Lucerne University of
Applied Sciences and Arts

HOCHSCHULE
LUZERN

Business

Spatially Varying Coefficients Models for Large Data with an Application to Real Estate Price Prediction

Swiss Real Estate Research Congress 2021 (online)

Prof. Dr. Fabio Sigrist
Lecturer & project leader
Institute of Financial Services Zug IFZ

20.05.2021

Hedonic Real Estate Pricing

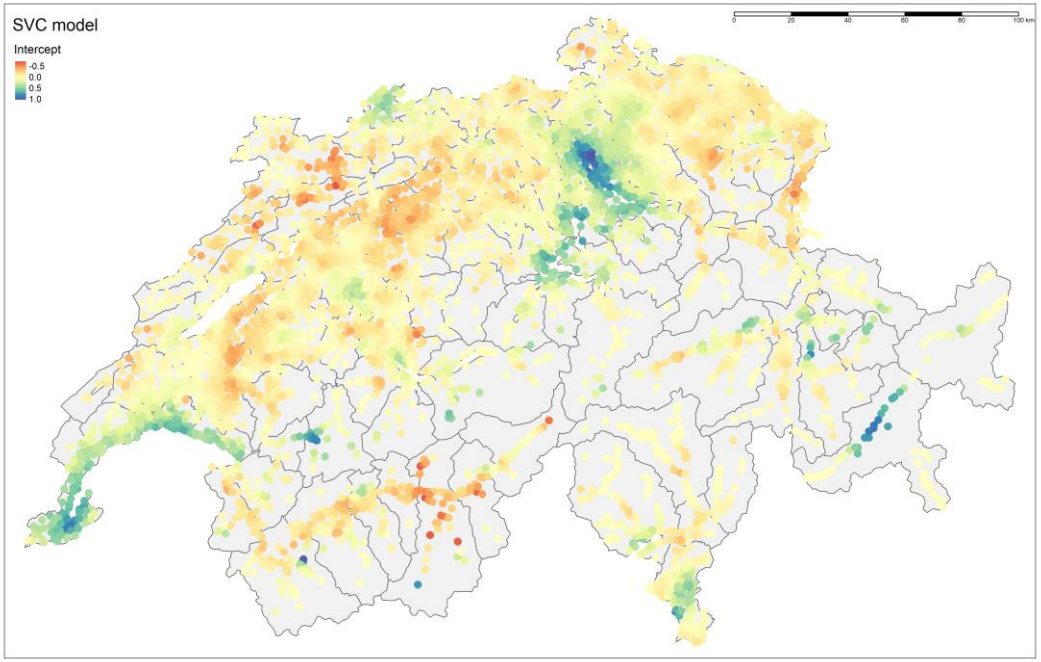
- Traditionally based on **linear regression model**

$$y_i = \beta^{(1)}(s_i) + \beta^{(2)}x_i^{(2)} + \ \ldots \ + \beta^{(p)}x_i^{(p)} + \epsilon_i$$

