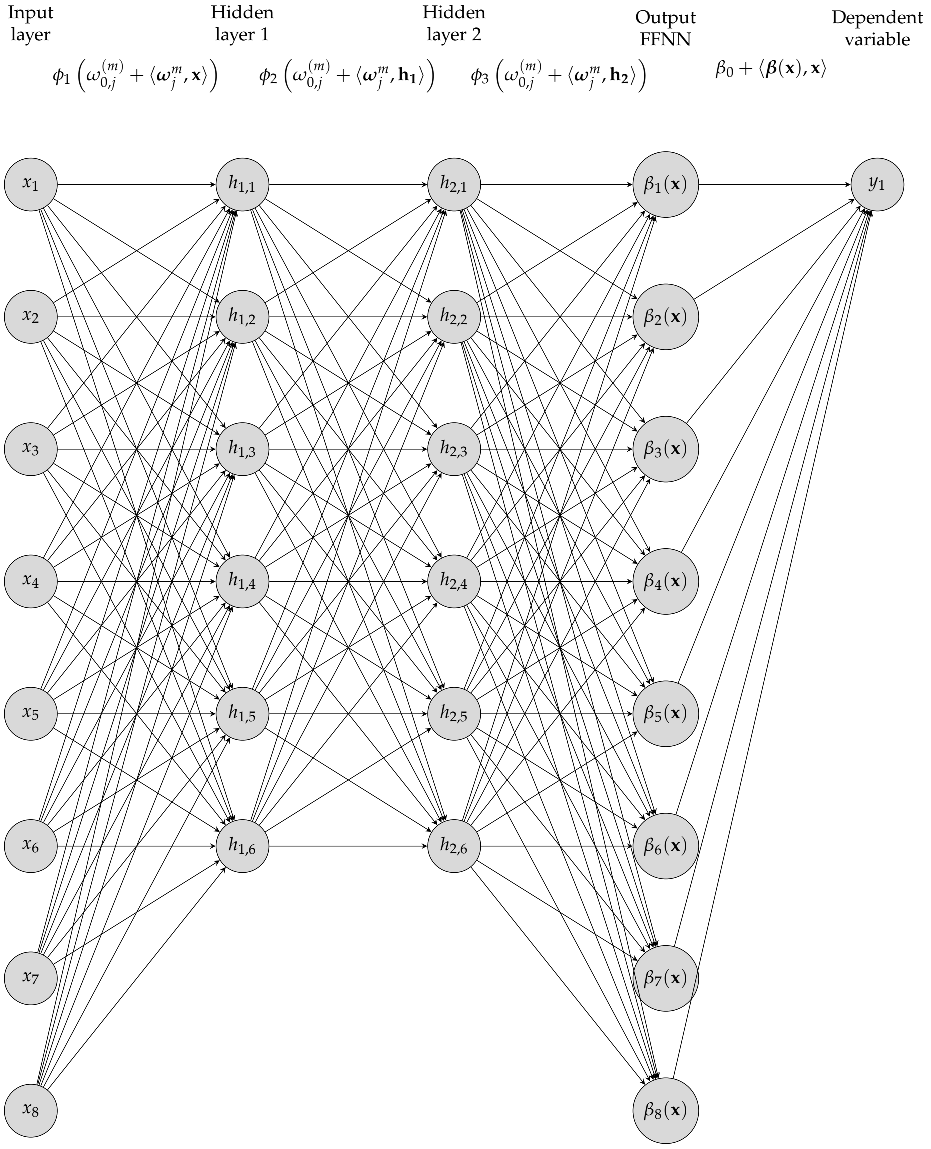

Input layer
Hidden layer 1
Hidden layer 2
Output FFNN
Dependent variable
$\phi_1\left(\omega_{0,j}^{(m)}+\langle\boldsymbol{\omega}_j^m,\mathbf{x}\rangle\right)$
$\phi_2\left(\omega_{0,j}^{(m)}+\langle\boldsymbol{\omega}_j^m,\mathbf{h_1}\rangle\right)$
$\phi_3\left(\omega_{0,j}^{(m)}+\langle\boldsymbol{\omega}_j^m,\mathbf{h_2}\rangle\right)$
$\beta_0+\langle\boldsymbol{\beta}(\mathbf{x}),\mathbf{x}\rangle$
x_1
x_2
x_3
x_4
x_5
x_6
x_7
x_8
$h_{1,1}$
$h_{1,2}$
$h_{1,3}$
$h_{1,4}$
$h_{1,5}$
$h_{1,6}$
$h_{2,1}$
$h_{2,2}$
$h_{2,3}$
$h_{2,4}$
$h_{2,5}$
$h_{2,6}$
$\beta_1(\mathbf{x})$
$\beta_2(\mathbf{x})$
$\beta_3(\mathbf{x})$
$\beta_4(\mathbf{x})$
$\beta_5(\mathbf{x})$
$\beta_6(\mathbf{x})$
$\beta_7(\mathbf{x})$
$\beta_8(\mathbf{x})$
y_1

Figure 3.2.: Schematic overview of a local generalized linear model network with an input dimension equal to 8 and two hidden layers with dimension 6.

4. Methodology

In this research, the hedonic price methodology will be used with the natural logarithm of the selling price as a dependent variable. This logarithmic structure is used for multiple reasons: it allows the value added to vary proportionally with the size and quality of the home, the variables can be interpreted as a percentage change, and it mitigates problems with heteroskedasticity (Malpezzi et al. [2003]).

Let y_{itm} be the log selling price of transaction i at time t at location m. Assume without loss of generality that the time of the transactions is modeled on quarterly level $t = 1, .., T$, and the location on a neighborhood level, with M neighborhoods $m \in \{1, ..., M\}$. In this research, the log selling price will be modeled as the sum of a property characteristics component, a spatial-temporal component, and a normally distributed error:

$$\begin{aligned} & y_{itm} = f(\mathbf{x}_{it}) + g(m,t) + \varepsilon_{itm}, \\ & \varepsilon_{itm} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \end{aligned} \tag{4.1}$$

where the component $f(\mathbf{x}_{it})$ is a function of the property characteristics $\mathbf{x}_{it} = (x_{i,t,1}, ..., x_{i,t,n})^{\mathrm{T}}$ and spatial-temporal component $g(m,t)$ is a function of space m and time t.

All models described in this chapter are different variants of the model in equation 4.1. For the spatial-temporal component, two types of SDFMs will be used. Section 4.1.1 describes the SDFM based on the spatial random walk (SRW) and Section 4.1.2 describes the SDFM based on the Gaussian random field (GRF). For the property characteristics component $f(\mathbf{x}_{it})$, a linear structure, a nonlinear structure, a FFNN structure, and a LGLMN structure will be used which are elucidated in Section 4.2.

4.1. Spatial Dynamic Factor Model

A spatial dynamic factor model (SDFM) is a DFM containing a spatial structure. There are multiple applications of SDFMs (Gamerman et al. [2008] ; Lopes et al. [2011] ; Strickland et al. [2011]), of which one in the repeat sales framework for real estate valuation (Francke et al. [2022]). All have a common purpose of incorporating spatial dependencies while reducing the dimensionality and hence the complexity of the problem. This section describes the methodology of the SDFM used in the setting of this research.

Choose a number of factors $K << M$. Let $g(m,t)$ be a spatial-temporal effect that depends on time $t \in \{1, ..., T\}$ and neighborhood $m \in \{1, ..., M\}$. Then $g(m,t)$ can be modeled using an SDFM by:

$$\begin{aligned} & g(m,t) = \mathbf{\Gamma}_m \mathbf{F}_t + \varepsilon_{tm}, \\ & \quad \varepsilon_{itm} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \end{aligned} \tag{4.2}$$

where Γ $(M \times K)$ is the matrix of the loadings and F $(K \times T)$ is the matrix of the factors.

In this research, the dynamic factors $\mathbf{F}_t$ follow a random walk

$$\mathbf{F}_{t,k} \sim \mathcal{N}\left(\mathbf{F}_{t-1,k}, 1\right), \tag{4.3}$$

for $t = 2, ..., T$, starting with $F_{1,k} = 0$ for all factors k. The structure of the loadings matrix is dependent on the locations of the observations and differs for the type of SDFM. The loadings matrices will be elaborated in detail in the following sections. Section 4.1.1 presents an SDFM with a 1D spatial structure on the loadings matrix and Section 4.1.2 introduces an SDFM with a 2D spatial structure on the loadings matrix. The 2 types of SDFMs that are used in this research are elucidated in the subsequent sections.

4.1.1. Spatial Random Walk

The first SDFM that is implemented in this research is an SDFM introduced by Francke et al. [2022]. In that research, an SDFM is presented to model condominium prices in Manhattan, New York, in a repeat sales framework. In this thesis, the SDFM is adapted to a HPM framework.

The SDFM has a 1D spatial structure which is specified by a spatial random walk (SRW). The methodology of this SDFM follows Francke et al. [2022]. Choose the number of factors $K \in \mathbb{N}$. First, define the loadings matrix Γ $(M \times K)$ to have an upper triangular structure with positive diagonal elements for identification purposes:

$$\Gamma = \begin{bmatrix} \gamma_{11}^{+} & 0 & \cdots & \cdots & 0 \\ \gamma_{21} & \gamma_{22}^{+} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots \\ \gamma_{K-1,1} & \gamma_{K-1,2} & \cdots & \gamma_{K-1,K-1}^{+} & 0 \\ \gamma_{K,1} & \gamma_{K,2} & \cdots & \gamma_{K,K-1} & \gamma_{K,K}^{+} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \gamma_{M,1} & \gamma_{M,2} & \cdots & \gamma_{M,K-1} & \gamma_{M,K} \end{bmatrix}, \tag{4.4}$$

where the superscript $+$ indicates a positive value.

The loadings γ_{km} have a spatial structure based on the SRW introduced by Francke and Van de Minne [2021]. The SRW is defined in two steps. First, the travelling sales person (TSP) algorithm (Lenstra and Kan [1975]) is applied to the coordinates of the centers of the neighborhoods to derive a 1-dimensional ordering of the neighborhoods. This step summarizes a 2D space into a 1D line that goes through all neighborhood centers using the shortest possible route. Unlike Francke et al. [2022], the starting point of the TSP-route is fixed at the neighborhood with the most observations. This is done because using a neighborhood with very few observations might give unstable results. The second step of the SRW assumes the loadings of the SDFM to have a spatial structure based on the following ordering:

$$\gamma_{k,(m)} \sim \mathcal{N}\left(\gamma_{k,(m-1)}, \sigma_{\gamma}^{2}\right), \tag{4.5}$$

where subscripts (m) denote the neighborhoods ordered by the TSP route. Note that, in this research, it is assumed that the variance is the same for every column of Γ, i.e. σ_γ^2 does not depend on k. However, it is possible to have a different variance $\sigma_\gamma^2(k)$ for every column.

The SDFM based on the SRW is defined by equations (4.2), (4.3), (4.4), and (4.5). A standard normal distribution is used as the prior distributions for σ_ε and σ_γ.

Note that due to the Bayesian structure of this SDFM, the spatial-temporal effect of a transaction is not only determined by the transactions at the same time and location but also by transactions at close times and locations. Both geographical similarity of neighborhoods (i.e. how close they are to each other) and categorical similarity of neighborhoods (i.e. neighborhoods that are similar in terms of characteristics, but that are not necessarily close to each other, might show similar behavior in selling prices of real estate) are captured by the model. This categorical similarity of neighborhoods is incorporated in a data-driven way. Therefore, the model is easily applicable in different situations and also applicable when no data about the characteristics of a neighborhood itself is available (i.e. when only the location of the neighborhood and characteristics of the houses in the neighborhood are available but not characteristics of the neighborhood itself are known, as this might not be the case for every application).

4.1.2. Gaussian Random Field

In Section 4.1.1, the loadings of the SDFM only contain a 1-dimensional structure based on the TSP-route through the neighborhood centers. Therefore, there could be neighborhoods that are close to each other in the sense of 2-dimensional distance but are very far away from each other in the TSP-ordering. Hence, intuitively a lot of spatial information between the neighborhoods gets lost when using the SDFM based on the SRW. Gamerman et al. [2008] introduced an SDFM in which the 2-dimensional distance between the neighborhoods is incorporated. In this model, the observations are modeled using a GRF, which is a distance-based Gaussian random process. In the research from Gamerman et al. [2008], the observations are modeled with each observation having its own sub-trend and without covariate effects. In this research, covariates or added and the observations are merged based on locational level (e.g. district level or neighborhood level).

The SDFM based on the GRF is defined as follows. Choose the number of factors $K \in \mathbb{N}$. Let loadings matrix $\Gamma = (\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_K)$. Denote Γ_k as the k^{th} column of Γ. Then Γ_k is modeled as a GRF:

$$\Gamma_k \sim \text{GRF}\left(\mu_k, \tau_k^2 \rho_{\phi_k}(\cdot)\right) := \mathcal{N}\left(\mu_k, \tau_k^2 \mathcal{R}_{\phi_k}\right), \tag{4.6}$$

where μ_k is a K-dimensional mean vector. The (m, n)-element of the matrix $\mathcal{R}_{\phi_k}$ $(M \times M)$ is defined by

$$r_{mn} = \exp \frac{-d(m,n)}{\phi_k}, \tag{4.7}$$

where $d(m, n)$ is the distance between neighborhood m and neighborhood n and parameter $\phi_k > 0$.

The SDFM based on the GRF is defined by equations (4.2), (4.3), and (4.6). A standard normal distribution is used as the prior distributions for σ_ε, ϕ_k (for $k = 1, ..., K$), and σ_F.

Note that the correlations between these spatial-temporal parts (linear combinations of factors and loadings) are not a direct function of Euclidean distance but that the information of these Euclidean distances is captured in the prior distributions of the loadings. Using this structure, near neighborhoods are stimulated to show similar spatial-temporal behavior, but still are allowed to show completely different behaviors when necessary. At the same time, spatially distant neighborhoods are not directly incentivized to show similar spatial-temporal behavior by the characterization of their priors, however, it is still feasible that they are. This is a desirable structure, as generally close neighborhoods should show more similar spatial-temporal trends than distant ones, yet, distant neighborhoods should have the possibility to show similar trends as distant neighborhoods might look alike in other terms (e.g. in a city with two distant neighborhoods in which both a university is located, the spatial-temporal indices of these housing prices might look very similar.).

4.2. Property Characteristics Component

This section elucidates the 4 different property characteristic components used in this research. These components are subsequently described in the order of their simplicity, starting with the linear component and closing with the LGLMN component.

Intuitively, an increase in lot size for a house with a small lot size should have more impact than an increase in lot size for a house with a large lot size. Therefore, instead of using the variable *lot size* as an input for the regression models in this research, linear splines of the variable are used. The structure of linear splines is described in Section A.1. The splines for the *lot size* are divided into the 3 intervals $[0, 200), [200, 500)$, and $[500, \infty)$. The sub-variables of these splines are used as variables for the linear component in Section 4.2.1 and the non-linear component in Section 4.2.2. The models in Section 4.2.3 and Section 4.2.4 have a data-driven structure, and therefore, only the value of *lot size* is used as input instead of using linear splines of the lot size.

4.2.1. Linear Component

The linear property characteristic component is a linear combination of the property characteristics $\mathbf{x}_{it}$ and the parameters $\boldsymbol{\beta} = (\beta_1, ..., \beta_n)^{\mathrm{T}}$ including an intercept α. The logarithm of the *house size* is added to the model and the *lot size* is decomposed into linear splines, as described in Section A.1. The representation of this linear component is given by

$$f(\mathbf{x}_{it}) = \alpha + \boldsymbol{\beta}\mathbf{x}_{it}. \tag{4.8}$$

4.2.2. Non-linear Component

A linear functional form might not always be the most optimal form to include property characteristics in a real estate valuation model. For instance, research from Goodman [1978] has shown a nonlinear relationship between house price and house size.

For the non-linear property characteristics component, this thesis follows a simplification of Francke and van de Minne [2017], by decomposing the property characteristics value in the structure and land value. This component is written as a scalar plus the natural logarithm of the sum of structure and land value

$$f(\mathbf{x}_{it}) = \alpha + \ln\left(SV_{it} + LV_{it}\right) + \epsilon_{it}, \tag{4.9}$$

where S_{it} is the structure value and LV_{it} the land value. Subsequently, the value of the structure can be written as

$$SV_{it} = (\mathit{house\ size}(i))^{\beta} \exp\left[\sum_{h=1}^{H} SC_h(i)\gamma_h\right], \tag{4.10}$$

where SC_h indicates the structure characteristics variable h such as building year and building type, and γ_h their corresponding coefficients for $h = 1, ..., H$, where H is the number of structure characteristics.

The land value is written as a weighted sum of the linear splines of the *lot size*, i.e.

$$LV_{it} = \sum_{s=1}^{S} \mathrm{ls}_s(i)\nu_s, \tag{4.11}$$

where S the number of linear splines, and $\mathrm{ls}_s(i)$ the value of sub-variable ls_s of transaction i.

