**varycoef:**
An R Package to Model Spatially Varying Coefficients

Jakob A. Dambon1,2, Fabio Sigrist2, Reinhard Furrer1,3

1Department of Mathematics, University of Zurich
2Institute of Financial Services Zug, Lucerne University of Applied Sciences and Arts
3Department of Computational Science, University of Zurich

1. Introduction

In regression models for spatial data, it is often assumed that the marginal effects on the response are constant over space. In practice, this assumption might often be questionable. Spatially varying coefficient (SVC) models extend the linear regression model, by allowing the coefficients to vary spatially. The underlying model equation is

\[ y(s) = \beta_0(s) + \beta_1(s) x_1(s) + \ldots + \beta_p(s) x_p(s) + \epsilon(s), \]

where \( s \) is a location in a domain \( D \subseteq \mathbb{R}^2 \). This model offers a higher degree of flexibility while being easy to interpret.

2. Data and Motivation

Our novel methodology is motivated by a data set of apartment transactions in Switzerland containing the selling price, six covariates and approximated coordinates for each transaction. Real estate mass appraisal is a classical application for SVC models. We define our model as

\[ \log(\text{price}) = \beta_1(\mathbf{x}) + \beta_2(\mathbf{x}) \log(\text{HNF03}) + \beta_3(\mathbf{x}) \text{AGE} + \beta_4(\mathbf{x}) \text{AGE}_2 + \beta_5(\mathbf{x}) \text{MICRO} + \beta_6(\mathbf{x}) \text{STAND} + \beta_7(\mathbf{x}) \text{RENOV} + \beta_8(\mathbf{x}) \text{OBS} + \epsilon, \]

training it on data from 6 consecutive quarters and giving predictions for a following 7th quarter. Tab. 1 gives an overview of the data and the SVCs. To the best of our knowledge, there exists no geostatistical methodology for our application as there are too many observations (ca. 1.5 x 10^4) and too many SVCs.

3. Methodology

Similarly to [Gelfand et al., 2003], we assume that the SVCs follow zero mean Gaussian processes with a stationary covariance structure. Each SVC is defined by

\[ \eta_j(\mathbf{s}) \sim N(0, \Sigma_j); \]

with \( \Sigma_j \) depending on some range \( \mu_j \) and variance \( \eta_j^2 \) parameters. Using two data matrices X and W, the general SVC model is given by

\[ y = X \beta + W \eta + \epsilon, \]

where \( \mu \) are the fixed effects, \( \eta \) are the joint SVCs modelling the random effects (RE), and \( \epsilon \) is the error term with no spatial structure. Further, we assume mutual prior independence of the Gaussian processes. The response \( y \) is normally distributed and we can state its likelihood function given some observed data \( X \) and \( W \). We use maximum likelihood estimation (MLE) to estimate the mean and covariance parameters. We apply covariance tapering as well as likelihood regularization suggested by [Furrer et al., 2006] and [Fuglstad et al., 2018], respectively, to deal with the large data set. Predictions for SVCs as well as the predictive variance can be computed using the empirical best linear unbiased predictor (EBLUP).

4. Availability

The new methodology has been implemented in the R package varycoef which is available on CRAN [Dambon et al., 2019].

5. Results

In an extensive simulation study we were able to show that our proposed novel methodology is one of few, if not the only one, to accurately estimate the covariance parameters of the Gaussian processes. We analyzed the predictive performance in the before mentioned real estate application using a temporal cross validation (CV) with 4 folds to compare our approach with other SVC methods such as GWR [Fotheringham et al., 2002] or ESF [Murakami and Griffith, 2015]. Fig.1 and Tab. 1 show the first fold results for our proposed MLE method. Fig. 2 shows the computed root-mean-squared errors (RMSE) over all CV folds.

### Conclusion and Future Work

Our new methodology offers a high degree of flexibility to model SVCs based on geostatistics. It scales to large data both in the number of observations and SVCs. The prediction accuracy outperforms most other available SVC methods. Future work includes runtime optimization, SVC selection, and the extension on spatio-temporal varying coefficients.

### Contact

jakob.dambon@math.uzh.ch
http://user.math.uzh.ch/dambon

### References

[Dambon et al., 2019] varycoef: Varying Coefficients, R package version 0.2.9. https://cran.r-project.org/package=varycoef


[Fuglstad et al., 2010] Constructing Priors that Penalize the Complexity of Gaussian Random Fields. JASA 105(491)


[Gelfand et al., 2003] Spatial Modeling with Spatially Varying Coefficient Processes. JASA 98(462)