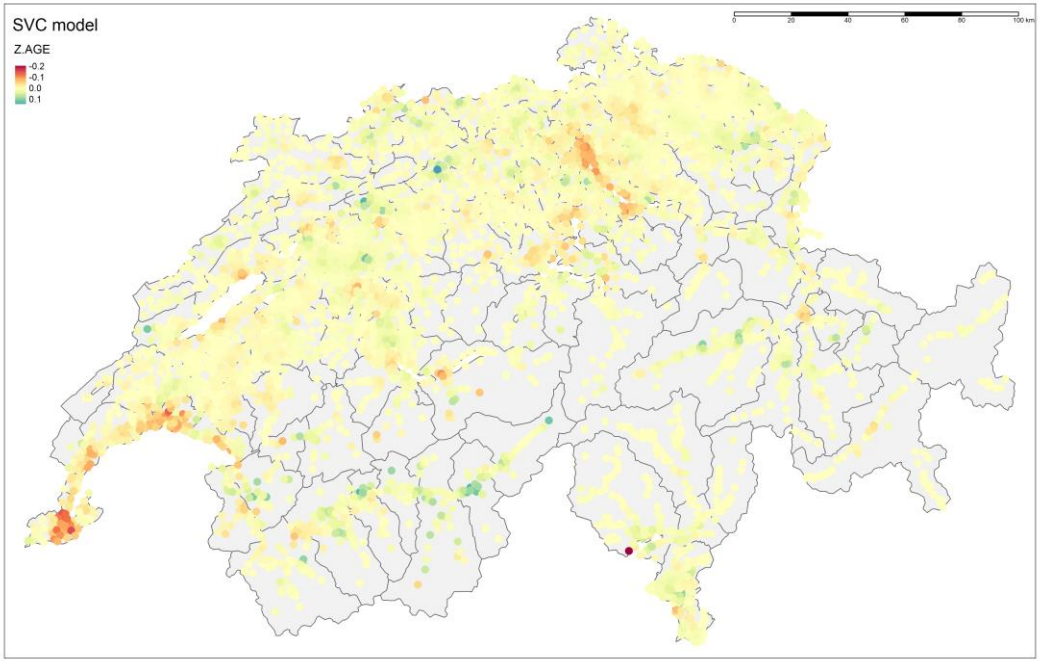
where
  - y_i: log(price) or log(rent)
  - $x_i^{(j)}$: properties of object i (size, age, standard, micro location, condition, etc.)
  - s_i: location of object i

- Strong conjecture from literature[6, 7] and domain experts that **coefficients $\beta^{(j)}$ are not constant over space**

Intercept (deviation from global mean)



Age Coefficient (deviation from global mean)



Spatially Varying Coefficients

- Linear regression model

$$y_i = \beta^{(1)}(s_i) + \beta^{(2)}x_i^{(2)} + \ \dots \ + \ \beta^{(p)}x_i^{(p)} + \epsilon_i$$

- Generalize to a **spatially varying coefficients (SVC) model**

$$y_i = \beta^{(1)}(s_i) + \beta^{(2)}(\boldsymbol{s}_i)x_i^{(2)} + \ \dots \ + \ \beta^{(p)}(\boldsymbol{s}_i)x_i^{(p)} + \epsilon_i$$

  - $\boldsymbol{s}_i$: location of observation i

Goal

- **Develop novel methodology for modelling of SVCs** that
  - **scales** in n (observations) and p (number of SVC)
  - is based on sound statistical methodology**: Gaussian processes** (GP)

- Innosuisse funded project in collaboration with Fahrländer Partner Raumentwicklung (FPRE), ZKB, BCV, and UBS

- Special thanks to **Jakob Dambon** (HSLU), **Reinhard Furrer** (UZH), **Manuel Lehner** (FPRE), **Jaron Schlesinger** (FPRE)

FP
RE

Existing Methods for Handling SVCs

Geographically Weighted Regression[1]

Very heuristic, bad properties

Bayesian SVC Processes[3]

$\boldsymbol{\beta}^{(j)}(\cdot) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$
Gaussian processes

- Gaussian Predictive Processes[4]
- Gaussian Markov Random Field Approximation[5]

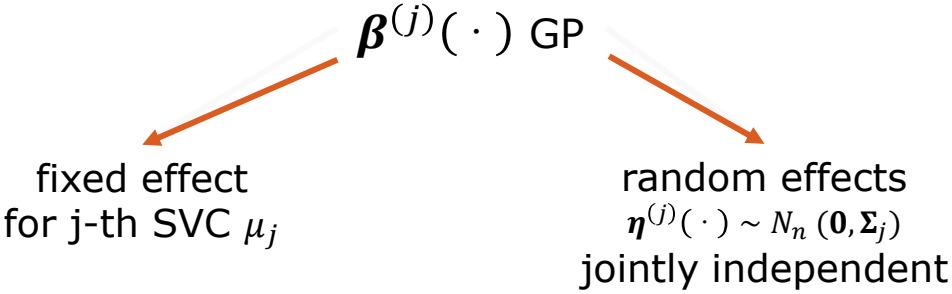
Existing model-based methods do not scale well