Equations (4.9), (4.10), and (4.11) form the non-linear structure of the property characteristics component. In this structure, the coefficients α, β, γ_h ($h = 1, ..., H$), ν_s ($s = 1, ..., S$) will be estimated.

4.2.3. Neural Network Component

Instead of using a pre-specified structure for the property characteristics component, an FFNN can be used. The FFNN has a data-driven structure. Only the hyper-parameters such as the number of layers d and the neurons per layer q_m have to be specified in advance. The neural network representation of the property characteristics component is given by is given by

$$f(\mathbf{x}_{it}) = \beta_0 + \langle \boldsymbol{\beta}, \mathbf{z}^{(d:1)}(\mathbf{x}_{it}) \rangle \tag{4.12}$$

where $\mathbf{z}^{(d:1)}(\cdot)$ is specified in Section 3.1 with an input layer of dimension $q_0 = n$.

4.2.4. Local Generalized Linear Model Network Component

Lastly, the LGLMN is used as the property characteristics component. This model allows for a data-driven and interpretable structure. The LGLMN representation of the property characteristics component is given by

$$f(\mathbf{x}_{it}) = \beta_0 + \langle \boldsymbol{\beta}(\mathbf{x}), \mathbf{x} \rangle, \tag{4.13}$$

where $\boldsymbol{\beta}(\mathbf{x})$ is defined in Section 3.2 where the input and output layers of the FFNN used in the LGLMN have a dimension $q_0 = q_d = n$.

4.3. Model Representation

The models in this research are based on the general framework (4.1), in which different combinations of the 4 property characteristics component $f(\mathbf{x}_{it})$ and the 2 spatial-temporal components $\mathbf{\Gamma}_m \mathbf{F}_t$ are used. This leads to 8 different models which are shown in Table 4.1.

Table 4.1.: Schematic representation of property characteristics components and the spatial-temporal components of the models.

Model	Property characteristics	Spatial-temporal
Linear adjusted SRW-DFM (L-SRW-DFM)	Linear	SRW
Nonlinear adjusted SRW-DFM (NL-SRW-DFM)	Non-linear	SRW
NN adjusted SRW-DFM (NN-SRW-DFM)	NN	SRW
LGLMN adjusted SRW-DFM (LGLMN-SRW-DFM)	LGLMN	SRW
Linear adjusted GRF-DFM (L-GRF-DFM)	Linear	GRF
Non-linear adjusted GRF-DFM (NL-GRF-DFM)	Non-linear	GRF
NN adjusted GRF-DFM (NN-GRF-DFM)	NN	GRF
LGLMN adjusted GRF-DFM (LGLMN-GRF-DFM)	LGLMN	GRF

4.3.1. Model Extensions

The models described in Table 4.1 can be extended in several ways, e.g. by adding a common trend or a by adding random effects. This section describes these extensions.

Common trend

One might suggest that besides sub-trends on a neighborhood level, also a common market trend, indifferent of location, affects the house price. This common trend, μ_t can be modeled by a random walk

$$\mu_{t+1} \sim \mathcal{N}(\mu_t, \sigma_\mu^2), \quad \mu_0 = 0. \tag{4.14}$$

Subsequently, equation (4.1) can be extended with this common trend:

$$\begin{aligned} y_{itm} &= f(\mathbf{x}_{it}) + \mathbf{\Gamma}_m \mathbf{F}_t + \mu_t + \varepsilon_{itm}, \\ \varepsilon_{itm} &\sim \mathcal{N}(0, \sigma_\varepsilon^2),. \end{aligned} \tag{4.15}$$

Random effects of geographical level

When using a very small geographical level (e.g. street level, census tract level, or neighborhood level) in the SDFM it can be heavy to compute. However, information can get lost when using a larger geographical level (e.g., district level). A possible solution for this is to use the SDFM on a large geographical level, such as a district, while adding a random effect of a smaller geographical level, such as a neighborhood to the model. Then the price dynamics are modeled only on a district level, while significant price differences between neighborhoods are still incorporated in the model.

Using this extension, the price of transaction i at time t in district d and neighborhood n noted as y_{itdn} can be modelled by

$$\begin{aligned} y_{itdn} &= f(\mathbf{x}_{it}) + \mathbf{\Gamma}_d \mathbf{F}_t + \nu_n + \varepsilon_{itm}, \\ \varepsilon_{itm} &\sim \mathcal{N}(0, \sigma_\varepsilon^2), . \\ \nu_n &\sim \mathcal{N}(0, \sigma_\nu^2), \end{aligned} \tag{4.16}$$

in which ν_n is the neighborhood random effect. Especially when using a large number of dynamic factors, this extension is faster to compute than the SDFM on neighborhood level, as fewer parameters have to be estimated.

4.4. Model Estimation

In this thesis, the models will be estimated with 2 different estimation techniques: Bayesian estimation and algorithmic estimation.

4.4.1. Bayesian Estimation

In the Bayesian estimation technique, the models are estimated using the NUTS, introduced by Hoffman et al. [2014]. This is a MCMC method implemented in high-performance statistical software Stan (Stan Development Team [2023]). The priors of the parameters are standard normal distributions. An advantage of this technique is that the models can be estimated directly using their given representations. However, a drawback of this estimation procedure is that it is ineffective to use a Bayesian estimation technique to estimate a FFNN. Therefore, this thesis only uses an FFNN with 1 hidden layer when using this procedure.

The NUTS-algorithm approximates the posterior distribution of the parameters by iterating over the sample set of the parameters. First, the algorithm uses warm-up iterations to learn the parameters. After that, it uses sampling iterations to create posterior draws. This is done over multiple chains. The settings of the NUTS-algorithm that is used in this research can be found in Table A.13.

High-performance computer (HPC)

The models based on the FFNN are computationally heavy to estimate and therefore, it can take a long time to estimate the models. For that reason, the models involving an FFNN are estimated using the high performance computer (HPC) of DelftBlue (Delft High Performance Computing Centre [DHPC]).

Uniqueness of loadings and factors

The output of the NUTS-algorithm is a distribution of each parameter in the model. In this research, the estimate of a parameter is determined by the mean over all instances of that parameter of the NUTS-algorithm. Only the estimates of the factors and loadings are determined differently. For the factors and the loadings of the models, the mean of the linear combinations over all instances is taken as the estimates in the models. This is done, as the factors and loadings were not unique for all chains (see figures A.1 and A.3). Thus instead of using $\hat{\mathbf{\Gamma}}_m\hat{\mathbf{F}}_t$, we use $\widehat{\mathbf{\Gamma}_m\mathbf{F}}_t$. For neighborhood m and time-step t, the linear combination of the factors and loadings is given by

$$\widehat{\mathbf{\Gamma}_m\mathbf{F}}_t = \text{mean}_{\text{chains}}\left[\text{mean}_{\text{iterations}}\left(\mathbf{\Gamma}_m\mathbf{F}_t\right)\right] = \frac{1}{N_{\text{chains}}}\sum_{k=1}^{N_{\text{chains}}}\left[\frac{1}{N_{\text{iterations}}(k)}\sum_{i=1}^{N_{\text{iterations}}(k)}\mathbf{\Gamma}_m(k,i)\mathbf{F}_t(k,i)\right], \tag{4.17}$$

where N_{chains} is the number of chains and $N_{\text{iterations}}(k)$ is the number of iterations in chain k.

4.4.2. Algorithmic Estimation

The estimation of an FFNN using the NUTS requires a lot of computing power, compared to other methods such as the stochastic gradient descent (SGD) algorithm (Amari [1993]). However, the SDFMs used in this research can not be estimated simultaneously with the FFNN when using the SGD algorithm. Therefore, inspired by Francke and Van de Minne [2023], in addition to the Bayesian approach described in Section 4.4.1, an algorithmic approach will be used to estimate the models.

The idea of this algorithmic approach is that the estimation of the model is split into 2 parts: the estimation of the spatial-temporal component and the estimation of the property characteristics component. Unlike the Bayesian estimation procedure, these 2 parts will not be estimated simultaneously, but they will be estimated consecutively repeatedly until convergence. This approach has 2 advantages compared to the Bayesian approach. The first advantage is that it is faster, since when estimating an FFNN separately from an SDFM, fast methods such as SGD can be used to estimate the FFNN. Therefore, a more complex FFNN, i.e. an FFNN with more layers, can be used, which could better capture the underlying structure of the data. Another advantage is that the FFNN can be replaced easily with other ML models, even models that can not be estimated, or models that take even longer to estimate in a Bayesian estimation procedure (Francke and Van de Minne [2023]). Therefore, this structure also allows for the usage of the LGLMN instead of the FFNN. This method, as described in Section 3.2, is more interpretable than a FFNN which is a desired feature. A limitation of this algorithmic approach is that it is hard or even impossible to know in advance whether the model converges. The algorithm used in this research is stated in Algorithm 1.

Algorithm 1 Iterative algorithm to estimate models with an FFNN or LGLMN

```
1. Estimate: L-SRW-DFM or L-GRF-DFM with NUTS.
2. Compute: in-sample MAPE and set ΔMAPE at an arbitrarily high number
While |ΔMAPE | ≥ C_0 do
    3. Compute: $\tilde{y}_{itm} = y_{itm} - \widehat{\Gamma_m F_t}$
    4. Train: FFNN or LGLMN over set $\{\tilde{y}_{itm}, x_{it}\}$, which gives $\hat{f}$
    5. Compute: FFNN or LGLMN algorithm over set $y^*_{itm} = y_{itm} - \hat{f}(x_{it})$
    6. Estimate: $y^*_{itm} = \Gamma_m F_t + \varepsilon_{it}$ to derive $\widehat{\Gamma_m F_t}$, using NUTS
    7. Compute: in-sample MAPE, mean(abs($y_{itm} - \bar{y}_{itm} - \hat{f}(\bar{x}_{it})$))
end while
```

4.5. Assessment of Performance

4.5.1. Prediction Accuracy Measures

Commonly used measures for prediction accuracy in real estate valuation are the mean absolute percentage error (MAPE) and the root mean squared error (RMSE) (Lee [2022] ; Wang et al. [2014] ; Abidoye and Chan [2018] ; Uzut and Buyrukoglu [2020] ; Khobragade et al. [2018]). Furthermore, when assessing the performance of an ANN, the MAPE and the RMSE are two of the most commonly used measures (Twomey and Smith [1995]). This section describes the 3 prediction accuracy measures that are used in this research: MAPE, RMSE, and RMSE of the log price.

Mean absolute percentage error (MAPE)

The MAPE is the mean of the absolute differences in percentage between the predicted value and the true value. Therefore, this measure is easy to interpret. Since the MAPE is a relative measure, it can still give an indication of the performance of valuation models when using different models on different data sets. The MAPE is formulated as

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|e^{y_i} - e^{\hat{y}_i}|}{e^{y_i}}, \tag{4.18}$$

with N the number of observations, y_i the true log transaction price of observation i, and $\hat{y}_i$ the predicted log transaction price of observation i.

Root mean squared error (RMSE)

In the field of real estate valuation, significant errors are particularly undesirable. Unlike the MAPE, the RMSE penalizes larger errors more than smaller errors, and therefore, it proves to be a useful measure for prediction accuracy. The RMSE is computed by

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(e^{y_i} - e^{\hat{y}_i}\right)^2} \tag{4.19}$$

for N the number of observations, y_i the true log transaction price of observation i, and $\hat{y}_i$ the predicted log transaction price of observation i.

Root mean squared error of the log price

In the models that are used in this research, the dependent variable is the natural logarithm of the house selling price. The models in this thesis are optimized based on this measure. Therefore, also this measure is used for the evaluation of the models. The RMSE log price is computed by

$$\text{RMSE log price} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4.20}$$

for N the number of observations, y_i the true log transaction price of observation i, and $\hat{y}_i$ the predicted log transaction price of observation i.

4.5.2. Validation Methods

Train-test split

In this research, we are particularly interested how the models perform on unseen data. Therefore, the data set is randomly split into a 70/30 train/test split. The training set is used to estimate the models and the test set is used to evaluate the performance of the models.

K-fold cross validation

A K-fold cross-validation is used on the training set, for the selection of some hyper-parameters. In k-fold cross-validation (CV) the data set will be randomly divided into k groups of approximately equal size. The first fold is treated as the validation set, and the model is fit on the data of the other $k-1$ folds. The mean squared error (MSE) is computed on the observations in the validation set. This is done k times, and every time a different fold is used as a validation set. This leads to k different estimates of the MSE. The k-fold CV estimate is the average of these values

$$\text{CV}(k) = \frac{1}{k}\sum_{i=1}^{k}\text{MSE}(i). \tag{4.21}$$

In this research a k-fold cross-validation is performed using $k = 5$, as it is shown empirically that this value results in estimates that suffer neither from extremely high bias nor from excessively high variance (James et al. [2013]). Note that the MSE in (4.21) can be interchanged by another loss function but in this thesis, the MSE will be used (Hastie et al. [2009]).

4.5.3. Convergence of Bayesian Estimation

To reflect on the convergence of the NUTS-algorithm, the convergence diagnostic $\hat{R}$ is analyzed. The $\hat{R}$ is a statistic that compares parameter estimates across the chains in the NUTS-algorithm. If chains have not mixed well the value for $\hat{R}$ is larger than 1. A rule-of-thumb is to only use samples when the values of $\hat{R} < 1.05$ (Vehtari et al. [2021]).

5. Data

The data used in this research is on residential real estate transactions in the Netherlands and is provided by Ortec Finance. For this research, only data from the municipality of Rotterdam is used. Section 5.1 describes the data preparation and Section 5.2 gives a detailed data analysis.

5.1. Data Preparation

There are 40 observations with a lot size larger than 2000. These observations are removed. Furthermore, there is data that is labeled as located in the municipality of Rotterdam but with latitude and longitude values that do not correspond to the latitude and longitude values of Rotterdam. These observations are omitted as well. After removing the described observations, the database contains 30,016 observations. All observations range from 02-01-2009 until 31-03-2021. The observations are quarterly.

5.2. Data Analysis

The dependent variable in this research is the natural logarithm of the house selling price. The independent variables used in this research are: *house size*, *lot size*, *building type*, and *building year class*. A limited number of independent variables is employed, as the usage of more independent variables increases the likelihood of misspecification. Furthermore, these independent variables are very basic and as a consequence, they are unusally always available. Therefore, this data set does not contain any missing data about these variables.

Table 5.1 shows the summary statistics of the quantitative variables used in this research. Both the statistics of the selling prices and the log selling prices are shown. The average selling price is about 270k euros. The summary statistics of the *house size* and of the *lot size* for the observations where the *lot size* is larger than 0 are shown in table 5.1. The smallest house in the data sample is 24 m^2 and the largest house is 750 m^2. Summary statistics of the quantitative variables per *building type* are shown in Appendix A.3.

Table 5.1.: Summary statistics of quantitative variables.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	30,016	270	190	32	155	325	3,100
Log price (in EUR)	30,016	12.3	0.6	10.4	12.0	12.7	15.0
House size (in m^2)	30,016	105	44	24	75	125	750
Lot size (obs. $>$ 0, m^2)	11,327	204	175	18	122	214	1,992

Linear splines are used for variable *lot size*. The interval of this variable is divided into 3 sub-intervals using $s_1 = 200$ and $s_2 = 500$ in equation A.1. The distribution of the *lot size* of the observations over the linear splines is shown in table 5.2. Note that more than half of the observations have a lot size equal to zero, which indicates that those observations do not have a piece of land next to the property, such as a garden, which seems reasonable as the observations are from Rotterdam which is an urban area. Table 5.2 shows the summary statistics of the sub-variables of the *lot size*. For the majority of the observations, these sub-variables are 0, as the majority of the observations have a lot size equal to 0.

Table 5.2.: Summary statistics of the linear splines variables

Statistic	*N*	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)	Max
Lot size = 0	18,689	0	0	0	0	0	0
ls_1	8,109	58.050	78.400	0	0	133	200
ls_2	2,679	13.059	51.053	0	0	0	300
ls_3	539	5.972	61.953	0	0	0	1,492

Figure 5.1 shows the correlation matrix of the numerical independent variables. As expected the correlation between the *house size* and the sub-variables of the *lot size* is significant. However, it is never very correlated, and thus both the *house size* and the sub-variables of the *lot size* will be used as input for the models.

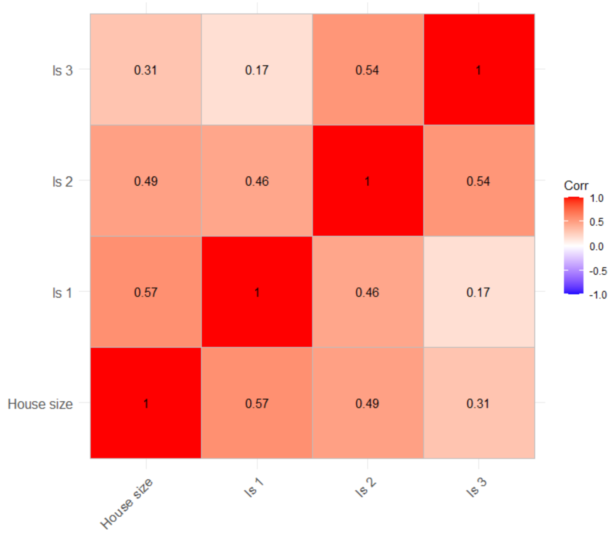


Figure 5.1.: Correlation matrix of numerical independent variables.

The data set contains 16 different secondary building types, which can be generalized into 5 primary building types: *detached*, *semi-detached*, *corner house*, *terraced house*, and *apartment*. Table A.1 shows the primary- and secondary-building types in the data set. A 5-fold CV is performed when using the 16 different secondary building types and when using the 5 primary building types in the L-SRW-DFM on the training set. The model performed slightly better when using the 5 building types instead of 16 based on the CV value (Table A.10). As the results are better for the 5 primary building types and the estimation of the model is faster, only the 5 primary building types will be used for the estimation and evaluation of the models in this thesis. The bar plot in figure 5.2 shows the distribution of the observations over the 5 different building types. Apartments are the most common building types, which is in line with a large number of houses with a *lot size* equal to 0 as shown in Table 5.2. Detached houses are the least common building type, which is hardly surprising in an urban area such as Rotterdam.

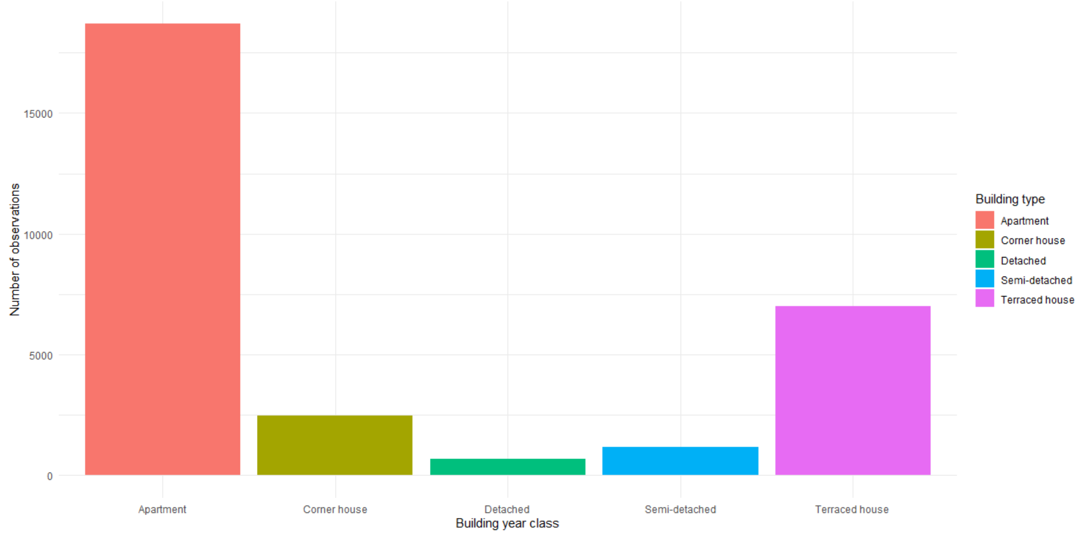


Figure 5.2.: Observations of the building types.

The distribution of the observations over the building year classes is shown in figure 5.3. The class with houses built after 2014 is the smallest. However, note that this is also one of the smallest time frames of the building year classes as the data reaches March 2021.

The data set contains observations in 76 different neighborhoods. Table 5.3 shows summary statistics on how many observations neighborhoods have. On average, the neighborhoods have 395 observations. The neighborhood with the least observations has one observation and the neighborhood with the most observations has 1,278.

Table 5.3.: Observations per neighborhood

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Observations	395	329	1	139	349	577	1,278

Figure 5.4 shows a histogram of the transactions over time. At first, each quarter has approximately the same amount of transactions. Later on, more transactions have been made but with more fluctuations. Furthermore, there are no quarters with very few observations. There are no seasonality patterns observed.

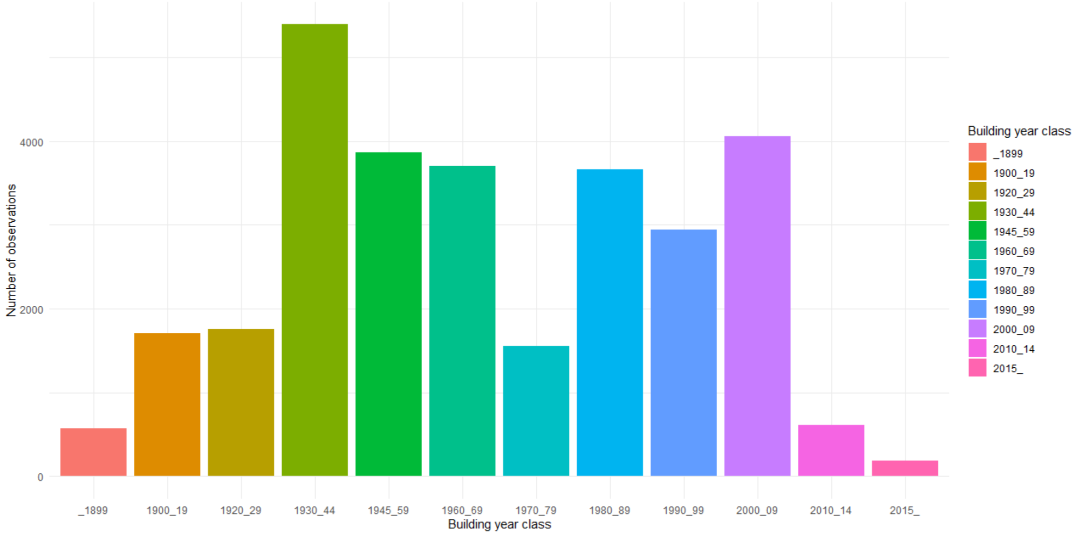


Figure 5.3.: Observations of the building year classes.

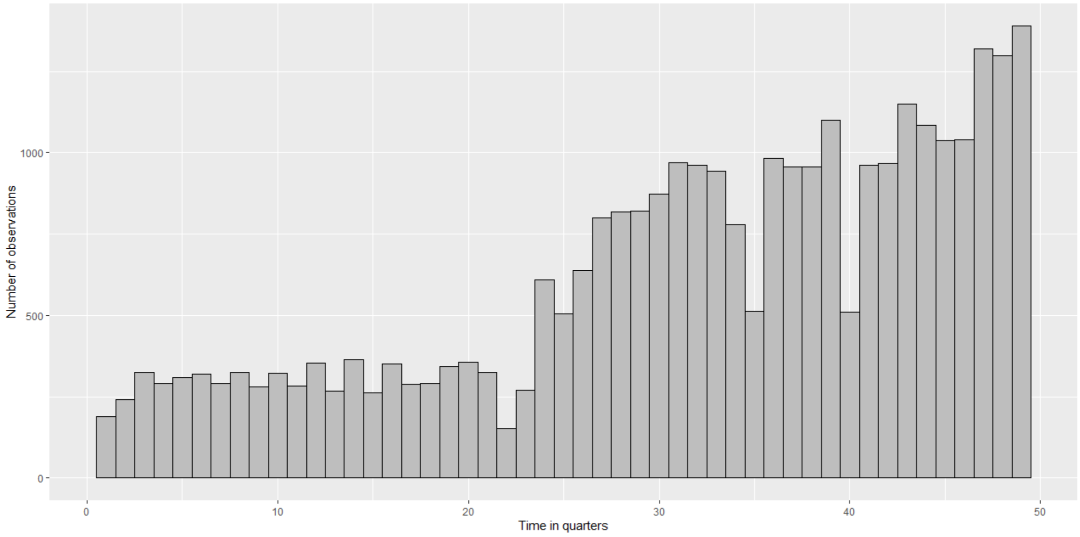


Figure 5.4.: Observations over time in quarters.

6. Simulation Study: Ability of the Neural Network in Capturing Linear and Non-linear Structures

Neural networks can capture both linear and non-linear structures (Goh [1995]). However, it is not clear in advance whether an FFNN can have those structures in combination with an SDFM. This section assesses the ability of the neural network to capture linear and non-linear structures in the property characteristics components of the models. An investigation of the model is conducted using realistic simulated data from a data generating process (DGP). The simulated data is obtained from a model containing a linear structure, as well as from a model containing a non-linear structure. The FFNNs are estimated using the Bayesian estimation procedure. Section 6.4 assesses the ability of an LGLMN to replicate regression coefficients of a linear structure when using an algorithmic estimation.

6.1. Data Generating Process

To generate data based on a linear and a non-linear structure, the L-SRW-DFM and the NL-SRW-DFM are estimated using the NUTS-algorithm on a sample of 10% of the total Rotterdam data set described in Chapter 5. The reason why only 10% of this data set is used is that it is faster to estimate the models on a smaller data set. Using the estimates of the parameters of these models, new data is created by sampling from the normal distribution. The location parameter of the normal distribution is equal to the combination of the estimated parameters and the data and the variance is equal to the variance of the errors of the model on the real data. Hence, for the model with the linear structure, the log transaction price $y_{\text{sim}}(x_{it}, m, t)$ for covariates x_{it}, neighborhood m, and time t, is simulated from

$$y_{\text{sim}}(x_{it}, m, t) \sim \mathcal{N}\left(\hat{\alpha} + \hat{\boldsymbol{\beta}}\mathbf{x}_{it} + \widehat{\boldsymbol{\Gamma}_m \mathbf{f}_t}, \widehat{\sigma_\varepsilon}^2\right), \tag{6.1}$$

where the estimated parameters are indicated with a hat.

Likewise, for the model with the nonlinear structure, the log transaction price is simulated from

$$y_{\text{sim}}(x_{it}, m, t) \sim \mathcal{N}\left(\hat{\alpha} + \ln\left((\textit{house size}(i))^{\hat{\beta}} \exp\left[\sum_{h=1}^{H} SC_h(i)\hat{\gamma}_h\right] + \sum_{s=1}^{S} \text{ls}_s(i)\hat{v}_s\right) + \widehat{\boldsymbol{\Gamma}_m \mathbf{f}_t}, \widehat{\sigma_\varepsilon}^2\right). \tag{6.2}$$

Note that two data sets are simulated in this DGP. One contains a linear structure for the property characteristics component and one contains a nonlinear structure. Both data sets contain 3,075 observations over a training set and a test set, which is split into a training set of 2,152 observations and a test set of 923 observations. The estimates $\widehat{\sigma_\varepsilon}^2$ of equations (6.1) and (6.2) are shown in Table 6.1.

Table 6.1.: Estimated variance of the models simulated by the DGP.

	L-SRW-DFM	NL-SRW-DFM
$\widehat{\sigma_\varepsilon}^2$	0.174	0.173

6.2. Assessment

To assess the ability of the neural network to capture linear and nonlinear structures, the NN-SRW-DFM is estimated on both the generated data with a linear structure and the generated data with a nonlinear structure. For comparison also the L-SRW-DFM and the NL-SRW-DFM are respectively estimated on the generated data with a linear structure and the generated data with a nonlinear structure. These models are the 'true' models for this data, and therefore, ideally the results of the NN-SRW-DFM on the data set should be close to the results of the L-SRW-DFM and the NL-SRW-DFM.

For the assessment of the models, the performance measures are compared to analyze whether the models that use a neural network perform nearly as well as their linear and nonlinear counterparts. Furthermore, differences between the estimated property characteristics components, and the differences between the estimated spatial-temporal components of the models are compared.

The difference between the property characteristics components of the models is given by

$$f_{\mathrm{T}}(\mathbf{x}_{it}) - f_{\mathrm{E}}(\mathbf{x}_{it}), \tag{6.3}$$

where $f_{\mathrm{T}}(\cdot)$ is the function of the property characteristics in the true model and $f_{\mathrm{E}}(\cdot)$ the function of the property characteristics in the estimated model.

Likewise, the difference between the spatial-temporal components of the models is given by

$$(\mathbf{\Gamma}_m \mathbf{F}_t)_{\mathrm{T}} - (\mathbf{\Gamma}_m \mathbf{F}_t)_{\mathrm{E}}\,, \tag{6.4}$$

where the subscript T indicates the true model and the subscript E indicates the estimated model.

6.3. Results

This section discusses the results of the NN-SRW-DFM on the artificial data compared to the results of respectively the L-SRW-DFM and the NL-SRW-DFM on respectively the data with the linear structure and the data with the nonlinear structure.

6.3.1. Linear Structure

Table 6.2 shows the in-sample and out-of-sample prediction accuracy measures of the L-SRW-DFM and the NN-SRW-DFM. The NN-SRW-DFM performs better on the in-sample data, which is an indicator of the over-fitting of the model. The out-of-sample performance is slightly worse for the NN-SRW-DFM compared to the L-SRW-DFM. However, the performance measures of the NN-SRW-DFM are very close to the ones of the L-SRW-DFM, which is the first sign that the FFNN is able to capture the linear structure well.

Table 6.2.: In-sample and out-of-sample performance measures of the models on data simulated from the L-SRW-DFM.

	L-SRW-DFM	NN-SRW-DFM
MAPE (in)	0.136	0.131
MAPE (out)	0.140	0.144
RMSE (in)	51,589	49,347
RMSE (out)	53,197	55,169
RMSE log price (in)	0.167	0.162
RMSE log price (out)	0.174	0.177

The estimated coefficients of the property characteristics part of the true L-SRW-DFM and the L-SRW-DFM on the generated data are shown in Table A.7. Note that the estimates are very similar which is as expected as the model on the right-hand-side is estimated on data generated from the coefficients on the left-hand-side.

Table 6.3 shows the distributions of the differences between the true property characteristics component and the estimated property characteristics components of the L-SRW-DFM and the NN-SRW-DFM on the simulated data. Note that the standard deviation is a bit higher and the absolute values of the minimum and maximum differences are greater when comparing the base model with the L-SRW-DFM. However, the distribution of the difference between the true property characteristics component and the property characteristics component estimated with the NN-SRW-DFM is clearly not far off from the distribution of the differences of the fixed part between the base model and the L-SRW-DFM.

Table 6.3.: In-sample and out-of-sample distributions of the differences of the true property characteristics components and the property characteristics components estimated from the L-SRW-DFM and NN-SRW-DFM on the artificial data.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
L-SRW-DFM (in)	−0.012	0.020	−0.149	−0.025	−0.011	0.003	0.053
L-SRW-DFM (out)	−0.010	0.020	−0.141	−0.022	−0.008	0.004	0.052
NN-SRW-DFM (in)	−0.004	0.039	−0.287	−0.023	0.006	0.015	0.212
NN-SRW-DFM (out)	−0.014	0.043	−0.228	−0.040	−0.006	0.011	0.189

Table 6.4 shows the distributions of the differences between the true spatial-temporal component and the spatial-temporal components of the L-SRW-DFM and the NN-SRW-DFM estimated on the simulated data. Note that both the in-sample and out-of-sample distributions of

these differences are very similar. This is what should be expected as the structure of the spatial-temporal component is the same for both models.

Table 6.4.: In-sample and out-of-sample distributions of the differences of the true spatial-temporal components and the spatial-temporal components estimated from the L-SRW-DFM and NN-SRW-DFM on the artificial data.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
L-SRW-DFM (in)	0.011	0.045	−0.172	−0.015	0.012	0.040	0.151
L-SRW-DFM (out)	0.011	0.046	−0.140	−0.014	0.012	0.039	0.149
NN-SRW-DFM (in)	0.003	0.045	−0.184	−0.024	0.003	0.030	0.156
NN-SRW-DFM (out)	0.003	0.046	−0.153	−0.022	0.002	0.032	0.151

In Figure A.1 the factors of the true L-SRW-DFM model, and of the L-SRW-DFM and the NN-SRW-DFM on the data are plotted for every chain of the NUTS iterations. Likewise, in Figure A.2 the loadings of the true L-SRW-DFM model, and of the L-SRW-DFM and the NN-SRW-DFM on the data are shown for every chain. Note that although the factors are not unique for every chain(see Section 4.4.1), there seem to be two 'flavors'. The factors of chains 1, 2, and 4 are very similar, while the factors of chain 3 are different from the ones in chains 1, 2, and 4. These flavors are recovered by the L-SRW-DFM and the NN-SRW-DFM when estimated on the data. The factors of chains 1 and 2 in Figure A.1b are similar to the factors of chains 1, 2, and 4 in Figure A.1a. Likewise, the factors of chains 3 and 4 in Figure A.1b are similar to the factors of chains 3 in Figure A.1a. Note that also the factors in Figure A.1c are similar to the ones in Figure A.1a. However, note that the scales on the axis are different, and thus, the structure is replicated on another scale.

6.3.2. Nonlinear Structure

Table 6.5 shows respectively the in-sample and out-of-sample prediction accuracy measures of the NL-SRW-DFM and the NN-SRW-DFM. As for the NL-SRW-DFM, the NN-SRW-DFM performs better on the in-sample data and is slightly worse on the out-of-sample data. Also for the nonlinear structure, the NN-SRW-DFM performs quite well as the performance measures are close to the ones of the true model.