Eigenvector Spatial Filtering[2]

Not model-based, heuristic

Model-based Approach using Gaussian Processes (GP)

- Model SVCs $\boldsymbol{\beta}^{(j)}(\cdot)$ using Gaussian Processes (GP)



- Covariance matrix $\boldsymbol{\Sigma}_j$ parametrized by a range ρ_j and marginal variance σ_j^2

- Model in matrix notation

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\mu} + \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

- Joint GP $\boldsymbol{\eta}(\cdot) \sim N_{np}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_\mathrm{p}))$ with according parametrization
  - $\boldsymbol{W}$ SVC data matrix and $\boldsymbol{X}$ fixed effect data matrix

Maximum Likelihood Estimation

- Log-likelihood: $\log L(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\mu}) \propto \log\det \boldsymbol{\Sigma}_{\mathbf{y}} + (\boldsymbol{y} - \boldsymbol{X\mu})^{\top} \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1} (\boldsymbol{y} - \boldsymbol{X\mu})$

  where $\boldsymbol{\theta} = (\rho_1, \sigma_1^2, \ldots, \rho_p, \sigma_p^2, \tau^2)$ and $\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta}) = \boldsymbol{W\Sigma_{\eta}W}^{\top} + \tau^2 I_{n\times n}$

- Numeric optimization required. Challenges:
  - High-dimensional parameter space $(\boldsymbol{\theta}, \boldsymbol{\mu}) \in \boldsymbol{\Omega} = (0, \infty)^{2p+1} \times \mathrm{R}^{p}$
  - Computationally intensive tasks

Ad-hoc MLE does **not** scale in p nor n

Proposed Approach

- Use **profile likelihood** (generalized least squares estimate) for $\boldsymbol{\mu}$

$$\hat{\boldsymbol{\mu}} = \left(\boldsymbol{X}^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}(\hat{\boldsymbol{\theta}})\, \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Sigma}_{\boldsymbol{y}}^{-1}(\hat{\boldsymbol{\theta}})\, \boldsymbol{y}$$

Dimension reduction of parameter space

- Prior **independence assumption** on GPs

$$\boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta}) = \sum_{j=1}^{p} \boldsymbol{\Sigma}_j \odot \left(\boldsymbol{x}^{(j)} (\boldsymbol{x}^{(j)})^\top\right) + \tau^2 I_{n\times n}$$

Efficient updating of $\boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta})$

- **Covariance matrix tapering**[8]

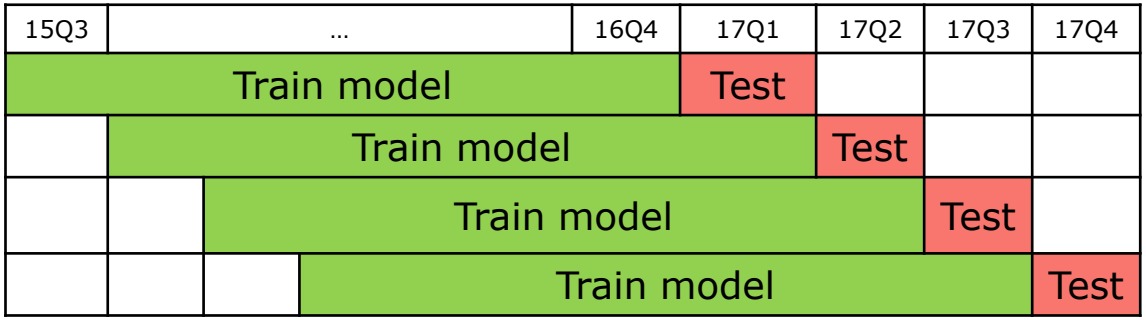
$$\boldsymbol{\Sigma}^{taper} \odot \boldsymbol{\Sigma}_{\boldsymbol{y}}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{taper} \odot \left(\sum_{j=1}^{p} \boldsymbol{\Sigma}_j \odot \left(\boldsymbol{x}^{(j)} (\boldsymbol{x}^{(j)})^\top\right)\right) + \tau^2 I_{n\times n}$$