Table 6.5.: In-sample and out-of-sample performance measures of the models on data simulated from the NL-SRW-DFM.

	NL-SRW-DFM	NN-SRW-DFM
MAPE (in)	0.129	0.123
MAPE (out)	0.149	0.156
RMSE (in)	50,995	45,946
RMSE (out)	53,034	56,721
RMSE log price (in)	0.161	0.154
RMSE log price (out)	0.185	0.191

The estimated coefficients of the property characteristics part of the true NL-SRW-DFM and the NL-SRW-DFM on the generated data are shown in Table A.8. Note that, as for the L-SRW-DFM, the estimates are very similar which is as expected as the model on the right-hand-side is estimated on data generated from the coefficients on the left-hand-side.

Table 6.6 shows the distributions of the differences between the true property characteristics component and the estimated property characteristics components of the NL-SRW-DFM and the NN-SRW-DFM on the simulated data. As for the linear structure, the standard deviation is a bit higher and the absolute values of the minimum and maximum differences are greater, when comparing the base model with the NL-SRW-DFM. However, the distribution of the difference between the true property characteristics component and the NN-SRW-DFM is clearly not far off from the distribution of the differences of the property characteristics components between the base and the NL-SRW-DFM.

Table 6.6.: In-sample and out-of-sample distributions of the differences of the true property characteristics components and the property characteristics components estimated from the NL-SRW-DFM and NN-SRW-DFM on the artificial data.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
NL-SRW-DFM (in)	0.019	0.021	−0.075	0.004	0.020	0.036	0.067
NL-SRW-DFM (out)	0.018	0.022	−0.063	0.004	0.018	0.036	0.063
NN-SRW-DFM (in)	0.018	0.043	−0.206	−0.001	0.019	0.043	0.242
NN-SRW-DFM (out)	0.011	0.058	−0.215	−0.024	0.019	0.049	0.277

Table 6.7 shows the distributions of the differences between the true spatial-temporal component and the spatial-temporal component of the NL-SRW-DFM and the NN-SRW-DFM on the artificial data. As for the linear structure, both the in-sample and out-of-sample distributions of these differences are very similar. Again, this is as expected since both models have the same structure for the trend part.

Table 6.7.: In-sample and out-of-sample distributions of the differences of the true spatial-temporal components and the spatial-temporal components estimated from the NL-SRW-DFM and NN-SRW-DFM on the artificial data.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
NL-SRW-DFM (in)	−0.018	0.046	−0.200	−0.048	−0.017	0.011	0.193
NL-SRW-DFM (out)	−0.016	0.045	−0.153	−0.043	−0.016	0.012	0.187
NN-SRW-DFM (in)	−0.017	0.047	−0.227	−0.048	−0.016	0.012	0.206
NN-SRW-DFM (out)	−0.014	0.046	−0.142	−0.043	−0.015	0.013	0.177

In Figure A.3 the factors of the true NL-SRW-DFM model, and of the NL-SRW-DFM and the NN-SRW-DFM estimated on the generated data, are plotted for every chain of the NUTS-algorithm. Likewise, in Figure A.4 the loadings of the true NL-SRW-DFM model, and of the NL-SRW-DFM and the NN-SRW-DFM estimated on the generated data are shown for every chain. As for the NL-SRW-DFM, the factors are not unique, as they are not the same for every chain. However, again there are two types of flavors. The factors of chain 3 in Figure A.3a are recovered by the NL-SRW-DFM in chains 3 and 4 and by the NN-SRW-DFM in chain 1.

Likewise, the factors of chains 1,2,4 in Figure A.3a are recovered by chains 1 and 2 in Figure A.3b and by chains 2, 3, and 4 in Figure A.3c. As for the L-SRW-DFM, the scales on the axis are different, and thus, the structure is replicated on another scale.

6.4. Replication of linear structure by the LGLMN

Another experiment is been performed on whether the LGLMN is able to replicate linear structures and their interpretability. The ability of the LGLMN to replicate non-linear structures is outside the scope of this thesis. Again, data is generated with the true L-SRW-DFM following (6.1), but this time on the whole data set, since the iterative model is more stable when using more data. Thereafter, both the L-SRW-DFM and the LGLMN-SRW-DFM are estimated on the generated data. Table A.9 shows the estimated coefficients of the true L-SRW-DFM and the L-SRW-DFM estimated on the generated data.

Table 6.8 shows the results of the L-SRW-DFM and the LGLMN-SRW-DFM on data generated from the L-SRW-DFM. The LGLMN-SRW-DFM performs quite well on the generated data since the performance measures are not far off from the performance measures of the L-SRW-DFM. Therefore, based on the performance measures, the LGLMN can handle the linear structure quite well.

Table 6.8.: In-sample and out-of-sample performance measures of the models on data simulated from the L-SRW-DFM.

	L-SRW-DFM	LGLMN-SRW-DFM
MAPE (in)	0.147	0.150
MAPE (out)	0.150	0.153
RMSE (in)	71,698	71,479
RMSE (out)	82,627	78,537
RMSE log price (in)	0.185	0.187
RMSE log price (out)	0.188	0.192

Since the data is generated from a model with a linear structure (the L-SRW-DFM), the LGLMN-SRW-DFM should replicate this linear structure, i.e. we should have

$$\beta_j(\mathbf{x_i}) \approx \hat{\beta}_j, \tag{6.5}$$

where $\beta_j(\mathbf{x_i})$ is the regression attention of variable j for transaction i for the LGLMN-SRW-DFM and $\hat{\beta}_j$ the estimated regression coefficient of the true L-SRW-DFM.

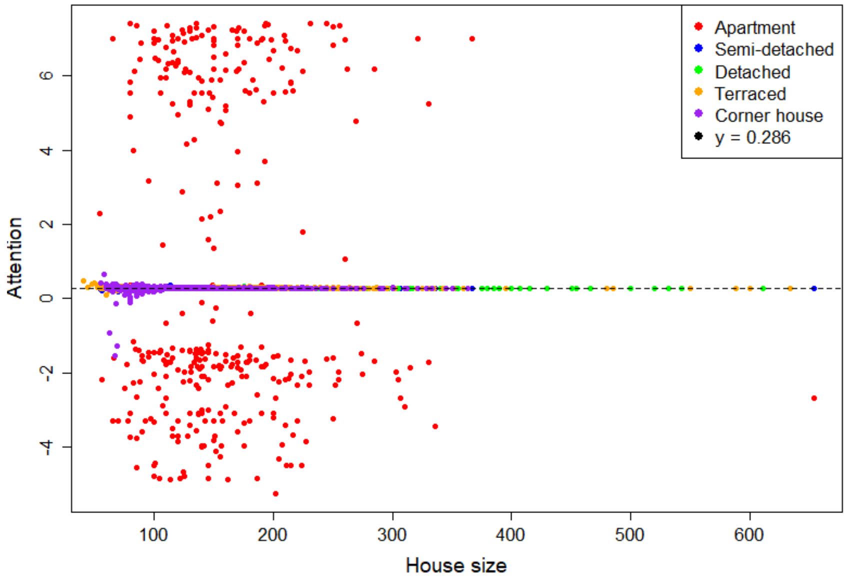
Figure 6.1a shows a plot of the regression attentions of the house size variable against the house size x_{ij} for every transaction i for different building types. Except for the building type 'Apartment', all regression attentions are around 0.286, which is the estimate of the coefficient of *house size* from the true L-SRW-DFM (See Table A.9). Hence, for the transactions that do not have building type 'Apartment', we have the desired property (6.5) with j being the index of the variable *house size*. Figure 6.1b shows a plot of the regression attentions of the *lot size* variable against the *lot size* x_{ij} for every transaction i for different building types with j being the index of the *lot is*. Except for the building type 'Apartment', all regression attentions are around 0.009, which is the estimate of the coefficient of the *lot size* from the true L-SRW-DFM (See Table A.9). The reason why the building type 'Apartment' does not have the (6.5) is probably due to unidentifiability since a *lot size* equal to zero implies that the building type is 'Apartment' and the other way around (Table A.2). For identifiability, not only the one-hot-encoded variable 'Apartment' should have been removed from the data set when estimating the models, but the one-hot-encoded building type variable in addition. Then the models can be uniquely identified, since the building type 'Apartment' will be incorporated in the model by the *lot size*.

Figure 6.2a shows box plots of the regression attentions of the different building types. Note that the regression attentions do not correspond with the estimated coefficients of the true L-SRW-DFM as shown in Table A.9. This is probably because the building types are not uniquely identified in the model. Figure 6.2b shows box plots of the regression attentions of the different building year classes. Note that the attentions are to a great extent around the coefficients of the building year classes of the true L-SRW-DFM as shown in Table A.9.

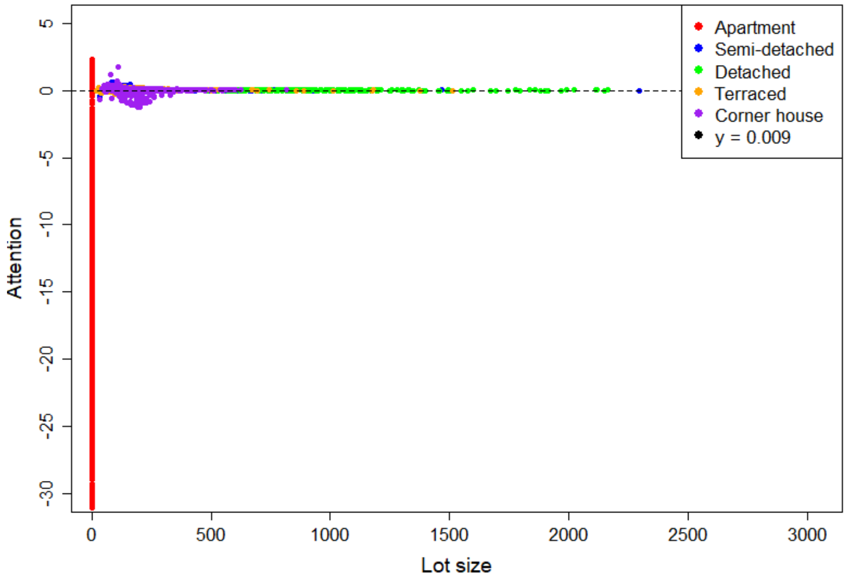
Figure 6.3 plots the interaction strength between house size (variable index j) and lot size (variable index k)

$$\frac{\partial}{\partial x_k}\beta_j(\mathbf{x_i}), \tag{6.6}$$

against the house size x_{ij} for all observations i. Since the data is generated from the L-SRW-DFM, there should be no interaction between the *lot size* and the *house size*. However, as shown in the plot, the interaction strength will increase with an increase in house size. Hence, the LGLMN-SRW-DFM does not perfectly replicate the structure with no interactions.

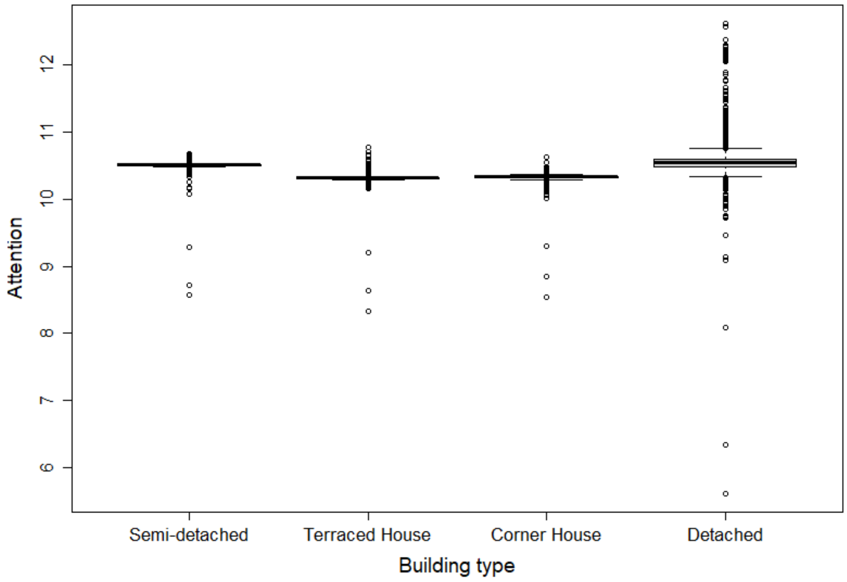


(a) Attentions of house size.

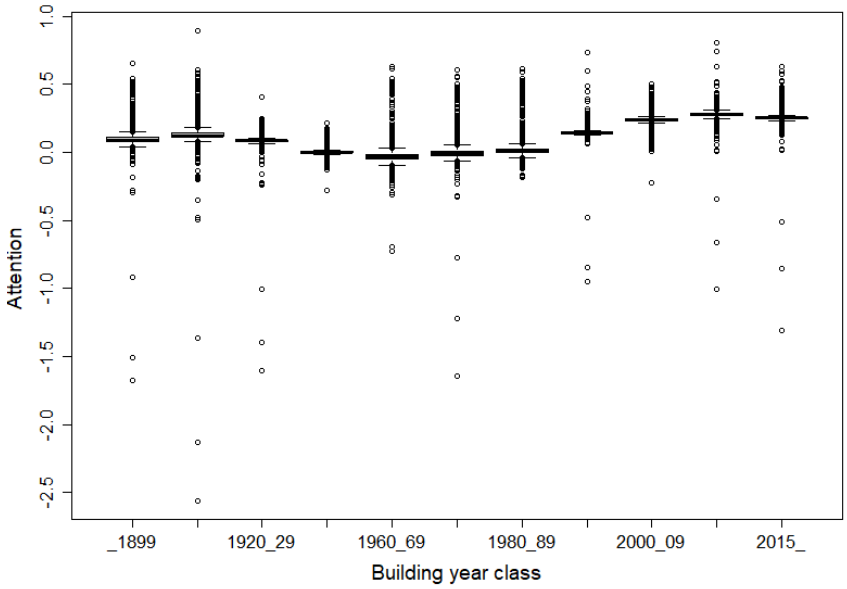


(b) Attentions of lot size.

Figure 6.1.: Feature contributions of the quantitative variables in the LGLMN-SRW-DFM based on data generated with the L-SRW-DFM.



(a) Attentions of building types.



(b) Attention of building year classes.

Figure 6.2.: Attentions of the categorical variables in the LGLMN-SRW-DFM based on data generated with the L-SRW-DFM.

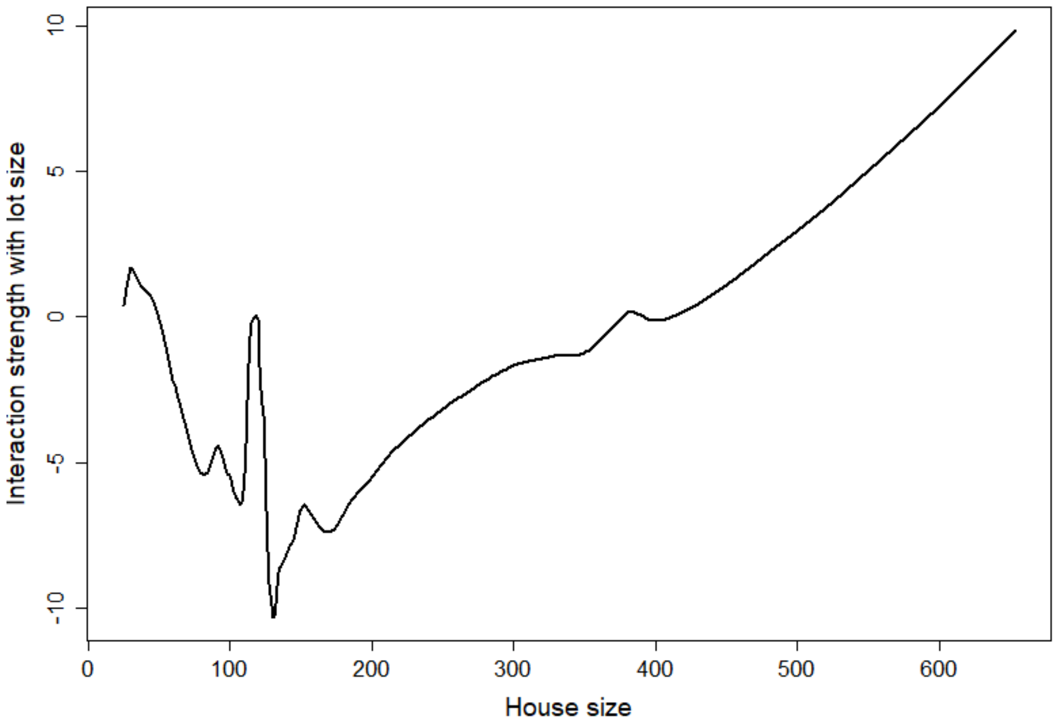


Figure 6.3.: Interaction strength of lot size over different values of house size in the LGLMN-SRW-DFM on data generated with the L-SRW-DFM.

7. Results

This section covers the assessment of the results of the models described in Chapter 4. Section 7.1 will cover the results of the performance measures on the models and Section 7.2 interprets the results of the structures of the LGLMN-SRW-DFM and the LGLMN-GRF-DFM.

An experiment with 5-fold CV on the L-SRW-DFM when using different numbers of dynamic factors showed the CV-value for the number of factors was approximately equal for $K \geq 2$. The results of this experiment are shown in Table A.11. Therefore, for evaluating the models in this chapter, 2 factors are used as it is computationally faster to use fewer factors and there is no significant improvement when using more than 2 factors. The variables *building type* and *building year class* are incorporated in the models as one-hot-encoded variables. Therefore, for identification purposes, the variables most-common of the most common observations for *building type* (which is the type *apartment*) and *building year class* (which is the class *1930-1944*) are removed from the training and test set. The geographical levels used in this section are based on neighborhoods as the results were the best when using this geographical level on the L-SRW-DFM. The models used in this research contain several hyper-parameters which can be found in Table A.12.

7.1. Performance Results

The L-SRW-DFM and NL-SRW-DFM are estimated on a normal computer, which took respectively 12 hours and 14 hours to estimate with the entire training set. The other models are computed on the HPC of DelftBlue because they are computationally heavier. The NN-SRW-DFM and the NN-GRF-DFM are estimated using the Bayesian approach with only one hidden layer for computational reasons. Furthermore, both models are also estimated using the algorithmic approach with four hidden layers, since extra layers do not add a lot of computational complexity when estimating the models using the algorithmic approach. The models using a LGLMN are only estimated using the algorithmic approach as they need at least two hidden layers to include both an FFNN structure in the first hidden layer and a linear structure in the last hidden layer.

Table 7.1 and Table 7.2 respectively show the in-sample and the out-of-sample performance of the models on the data from the municipality of Rotterdam described in Chapter 5. The models with a linear parametric structure for the property characteristics outperform the models with a non-linear parametric structure. The LGLMN-SRW-DFM estimated using an algorithmic approach performs the best on the in-sample performance measures, while the NN-SRW-DFM performs the best on the out-of-sample performance measures, except for the L-SRW-DFM and L-GRF-DFM, which perform better on out-of-sample RMSE. Furthermore, the models using a GRF as a spatial-temporal component show similar results to their analogs using an SRW. Note that the NN-SRW-DFM performs better when estimated using the algorithmic approach compared to when estimated using the Bayesian approach. This can be explained by two reasons. Firstly, the Bayesian approach only uses a limited number of iterations because of computational reasons, and therefore, using more iterations might improve the performance of the model. Secondly, the FFNN in the Bayesian approach only has one hidden layer for computational reasons, while the FFNN in the algorithmic approach has 4 hidden layers and therefore, it might be able to capture more complex structures. Observe that, the models using an SDFM based on the SRW perform about as well as the models using an SDFM based on the GRF, and sometimes even slightly better. Therefore, the 2-dimensional spatial structure of the SDFM based on the GRF does not seem to have a lot of added value compared to the 1-dimensional structure of the SRW.

Table 7.1.: In-sample performance measures of the models on data from the municipality of Rotterdam.

	MAPE	RMSE	RMSE log price
L-SRW-DFM	0.129	66,614	0.168
NL-SRW-DFM	0.131	69,660	0.171
NN-SRW-DFM	0.126	66,379	0.167
NN-SRW-DFM (alg.)	0.125	64,799	0.164
LGLMN-SRW-DFM (alg.)	**0.124**	**61,731**	**0.162**
L-GRF-DFM	0.129	66,449	0.168
NL-GRF-DFM	0.131	69,526	0.171
NN-GRF-DFM	0.128	63,077	0.167
NN-GRF-DFM (alg.)	0.126	66,034	0.165
LGLMN-GRF-DFM (alg.)	0.127	65,226	0.166

Note: The models are estimated with the Bayesian estimation procedure except for the models with (alg.). The preferred model for each metric is in **bold**.

Table 7.2.: Out-of-sample performance measures of the models on data from the municipality of Rotterdam.

	MAPE	RMSE	RMSE log price
L-SRW-DFM	0.129	**65,685**	0.171
NL-SRW-DFM	0.133	68,648	0.174
NN-SRW-DFM	0.129	65,786	0.169
NN-SRW-DFM (alg.)	**0.128**	66,101	**0.168**
LGLMN-SRW-DFM (alg.)	0.130	66,704	0.172
L-GRF-DFM	0.131	65,555	0.171
NL-GRF-DFM	0.133	68,521	0.174
NN-GRF-DFM	0.131	68,421	0.171
NN-GRF-DFM (alg.)	0.129	68,694	0.169
LGLMN-GRF-DFM (alg.)	0.133	72,969	0.171

Note: The models are estimated with the Bayesian estimation procedure except for the models with (alg.). The preferred model for each metric is in **bold**.

Appendix A.9 shows tables of the estimations of the parameters regarding the property characteristics, i.e. the spatial-temporal invariant parameters. The estimated parameters are in the 'mean' columns of the tables. Furthermore, quantiles of the posterior samples of the parameters are shown. As can be seen in the tables, the confidence in the parameters is high, as the standard deviation (StdDev) is low and the difference between the 5% and 95% quantiles is low in general. The $\hat{R}$-values are close to 1 for all estimations of the parameters regarding property characteristics which indicates that the algorithm has converged to these values. Note that the parameters of the L-SRW-DFM are similar to those of the L-GRF-DFM. Likewise, the estimated parameters of the NL-SRW-DFM and the NL-GRF-DFM are comparable.

In Table A.15, the results of the L-SRW-DFM based on districts are shown and the results of the L-SRW-DFM include neighborhood random effects. The L-SRW-DFM therefore includes neighborhood information but only uses trends on a district level. Note that the L-SRW-DFM performs worse when using districts compared to when using neighborhoods in the SDFM, and therefore, it is good to incorporate neighborhood information in the models. Furthermore, note that the L-SRW-DFM on districts with neighborhood random effects performs worse than the L-SRW-DFM on neighborhoods. Therefore, it is better not to include information on neighborhoods as random effects but to include neighborhood information in the SDFM to capture the price dynamics of the neighborhoods.

7.2. Interpretation of the LGLMN Results

7.2.1. LGLMN-SRW-DFM

The variable importance of the quantitative variables is measured by (3.11). The variable importance of *house size* is 3.37 and the importance of *lot size* is 8.81. Therefore, the variable *lot size* is more important for the LGLMN-SRW-DFM than the *house size*. This may be because *lot size* contains a lot of information since a *lot size* equal to zero implies that the house is an apartment and high *lot size* indicates a house is presumably a detached house.

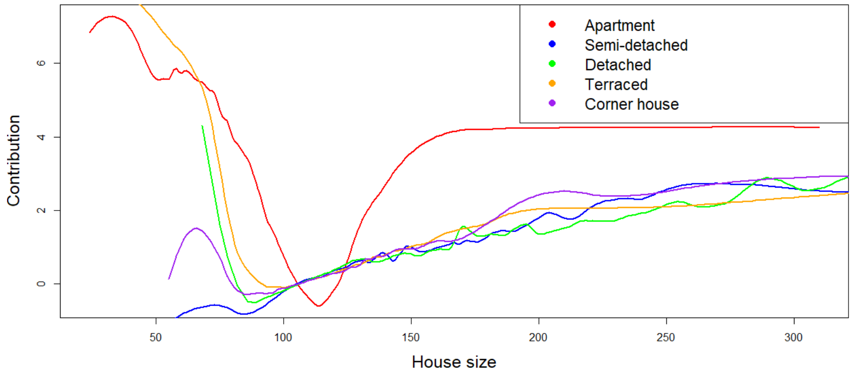
The feature contribution is the product of the regression attention and the variable, i.e.

$$\beta_j(\mathbf{x_i})x_{ij}, \tag{7.1}$$

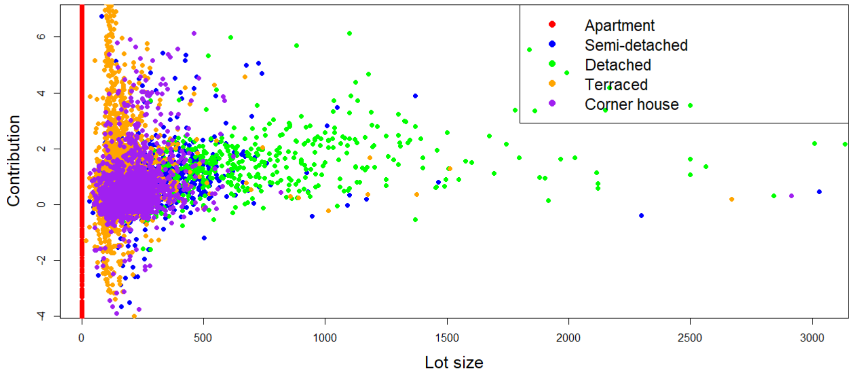
with j the index of the variable and i the observation. Figure 7.1 shows the feature contributions of the *house size* and the *lot size* estimated by the LGLMN-SRW-DFM. For all building types, the contribution of *house size* starts high, then gets lower, and subsequently increases again. The increase in contribution is what should be expected since a higher house size should contribute more than the selling price of the house. The decrease at the start is not what is expected but that can be explained by a small number of transactions with a very low house size. Note that, apart from the house type 'Apartment', the contribution for all house types is quite similar. This suggests that the regression attention for *house size* is independent of the building types. The dissimilar behavior of the contribution of *house size* for the building type apartment can be explained by the fact that the model is not uniquely identifiable for the building type apartment, as described in Section 6.4.

There is no clear pattern between the *lot size* and the contribution of the different building types. However, observe that the contribution of the *lot size* on the house price is mostly positive, especially when the *lot size* is high.

Figure 7.3 shows the feature contributions of the categorical variables estimated by the LGLMN-SRW-DFM. The median of the contribution of all building types is positive, but a lot of observations are negative since the 25% quantile is negative. The contributions of the building year classes are quite decisive in a positive or negative contribution since for most

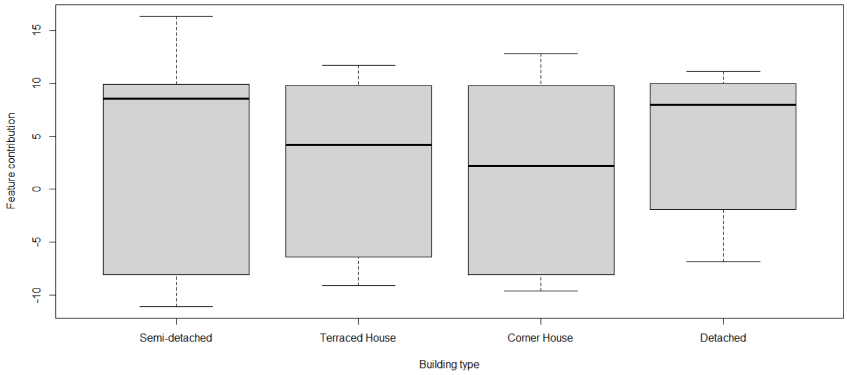


(a) Feature contribution of house size.

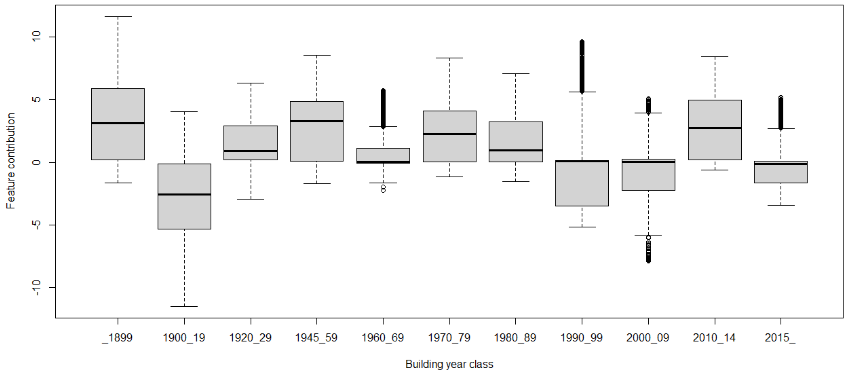


(b) Feature contribution of lot size.

Figure 7.1.: Feature contributions of the quantitative variables in the LGLMN-SRW-DFM.

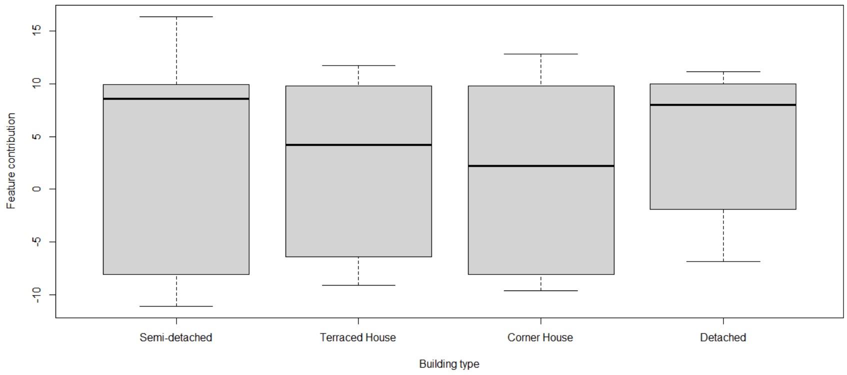


(a) Feature contributions of building types.

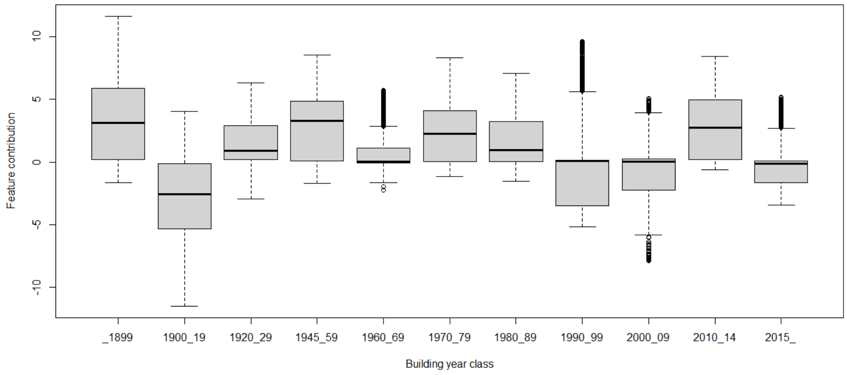


(b) Feature contributions of building year classes.

Figure 7.2.: Feature contributions of the categorical variables in the LGLMN-SRW-DFM.



(a) Feature contributions of building types.



(b) Feature contributions of building year classes.

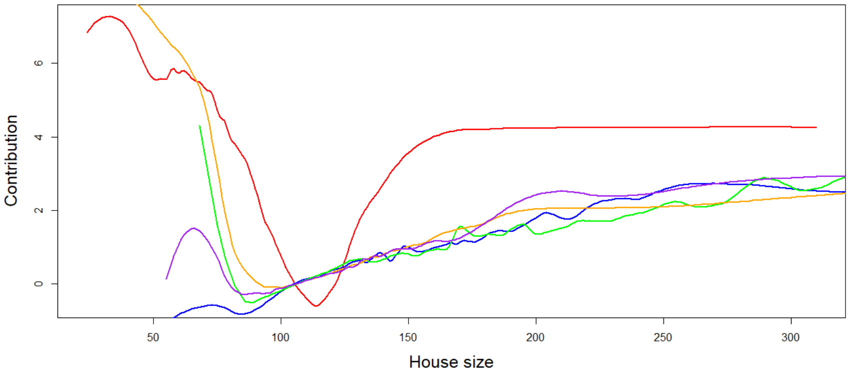
Figure 7.3.: Attentions of the categorical variables in the LGLMN-SRW-DFM.

building year classes, the 25% quantile of the contribution is above zero or the 75% quantile of the contribution is below zero.

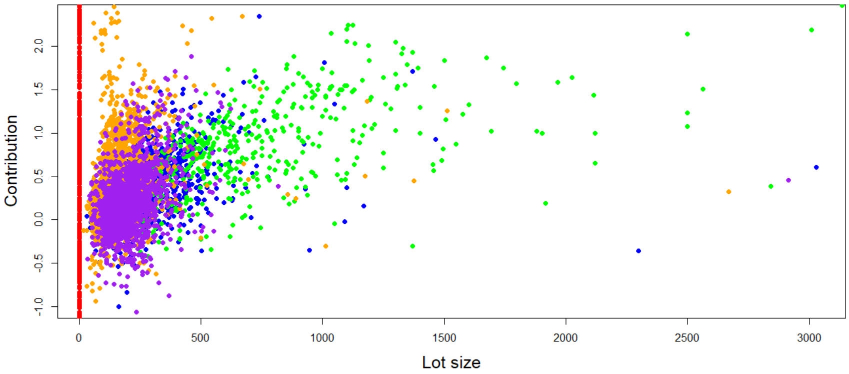
7.2.2. LGLMN-GRF-DFM

The variable importance of *house size* is 3.28 and the importance of *lot size* is 6.93. Therefore, as with the LGLMN-SRW-DFM, the variable *lot size* is more important for the LGLMN-GRF-DFM than the *house size*.