Sparse matrices to speed up Cholesky decomposition

- Further regularize likelihood using priors[9, 10]

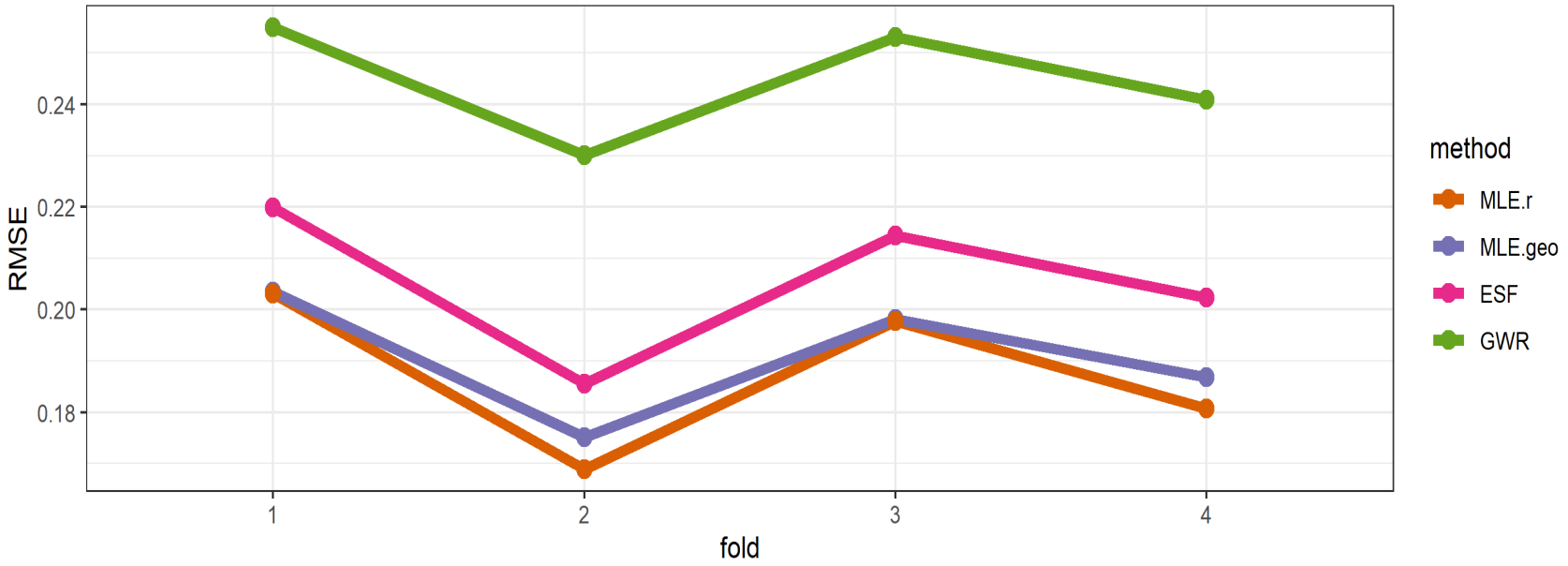
Comparison Using Cross-Validation

- Transaction prices for apartments in Switzerland
- Temporal "cross-validation" for model comparison