Figure 7.4 shows the feature contributions of the *house size* and the *lot size* estimated by the LGLMN-GRF-DFM. As with the LGLMN-SRW-DFM, for all building types, the contribution of *house size* starts high, then decreases, and subsequently increases again. As for the

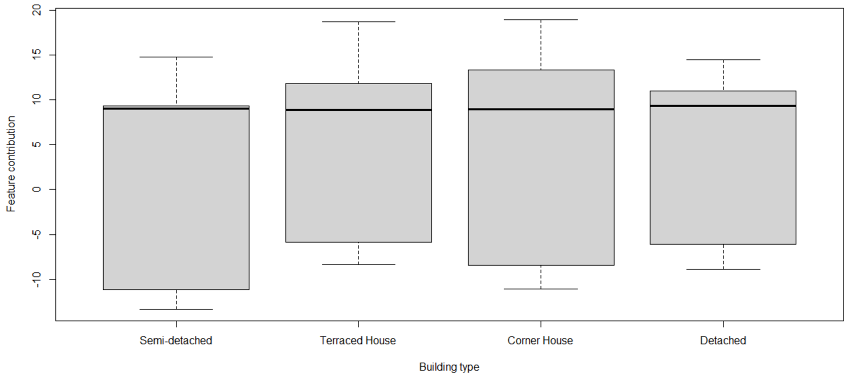


(a) Feature contribution of house size.

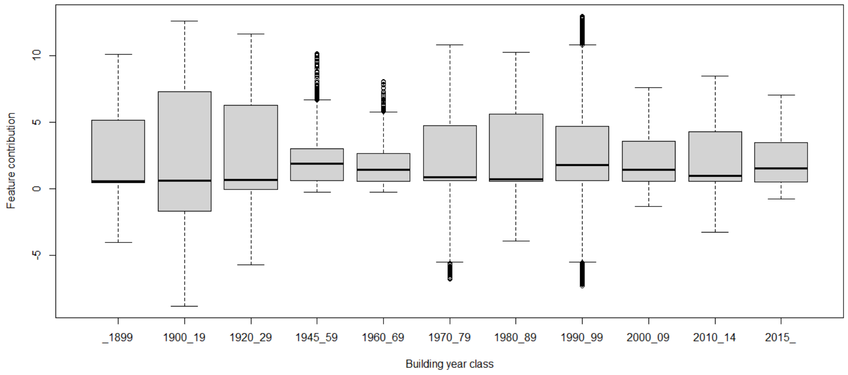


(b) Feature contribution of lot size.

Figure 7.4.: Feature contributions of the quantitative variables in the LGLMN-GRF-DFM.



(a) Feature contributions of building types.



(b) Feature contributions of building year classes.

Figure 7.5.: Feature contributions of the categorical variables in the LGLMN-GRF-DFM.

LGLMN-SRW-DFM, there is no clear pattern between the *lot size* and the contribution of the different building types.

Figure 7.5 shows the feature contributions of the categorical variables estimated by the LGLMN-GRF-DFM. As with the LGLMN-SRW-DFM, the median of the contribution of all building types is positive, but a lot of observations are negative since the 25% quantile is negative. Moreover, the contributions of the building year classes are quite decisive in a positive or negative contribution since for most building year classes, the 25% quantile of the contribution is above zero or the 75% quantile of the contribution is below zero.

7.2.3. Remarks on Interpretability of LGLMN

The results from the models using the LGLMN do show some patterns. The contribution of the *house size* increases when the *house size* increases. Furthermore, the LGLMN is quite decisive in whether a *building year class* positively or negatively affects the house price. No clear patterns are shown on the contributions of the *lot size* and the interactions along the variables. Furthermore, the LGLMN structure can give insights into how variables contribute and interact, yet no strong quantification can be made on how an increase/decrease of a specific variable affects the housing price.

8. Discussion

8.1. Conclusion

Data on real estate transactions is very sparse. That is, when considering a small time scale and location, e.g. quarterly transactions on a neighborhood level, there are a lot of spatial-temporal combinations with few or no observations. Nevertheless, price dynamics over fine geographies are highly desirable. This research provides a solution to the problem of spatial-temporal sparse real estate data by incorporating spatial-temporal dependencies using an SDFM, while including property characteristics in a data-driven way.

The natural logarithm of the selling price is modeled by the sum of a spatial-temporal component and a property characteristics component. Two different types of SDFMs are used as spatial-temporal components. Both parametric structures (linear and non-linear), and data-driven driven structures (an FFNN and an interpretable adjustment of the FFNN, the LGLMN), are used as property characteristics components. One SDFM is based on the SRW and has a 1-dimensional spatial structure, while the other SDFM is based on the GRF and has a 2-dimensional spatial structure. The estimation of the FFNN or the LGLMN using the NUTS-algorithm is computationally very heavy. Therefore, both an algorithmic procedure and the HPC of DelftBlue are used to estimate the models. The models are estimated with a Bayesian estimation procure and with an algorithmic procedure that algorithmically estimates the property characteristics component and the spatial-temporal component of the log price. In Chapter 6 a simulation study was performed to assess the ability of a neural network to capture linear and non-linear structures when combining it with an SDFM in a Bayesian representation. The neural network was able to properly replicate both linear and non-linear structures and to clone the dynamic factors (although on another scale). Furthermore, the LGLMN was able to replicate the linear structure quite well, although the models were not uniquely identifiable.

The models with a linear parametric structure for the property characteristics outperform the models with a non-linear parametric structure. Furthermore, the results of the models with a GRF as a spatial-temporal component performed similarly as their analogs using a SRW to capture spatial-temporal dependencies. Therefore, it is recommended to use the SRW as a spatial-temporal component as this is computationally faster to compute. The NN-SRW-DFM that was estimated with the iterative approach performed better than its linear and non-linear analogs. Hence, the NN-SRW-DFM might be the preferred choice of the models for the prediction of real estate selling prices. However, a drawback of this method is that it is uninterpretable, and therefore, one might still prefer to use the LGLMN-SRW-DFM or the L-SRW-DFM instead. Hence, the recommended method depends on a trade-off between accuracy on the one hand and speed and interpretability on the other.

8.2. Limitations and Future Research

The training of the models that involved an FFNN took a lot of computational power. This might not be a large drawback as the prediction using the model on new data is fast and does not have to be estimated on a daily basis. Although when considering larger data sets, which is quite obvious to do as this data set only considers data of one municipality, this might be a problem.

Because of the large amount of computational power needed for the models, only 400 warm-up and 400 sampling iterations were used on 4 different chains to estimate the models. The prediction accuracy of a model might improve when using more iterations, but this will lead to an even longer time to estimate the models.

In this research, some assumptions have been made that applied to all models based on results from the L-SRW-DFM. Based on an experiment in which the L-SRW-DFM is estimated with different numbers of dynamic factors, the choice was made to use 2 factors as it was optimal when considering accuracy and estimation time. When using the SRW with another model for property characteristics or an SDFM based on the GRF, using 2 factors might not be the optimal number, and therefore, these models might improve when using a different number of factors. Likewise, other assumptions based on results of the L-SRW-DFM were the usage of 5 building types instead of 16, and using neighborhoods as a geographical level instead of districts. These choices could have turned out differently if the experiments were performed on the other models. Since the models were computationally heavy, this was outside the scope of this thesis. Future research could analyze whether the accuracy of the model improves when using a different number of factors.

The models in this research are analyzed only on data from the municipality of Rotterdam. Therefore, when considering only the neighborhoods in this municipality an SDFM with a 1-dimensional spatial structure, might be enough to capture the relevant spatial information. However, when considering data from a larger geography, e.g. a country, an SDFM with a 2-dimensional structure might be more suitable as it might be harder to capture spatial information in one path. Future research can investigate whether more complex spatial structures (e.g. the SDFM based on the GRF) are preferable when using a larger geography.

The models in this thesis can only be used to make predictions of house selling prices and not to make forecasts. Future research could focus on the extension of these models to also make forecasts.

Bibliography

Abidoye, R. B. and Chan, A. P. (2018). Hedonic valuation of real estate properties in nigeria: Hedonic valuation of properties in nigeria. *Journal of African Real Estate Research*, 3(1):122–140.

Amari, S.-i. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5):185–196.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Bidanset, P. E. and Lombard, J. R. (2014). Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *Cityscape*, 16(3):169–182.

Bottero, M., Bravi, M., Mondini, G., and Talarico, A. (2017). Buildings energy performance and real estate market value: An application of the spatial auto regressive (sar) model. *Appraisal: From Theory to Practice: Results of SIEV 2015*, pages 221–230.

Bourassa, S. C., Haurin, D. R., Haurin, J. L., Hoesli, M., and Sun, J. (2009). House price changes and idiosyncratic risk: The impact of property characteristics. *Real Estate Economics*, 37(2):259–278.

Brunauer, W., Lang, S., and Umlauf, N. (2013). Modelling house prices using multilevel structured additive regression. *Statistical Modelling*, 13(2):95–123.

Cafarella, M., Ehrlich, G., Gao, T., Haltiwanger, J. C., Shapiro, M. D., and Zhao, L. (2023). Using machine learning to construct hedonic price indices. Technical report, National Bureau of Economic Research.

Calainho, F. D., van de Minne, A. M., and Francke, M. K. (2022). A machine learning approach to price indices: Applications in commercial real estate. *The Journal of Real Estate Finance and Economics*, pages 1–30.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.

Chau, K. W. and Chin, T. (2003). A critical review of literature on the hedonic price model. *International Journal for Housing Science and its applications*, 27(2):145–165.

Chiarazzo, V., Caggiani, L., Marinelli, M., and Ottomanelli, M. (2014). A neural network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia*, 3:810–817.

Delft High Performance Computing Centre (DHPC) (2022). DelftBlue Supercomputer (Phase 1). `https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase1`.

Fan, F., Xiong, J., and Wang, G. (2020). On interpretability of artificial neural networks. *arXiv preprint arXiv:2001.02522*, 2(10).

Fan, K., Peng, W., et al. (2005). Real estate indicators in hong kong sar. *Press & Communications CH-4002 Basel, Switzerland E-mail: publications@ bis. org Fax:+ 41 61 280 9100 and+ 41 61 280 8100*, page 124.

Firstenberg, P. M., Ross, S. A., and Zisler, R. C. (1988). Real estate: the whole story. *Journal of Portfolio Management*, 14(3):22.

Francke, M., Rolheiser, L., and Van de Minne, A. (2022). A spatial dynamic factor repeat sales model for real estate. *Available at SSRN 4260650*.

Francke, M. and Van de Minne, A. (2021). Modeling unobserved heterogeneity in hedonic price models. *Real Estate Economics*, 49(4):1315–1339.

Francke, M. and Van de Minne, A. (2023). Combining algorithmic and stochastic data models. *MIT Center for Real Estate Research Paper*, (23/05).

Francke, M. K. and van de Minne, A. M. (2017). Land, structure and depreciation. *Real Estate Economics*, 45(2):415–451.

Francke, M. K. and Vos, G. A. (2004). The hierarchical trend model for property valuation and local price indices. *Journal of Real Estate Finance and Economics*, 28(2-3):179.

Gamerman, D., Lopes, H. F., and Salazar, E. (2008). Spatial dynamic factor analysis.

Geltner, D. (2015). Real estate price indices and price dynamics: an overview from an investments perspective. *Annual Review of Financial Economics*, 7:615–633.

Geltner, D., Miller, N. G., Clayton, D. J., and Eichholtz, P. (2001). *Commercial real estate analysis and investments*, volume 1. South-western Cincinnati, OH.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Goh, A. T. (1995). Back-propagation neural networks for modeling complex systems. *Artificial intelligence in engineering*, 9(3):143–151.

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of urban economics*, 5(4):471–484.

Haider, M. and Miller, E. J. (2000). Effects of transportation infrastructure and location on residential real estate values: Application of spatial autoregressive techniques. *Transportation Research Record*, 1722(1):1–8.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Herath, S. and Maier, G. (2010). The hedonic price method in real estate and housing market research: a review of the literature.

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.

Hui, S. K., Cheung, A., Pang, J., et al. (2010). A hierarchical bayesian approach for residential property valuation: Application to hong kong housing market. *International Real Estate Review*, 13(1):1–29.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Khobragade, A. N., Maheswari, N., and Sivagami, M. (2018). Analyzing the housing rate in a real estate informative system: a prediction analysis. *Int. J. Civil Engine. Technol*, 9(5):1156–1164.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Lee, C. (2022). Enhancing the performance of a neural network with entity embeddings: an application to real estate valuation. *Journal of Housing and the Built Environment*, 37(2):1057–1072.

Lenstra, J. K. and Kan, A. R. (1975). Some simple applications of the travelling salesman problem. *Journal of the Operational Research Society*, 26(4):717–733.

Lopes, H. F., Gamerman, D., and Salazar, E. (2011). Generalized spatial dynamic factor models. *Computational Statistics & Data Analysis*, 55(3):1319–1330.

Malpezzi, S. et al. (2003). Hedonic pricing models: a selective and applied review. *Housing economics and public policy*, 1:67–89.

Peng, H., Li, J., Wang, Z., Yang, R., Liu, M., Zhang, M., Yu, P., and He, L. (2021). Lifelong property price prediction: A case study for the toronto real estate market. *IEEE Transactions on Knowledge and Data Engineering*.

Peterson, S. and Flanagan, A. (2009). Neural network hedonic pricing models in mass real estate appraisal. *Journal of real estate research*, 31(2):147–164.

Ren, Y., Fox, E. B., and Bruce, A. (2017). Clustering correlated, sparse data streams to estimate a localized housing price index. *The Annals of Applied Statistics*, pages 808–839.

Richman, R. and Wüthrich, M. V. (2023). Localglmnet: interpretable deep learning for tabular data. *Scandinavian Actuarial Journal*, 2023(1):71–95.

Ripley, B. D. (2005). *Spatial statistics*. John Wiley & Sons.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, 82(1):34–55.

Schwann, G. M. (1998). A real estate price index for thin markets. *The Journal of Real Estate Finance and Economics*, 16:269–287.

Silver, M. and Graf, B. (2014). *Commercial property price indexes: problems of sparse data, spatial spillovers, and weighting*. International Monetary Fund.

Sirmans, G. S., MacDonald, L., Macpherson, D. A., Zietz, E. N., et al. (2006). The value of housing characteristics: a meta analysis. *Journal of Real Estate Finance and Economics*, 33(3):215.

Stan Development Team (2023). Stan.

Strickland, C., Simpson, D., Turner, I., Denham, R., and Mengersen, K. (2011). Fast bayesian analysis of spatial dynamic factor models for multitemporal remotely sensed imagery. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(1):109–124.

Tay, D. P. and Ho, D. K. (1992). Artificial intelligence and the mass appraisal of residential apartments. *Journal of Property Valuation and Investment*, 10(2):525–540.

Twomey, J. and Smith, A. (1995). Performance measures, consistency, and power for artificial neural network models. *Mathematical and computer modelling*, 21(1-2):243–258.

Uzut, O. and Buyrukoglu, S. (2020). Prediction of real estate prices with data mining algorithms. *Euroasia Journal of Mathematics, Engineering, Natural and Medical Sciences*, 8(9):77–84.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc (with discussion). *Bayesian analysis*, 16(2):667–718.

Wang, X., Wen, J., Zhang, Y., and Wang, Y. (2014). Real estate price forecasting based on svm optimized by pso. *Optik*, 125(3):1439–1443.

Worzala, E., Lenk, M., and Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*, 10(2):185–201.

A. Appendix

A.1. Linear Splines

The interval of the *lot size* of a house is $[0, \infty)$. This interval can be divided into multiple sub-intervals by choosing S numbers $s_1, ..., s_S$ such that:

$$[0, \infty) = [0, s_1) \cup [s_1, s_2) \cup ... \cup [s_S, \infty), \tag{A.1}$$

where S is the number of splines.

The splines are created by dividing the variable *lot size* into S sub-variables $ls_1, ..., ls_S$ computed by:

$$\begin{aligned} ls_1 &= \min(s_1, \text{lot size}), \\ ls_2 &= \min(s_2 - s_1, (\text{lot size} - s_1)^+), \\ &\vdots \\ ls_{S-1} &= \min(s_S - s_{S-1}, (\text{lot size} - s_{S-1})^+), \\ ls_S &= (\text{lot size} - s_S)^+ \end{aligned}$$

Note that the sub-variables $s_1, ..., s_S$ sum up to the *lot size*, i.e. lot size $= \sum_{s=1}^{S} ls_s$.

A.2. Building types

Table A.1.: Primary and secondary building types in the Rotterdam data set.

Primary building type	Secondary building type (FW type)
Detached	Detached (1)
Semi-detached	Semi-detached (2)
Semi-detached	Linked semi-detached (3)
Semi-detached	Half semi-detached house (5)
Terraced	Linked house (4)
Terraced	Terraced house (6)
Corner house	Corner house (7)
Corner house	End-of-terrace house (8)
Apartment	Gallery flat (9)
Apartment	Porch flat (10)
Apartment	Corridor flat (11)
Apartment	Maisonette (12)
Apartment	Ground floor apartment (13)
Apartment	Upstairs apartment (14)
Apartment	Apartment in a building with a shared entrance (15)
Apartment	Penthouse (51)

A.3. Quantitative variables per building type

Table A.2.: Summary statistics of quantitative variables of building type Apartment

Statistic	*N*	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	18,689	222	129	32	135	280	1,795
Log price (in EUR)	18,689	12.2	0.5	10.4	11.8	12.5	14.4
House size (in m^2)	18,689	87	30	24	65	102	310
Lot size (obs. > 0, m^2)	0	0	0	0	0	0	0

Table A.3.: Summary statistics of quantitative variables of building type Semi-detached

Statistic	*N*	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	1,159	460	248	92	290	560	2,500
Log price (in EUR)	1,159	12.9	0.5	11.4	12.6	13.2	14.7
House size (in m^2)	1,159	154	49	54	123	179	654
Lot size (obs. > 0, m^2)	1,159	303	149	32	214	363	1,790

Table A.4.: Summary statistics of quantitative variables of building type Terraced

Statistic	*N*	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	7,015	301	170	45	193	350	2,430
Log price (in EUR)	7,015	12.5	0.5	10.7	12.2	12.8	14.7
House size (in m^2)	7,015	127	40	40	104	140	634
Lot size (obs. > 0, m^2)	7,015	145	66	18	135	160	1,662

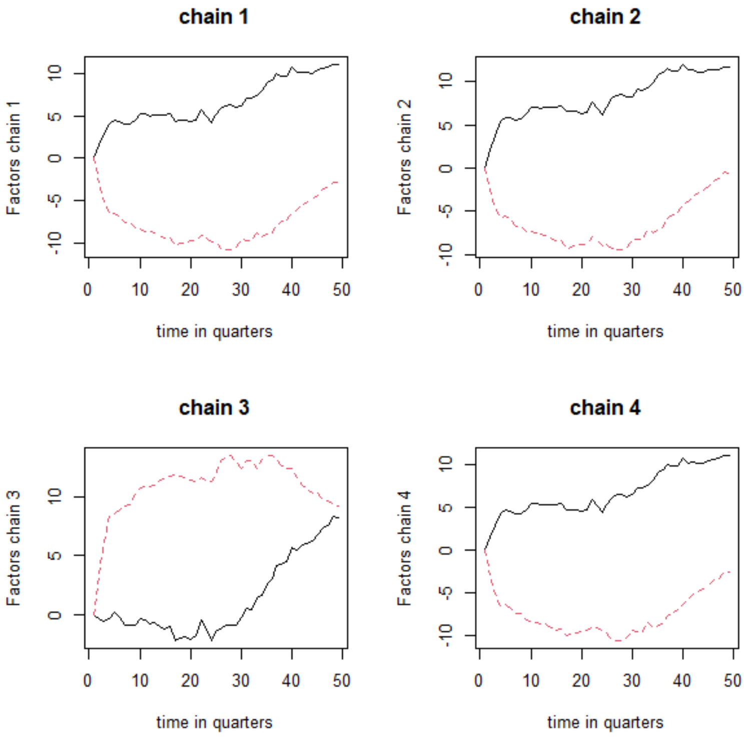
Table A.5.: Summary statistics of quantitative variables of building type Corner house

Statistic	*N*	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	2,477	324	187	84	205 0	384	1,925
Log price (in EUR)	2,477	12.6	0.5	11.3	12.2	12.9	14.5
House size (in m^2)	2,477	128	38	50	105	145	390
Lot size (obs. > 0, m^2)	2,477	200	84	33	186	242	818

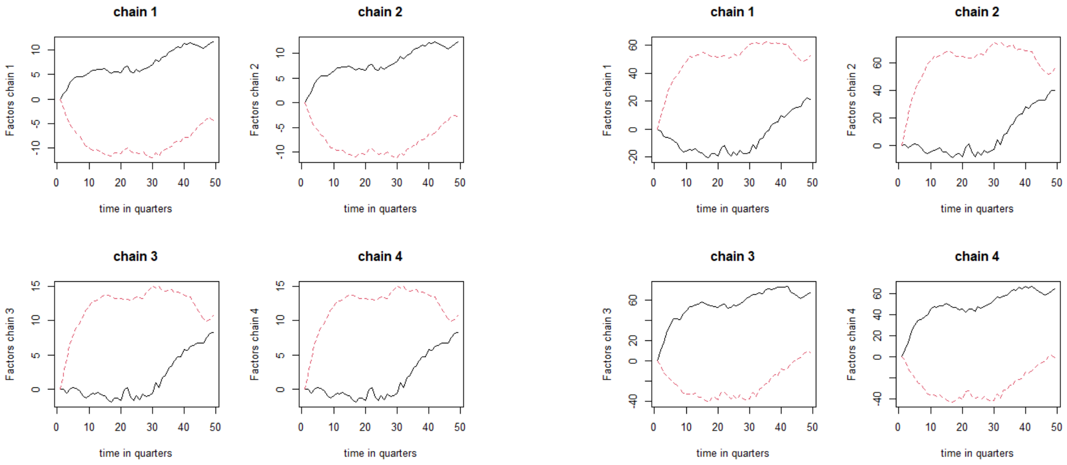
Table A.6.: Summary statistics of quantitative variables of building type Detached

Statistic	*N*	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Price (in 1000 EUR)	676	761	471	98	435	950	3,100
Log price (in EUR)	676	13.4	0.6	11.5	13.0	13.8	14.9
House size (in m^2)	676	205	85	58	150	245	750
Lot size (obs. > 0, m^2)	676	663	378	80	391	857	1,992

A.4. Multimodality of factors and loadings



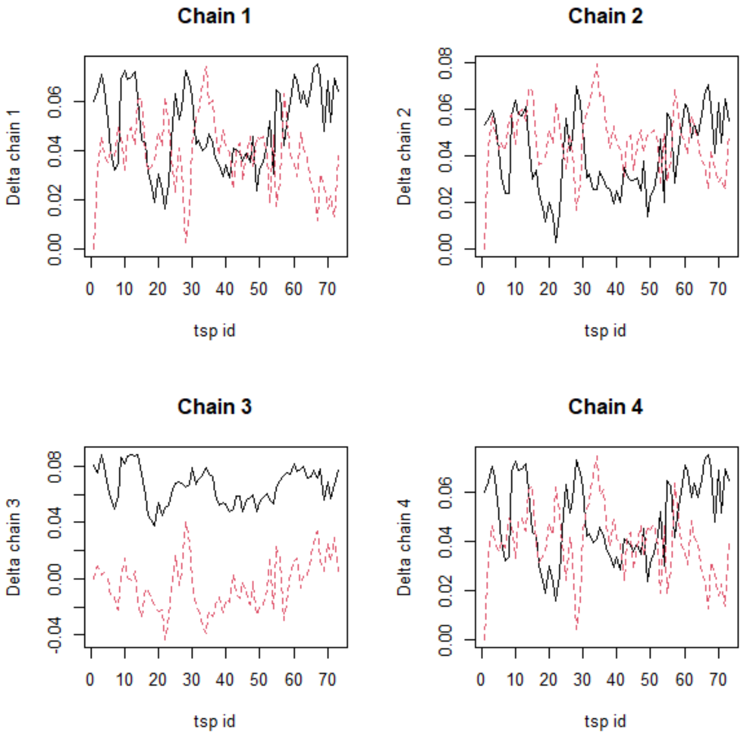
(a) Factors of base model: L-SRW-DFM.



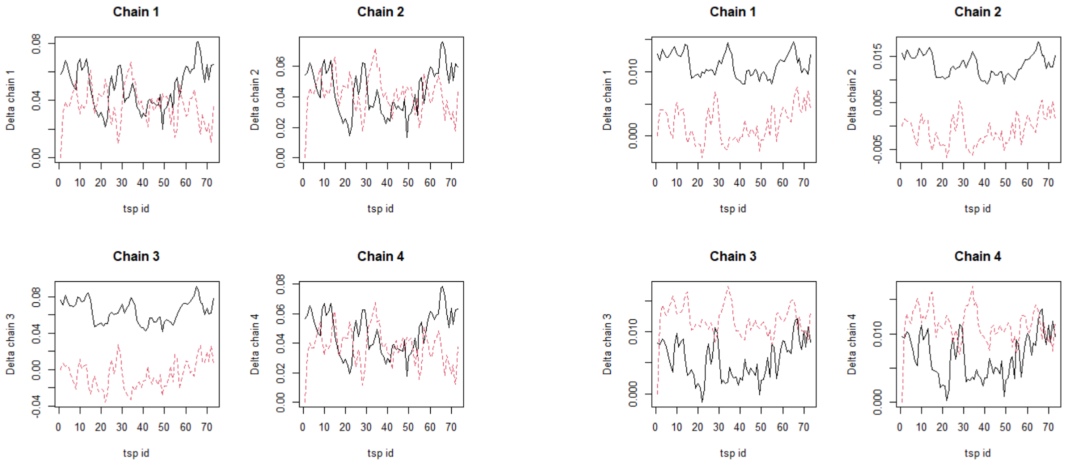
(b) Factors of L-SRW-DFM on generated data from base model.

(c) Factors of NN-SRW-DFM on generated data from base model.

Figure A.1.: Factors of the base model: L-SRW-DFM and of the L-SRW-DFM and NN-SRW-DFM estimated on data simulated from the base model.



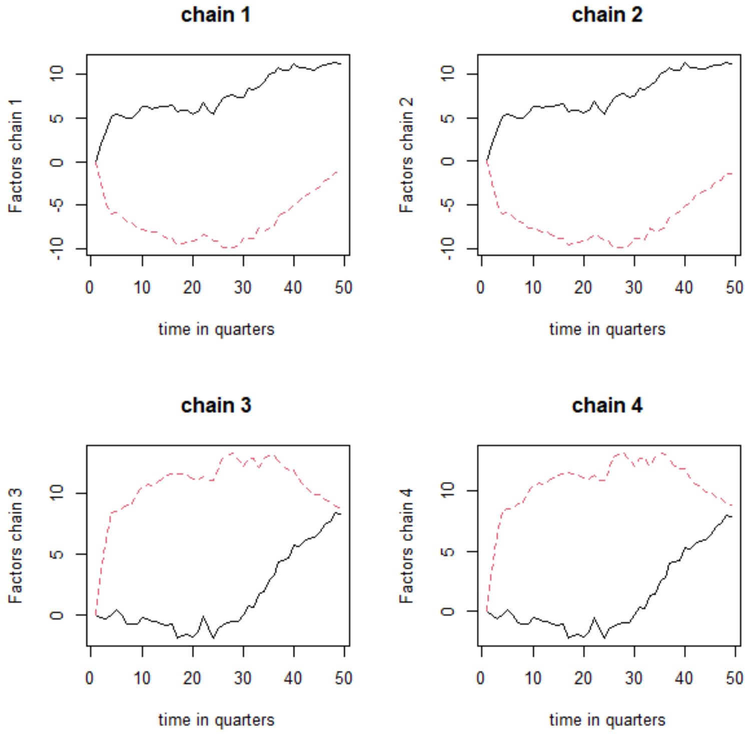
(a) Loadings of base: TSP SDFM lincomp



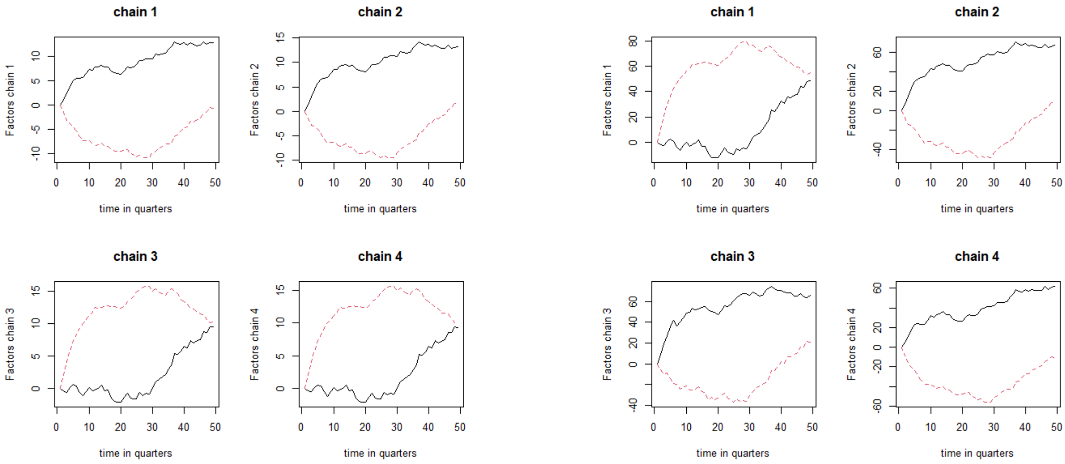
(b) Loadings of TSP SDFM lincomp on base

(c) Loadings of TSP SDFM NN on base

Figure A.2.: Loadings of the base model: TSP SDFM with linear components and of the models estimated on data simulated from the base model.



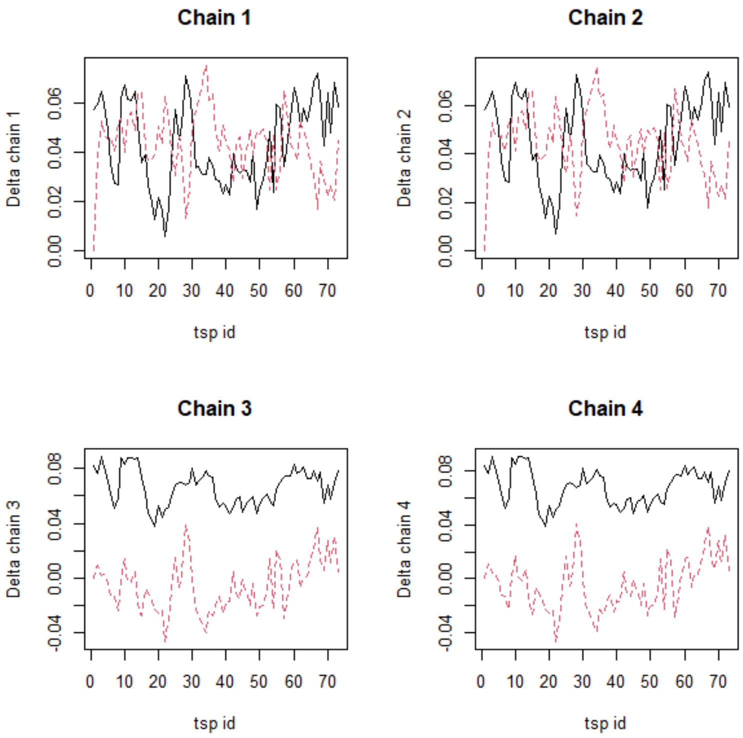
(a) Factors of base: TSP SDFM nonlincomp



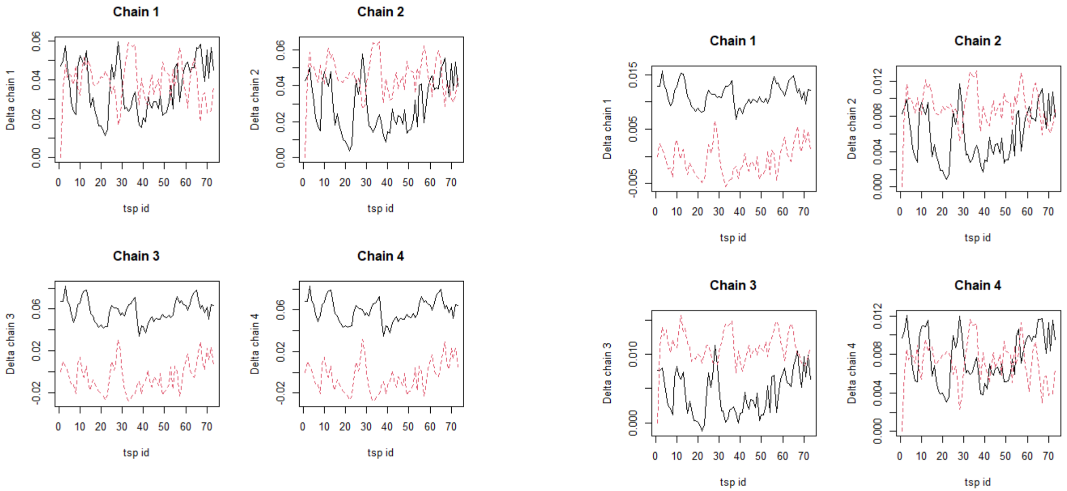
(b) Factors of TSP SDFM nonlincomp on base

(c) Factors of TSP SDFM NN on base

Figure A.3.: Factors of the base model: TSP SDFM with nonlinear components and of the models estimated on data simulated from the base model.