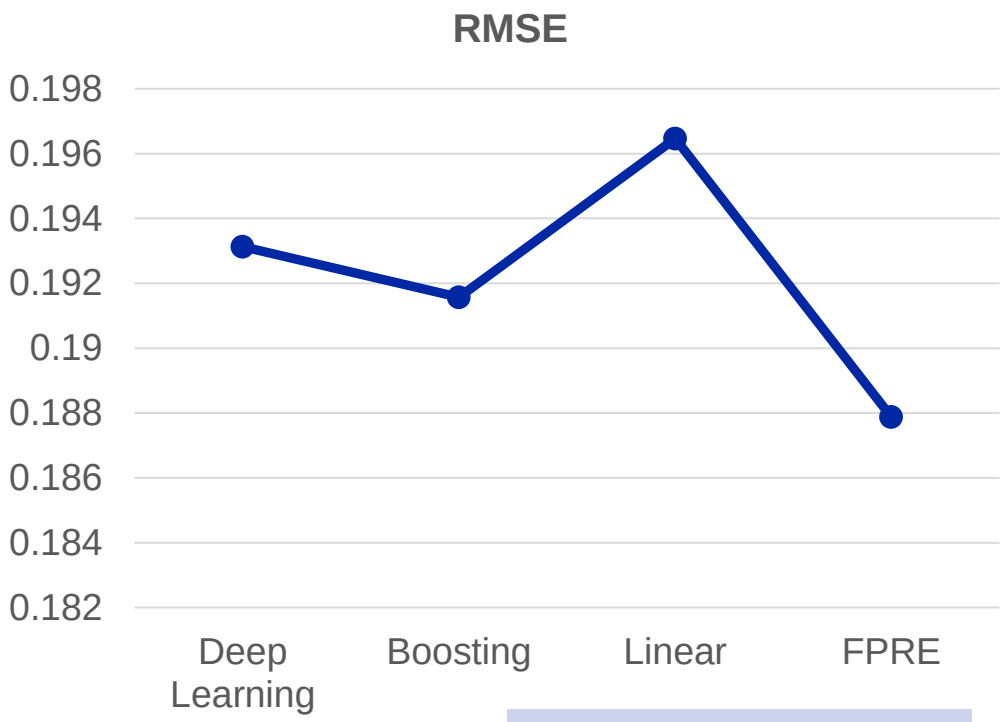


$p = 8$
$n \approx 15'000$

Out-of-sample RMSE



What About Machine Learning (ML) and Deep Learning?



Black-box ML is not the answer!

Comments

- **RMSE over 4 quarters**
- Models
  - **Linear**: linear model with spatial fixed effects (dummies for locations)
  - **FPRE**: linear model with spatial modelling used by FPRE (spline based but using additional historical data)
  - **Deep Learning**[11]: large amount of network architectures and tuning parameters considered
- *Note*: Slightly different data and linear model specifications compared to above results

Thank you for your attention!

- R-package **`varycoef`**

  https://github.com/jakobdambon/varycoef

- Article: Dambon, Sigrist, and Furrer (2021)[12]

References

(1) Fotheringham *et al.* (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. *Wiley.*
(2) Murakami and Griffith (2018). Spatially varying coefficient modeling for large datasets: Eliminating N from spatial regressions. *arXiv: 1807.09681*.
(3) Gelfand *et al.* (2003). Spatial Modeling with Spatially Varying Coefficient Processes. *JASA* **98**, 387-396.
(4) Banerjee *et al.* (2008). Gaussian Predictive Process Models for Large Spatial Data Sets. *JRSS: Series B* **70**, 825 - 848.
(5) Lindgren *et al.* (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *JRSS: Series B* **73**, 423 - 498.
(6) Cao *et al.* (2019). A Big Data Based Geographically Weighted Regression Model for Public Housing Prices: A Case Study in Singapore. *Ann. Am. Assoc. Geogr.* **109**, 173 - 186.
(7) Geng *et al.* (2011). Geographically Weighted Regression Model (GWR) based Spatial Analysis of House Price in Shenzhen. In *2011 19th International Conference on Geoinformatics*, pp. 1 - 5.
(8) Furrer *et al.* (2006). Covariance Tapering for Interpolation of Large Spatial Datasets. *J. Comput. Graph. Stat.* **15**, 502 - 523.
(9) Simpson *et al.* (2017). Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statist. Sci.* **32**, 1 - 28.
(10) Fuglstad *et al.* (2018). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *JASA* **114**, 445 - 452.
(11) Walthert and Sigrist (2019). Deep Learning for Real Estate Price Prediction. Available at SSRN 3393434
(12) Dambon, J. A., Sigrist, F., and Furrer, R. (2021). Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics*, *41*, 100470.