(a) Loadings of base: TSP SDFM nonlincomp



(b) Loadings of TSP SDFM nonlincomp on base

(c) Loadings of TSP SDFM NN on base

Figure A.4.: Loadings of the base model: TSP SDFM with nonlinear components and of the models estimated on data simulated from the base model.

A.5. Estimated Coefficients Simulation Study

Table A.7.: Estimates of the coefficients of the property characteristics component for the true L-SRW-DFM and the L-SRW-DFM estimated on the data generated from the DGP.

	True L-SRW-DFM			**Estimated L-SRW-DFM**		
	Mean	St. Dev.	$\hat{R}$	Mean	St. Dev.	$\hat{R}$
$\widehat{\alpha}$	8.334	0.086	1.000	8.209	0.083	1.001
$\widehat{\beta_1}$	0.815	0.016	1.000	0.843	0.016	1.001
$\widehat{\beta_2}$	-0.012	0.029	1.000	-0.034	0.029	0.999
$\widehat{\beta_3}$	-0.066	0.050	1.000	-0.056	0.049	1.000
$\widehat{\beta_4}$	0.002	0.016	0.999	0.027	0.016	0.999
$\widehat{\beta_5}$	-0.036	0.019	0.999	-0.063	0.019	1.000
$\widehat{\beta_6}$	0.041	0.014	1.000	0.060	0.014	1.000
$\widehat{\beta_7}$	0.006	0.023	1.000	0.007	0.022	1.000
$\widehat{\beta_8}$	0.138	0.056	1.000	0.241	0.054	1.000
$\widehat{\beta_9}$	0.087	0.034	0.999	0.083	0.034	1.000
$\widehat{\beta_{10}}$	0.078	0.023	1.000	0.071	0.023	1.000
$\widehat{\beta_{11}}$	0.011	0.019	1.000	0.022	0.019	1.000
$\widehat{\beta_{12}}$	-0.004	0.017	1.000	-0.002	0.017	0.999
$\widehat{\beta_{13}}$	-0.016	0.020	1.000	-0.013	0.020	1.000
$\widehat{\beta_{14}}$	-0.009	0.024	1.000	0.002	0.025	1.001
$\widehat{\beta_{15}}$	0.044	0.020	1.000	0.059	0.020	1.000
$\widehat{\beta_{16}}$	0.188	0.018	1.000	0.205	0.018	1.001
$\widehat{\beta_{17}}$	0.261	0.031	1.000	0.234	0.032	1.000
$\widehat{\beta_{18}}$	0.252	0.052	1.000	0.224	0.052	1.001
$\widehat{\nu}_1$	0.001	0.0001	0.999	0.001	0.0001	1.001
$\widehat{\nu}_2$	0.001	0.0001	1.000	0.001	0.0001	0.999
$\widehat{\nu}_3$	0.00003	0.0001	1.000	0.0002	0.0001	1.000

Table A.8.: Estimates of the coefficients of the property characteristics component for the true NL-SRW-DFM and the NL-SRW-DFM estimated on the data generated from the DGP.

	True NL-SRW-DFM			**Estimated NL-SRW-DFM**		
	Mean	St. Dev.	$\hat{R}$	Mean	St. Dev.	$\hat{R}$
$\widehat{\alpha}$	8.129	0.085	1.000	8.055	0.077	0.999
$\widehat{\beta}$	0.859	0.016	1.000	0.870	0.014	1.000
$\widehat{\gamma_1}$	-0.020	0.034	1.000	0.010	0.032	0.999
$\widehat{\gamma_2}$	-0.010	0.061	0.999	0.019	0.059	1.000
$\widehat{\gamma_3}$	0.004	0.016	0.999	-0.002	0.015	1.000
$\widehat{\gamma_4}$	-0.044	0.020	1.000	-0.034	0.019	1.000
$\widehat{\gamma_5}$	0.043	0.014	1.000	0.032	0.013	1.000
$\widehat{\gamma_6}$	0.011	0.023	0.999	0.047	0.023	0.999
$\widehat{\gamma_7}$	0.128	0.054	0.999	0.123	0.054	1.000
$\widehat{\gamma_8}$	0.096	0.036	1.000	0.143	0.034	0.999
$\widehat{\gamma_9}$	0.077	0.025	1.001	0.101	0.023	0.999
$\widehat{\gamma_{10}}$	0.004	0.021	1.001	0.007	0.019	1.000
$\widehat{\gamma_{11}}$	0.005	0.017	1.001	-0.006	0.016	1.000
$\widehat{\gamma_{12}}$	-0.002	0.022	1.001	0.004	0.020	0.999
$\widehat{\gamma_{13}}$	0.009	0.026	1.001	0.059	0.025	1.000
$\widehat{\gamma_{14}}$	0.058	0.022	1.001	0.046	0.020	1.000
$\widehat{\gamma_{15}}$	0.144	0.021	1.002	0.166	0.020	1.000
$\widehat{\gamma_{16}}$	0.201	0.019	1.002	0.222	0.018	0.999
$\widehat{\gamma_{17}}$	0.291	0.034	1.000	0.313	0.032	0.999
$\widehat{\gamma_{18}}$	0.280	0.056	1.001	0.350	0.053	1.001
$\widehat{\nu_1}$	0.053	0.006	0.999	0.047	0.006	0.999
$\widehat{\nu_2}$	0.153	0.016	1.000	0.154	0.017	0.999
$\widehat{\nu_3}$	0.032	0.015	1.000	0.053	0.017	1.001

Table A.9.: Estimate of the coefficients of the true L-SRW-DFM and the L-SRW-DFM estimated on the data generated from the true L-SRW-DFM used for the experiment of the LGLMN-SRW-DFM.

	True L-SRW-DFM			**Estimated L-SRW-DFM**		
Coefficient (variable)	Mean	St. Dev.	$\hat{R}$	Mean	St. Dev.	$\hat{R}$
α (intercept)	12.112	0.015	0.998	12.137	0.015	0.999
β_1 (house size)	0.286	0.002	1.000	0.285	0.002	1.003
β_2 (lot size)	0.009	0.002	1.000	0.006	0.002	0.999
β_3 (semi-detached)	0.317	0.008	1.000	0.323	0.008	0.998
β_4 (terraced)	0.146	0.004	0.998	0.150	0.004	1.000
β_5 (corner house)	0.209	0.006	0.998	0.206	0.006	0.999
β_6 (detached)	0.386	0.011	0.999	0.390	0.011	1.000
β_7 (- 1899)	0.060	0.011	0.998	0.059	0.011	0.998
β_8 (1900 - 1919)	0.071	0.008	0.999	0.085	0.007	0.999
β_9 (1920 - 1929)	0.044	0.007	1.003	0.038	0.007	1.002
β_{10} (1945 - 1959)	-0.041	0.005	0.999	-0.047	0.005	0.999
β_{11} (1960 - 1969)	-0.082	0.007	0.999	-0.080	0.006	0.998
β_{12} (1970 - 1979)	-0.027	0.009	0.999	-0.016	0.009	0.999
β_{13} (1980 - 1989)	0.030	0.007	1.000	0.035	0.007	0.999
β_{14} (1990 - 1999)	0.123	0.007	0.999	0.126	0.007	0.999
β_{15} (2000 - 2009)	0.205	0.006	0.998	0.210	0.006	1.001
β_{16} (2010 - 2014)	0.241	0.011	0.999	0.245	0.011	0.999
β_{17} (2015 -)	0.211	0.017	0.999	0.223	0.018	0.999

A.6. Experiments on Number of Factors and Building types

Table A.10.: CV values for the SRW-DFlin when using the 5 primary building types and 16 secondary building types.

	Using 5 primary building types	Using 16 secondary building types
CV value	4, 196, 972, 208	4, 216, 531, 198

Table A.11.: CV values on an experiment with 5-fold CV to see how many dynamic factors to use for the SRW-DFlin. The model is estimated for a different number of factors $K = 1,...,6$.

K = 1	K = 2	K = 3	K = 4	K = 5	K = 6
4, 518, 623, 937	4, 196, 972, 208	4, 189, 297, 819	4, 208, 286, 544	4, 207, 119, 291	4, 210, 591, 623

A.7. Hyper-parameters

Table A.12.: Hyper-parameters of the Bayesian estimation procedure.

Hyper-parameter	Hyper-parameter value
Number of factors	2
Number of knots of FFNN with 1 hidden layer	20

Table A.13.: Settings of the NUTS-algorithm.

Hyper-parameter	Hyper-parameter value
Number of sampling iterations	500
Number of warm-up iterations	500
Maximum tree-depth	12
Number of chains	4

Table A.14.: Hyper-parameters of the iterative estimation procedure.

Hyper-parameter	Hyper-parameter value
Number of layers	4
Number of knots per layer	20
Batch size	500
Number of epochs	1000

A.8. Result for variations in spatial geographies

Table A.15.: In-sample and out-of-sample performance measures of the models with different spatial geographies on data from the municipality of Rotterdam.

	MAPE	RMSE	RMSE log price
SRW-DFlin on districts (in)	0.132	69,493	0.173
SRW-DFlin on districts with neighborhood RE (in)	0.129	66,956	0.171
SRW-DFlin on districts (out)	0.133	73,263	0.175
SRW-DFlin on districts with neighborhood RE (out)	0.129	71,477	0.170

A.9. Estimates of property characteristic component parameters

Table A.16.: Estimation of the time-invariant parameters of the L-SRW-DFM

Parameter	Mean	StdDev	5%	50%	95%	$\hat{R}$
α	8.465876	0.024439	8.425966	8.465645	8.505752	0.999266
β_1	0.792944	0.004651	0.785090	0.792902	0.800410	0.997994
β_2	-0.019612	0.013924	-0.041725	-0.019894	0.002785	1.002402
β_3	-0.081630	0.009670	-0.097117	-0.081743	-0.065547	1.002206
β_4	-0.081470	0.012092	-0.101050	-0.081566	-0.061924	1.002196
β_5	0.061691	0.015287	0.036158	0.061996	0.087193	1.002709
β_6	0.087161	0.009448	0.071021	0.087470	0.102240	0.997868
β_7	0.080840	0.006723	0.069933	0.081047	0.091669	0.999969
β_8	0.031871	0.006237	0.021577	0.031768	0.041969	1.000334
β_9	-0.016862	0.005028	-0.025319	-0.016793	-0.008430	1.002462
β_{10}	-0.059182	0.006325	-0.069556	-0.059169	-0.048773	1.001007
β_{11}	-0.028267	0.007660	-0.041101	-0.028326	-0.016523	1.000422
β_{12}	0.028435	0.006092	0.018298	0.028332	0.038674	1.002816
β_{13}	0.105162	0.006112	0.094873	0.105022	0.115210	0.999248
β_{14}	0.183835	0.005629	0.174697	0.183813	0.193226	1.001882
β_{15}	0.239947	0.009884	0.223671	0.239806	0.256080	1.002327
β_{16}	0.227928	0.016145	0.200733	0.228168	0.254610	0.997803
β_{17}	0.001415	0.000069	0.001301	0.001415	0.001527	1.002070
β_{18}	0.000792	0.000039	0.000728	0.000792	0.000854	0.999497
β_{19}	0.000022	0.000007	0.000011	0.000021	0.000032	1.000328

Table A.17.: Estimation of the time-invariant parameters of the NL-SRW-DFM

Parameter	Mean	StdDev	5%	50%	95%	$\hat{R}$
α	6.393192	0.035957	6.335853	6.393226	6.454386	1.000774
β	3.753167	0.022067	3.716420	3.753094	3.788175	1.001303
ν_1	0.393236	0.027489	0.347797	0.393487	0.439790	1.002019
ν_2	0.591594	0.031885	0.539096	0.591112	0.643140	1.000785
ν_3	0.029457	0.008645	0.015768	0.029255	0.044140	1.000463
γ_1	0.261361	0.015191	0.236784	0.261333	0.286874	1.001703
γ_2	0.089709	0.012875	0.068151	0.089917	0.110566	1.000181
γ_3	-0.034642	0.009601	-0.050414	-0.034368	-0.018866	1.000855
γ_4	-0.024717	0.011722	-0.044379	-0.024769	-0.005498	1.000643
γ_5	0.093402	0.010494	0.076401	0.093341	0.111043	0.999743
γ_6	0.086184	0.007504	0.073784	0.086285	0.098548	0.999926
γ_7	0.023754	0.007206	0.011719	0.023898	0.035592	1.002816
γ_8	-0.010279	0.005170	-0.018608	-0.010270	-0.001701	1.000978
γ_9	-0.050550	0.006543	-0.061280	-0.050400	-0.040027	1.000366
γ_{10}	-0.013757	0.008534	-0.027547	-0.013758	0.000782	1.001105
γ_{11}	0.034192	0.006808	0.023292	0.034002	0.045787	1.001167
γ_{12}	0.113470	0.006891	0.102000	0.113338	0.124946	1.002488
γ_{13}	0.194670	0.006090	0.184478	0.194691	0.204582	1.000824
γ_{14}	0.262041	0.010231	0.246026	0.261736	0.279329	1.000469
γ_{15}	0.253817	0.017377	0.224591	0.253765	0.282295	0.999588

Table A.18.: Estimation of the time-invariant parameters of the -LGRF-DFM

Parameter	Mean	StdDev	5%	50%	95%	$\hat{R}$
α	8.470578	0.025510	8.426654	8.471118	8.512071	1.000284
β_1	0.792208	0.004845	0.784406	0.792134	0.800485	1.000158
β_2	0.001409	0.000066	0.001302	0.001408	0.001518	1.001667
β_3	0.000787	0.000040	0.000723	0.000787	0.000852	0.999067
β_4	0.000026	0.000007	0.000015	0.000026	0.000037	0.999393
β_5	0.062667	0.015847	0.036059	0.063072	0.088504	1.000714
β_6	-0.018461	0.013440	-0.040348	-0.018458	0.003808	1.000711
β_7	-0.080479	0.009720	-0.096547	-0.080512	-0.064456	1.001513
β_8	-0.080252	0.011776	-0.100455	-0.080143	-0.061213	1.003226
β_9	0.086148	0.009930	0.069788	0.086014	0.102444	0.997857
β_{10}	0.079417	0.006951	0.068339	0.079241	0.091064	0.999555
β_{11}	0.031892	0.006324	0.021561	0.031816	0.042058	1.000601
β_{12}	-0.017416	0.004907	-0.025723	-0.017099	-0.009693	1.005887
β_{13}	-0.059991	0.006144	-0.070039	-0.059987	-0.049937	0.998604
β_{14}	-0.028956	0.007654	-0.041495	-0.028872	-0.015929	0.998432
β_{15}	0.028360	0.005907	0.018702	0.028508	0.037943	0.997961
β_{16}	0.103658	0.006244	0.093428	0.103840	0.113824	1.000050
β_{17}	0.182393	0.005477	0.173264	0.182317	0.191287	1.000725
β_{18}	0.238454	0.009471	0.223084	0.238243	0.254290	0.998496
β_{19}	0.226946	0.016480	0.200221	0.226781	0.255313	0.998300

Table A.19.: Estimation of the time-invariant parameters of the NL-GRF-DFM

Parameter	Mean	StdDev	5%	50%	95%	$\hat{R}$
α	6.403259	0.037513	6.342585	6.403067	6.466521	0.999858
β	3.747285	0.023110	3.709171	3.747202	3.785224	0.998931
ν_1	0.389489	0.027721	0.344995	0.388378	0.434364	1.001682
ν_2	0.584177	0.031343	0.534618	0.583628	0.637441	1.001505
ν_3	0.031529	0.009065	0.017404	0.031406	0.046898	0.999324
γ_1	0.261422	0.015194	0.236461	0.262121	0.286628	1.004477
γ_2	0.089860	0.012656	0.068639	0.089908	0.110410	1.000504
γ_3	-0.034116	0.009587	-0.050829	-0.034056	-0.018986	1.000926
γ_4	-0.024288	0.011807	-0.044082	-0.024226	-0.005667	1.000855
γ_5	0.093459	0.009648	0.078360	0.093357	0.109859	0.999075
γ_6	0.086004	0.007295	0.074251	0.085789	0.098174	0.999840
γ_7	0.023803	0.007053	0.012489	0.023669	0.035320	1.000262
γ_8	-0.010748	0.005157	-0.019443	-0.010892	-0.002192	1.001383
γ_9	-0.050705	0.006617	-0.061365	-0.050598	-0.040026	0.999362
γ_{10}	-0.013763	0.008327	-0.027549	-0.013855	-0.000242	0.999378
γ_{11}	0.034723	0.006858	0.022882	0.034709	0.046741	0.999493
γ_{12}	0.112497	0.006611	0.101649	0.112345	0.123414	1.001837
γ_{13}	0.193923	0.006009	0.184084	0.193929	0.203499	1.001193
γ_{14}	0.261267	0.010605	0.243996	0.261847	0.279009	0.998996
γ_{15}	0.254620	0.017761	0.224875	0.254638	0.282949	0.998775