Article:

Blomqvist, BRH, Sumpter, DJT and Mann, RP orcid.org/0000-0003-0701-1274 (2019) Inferring the dynamics of rising radical right-wing party support using Gaussian processes. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 377 (2160). ISSN 1364-503X

https://doi.org/10.1098/rsta.2019.0145

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

Inferring the dynamics of rising radical right-wing party support using Gaussian processes

Journal:	*Philosophical Transactions A*
Manuscript ID	RSTA-2019-0145.R2
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Blomqvist, Björn; Uppsala University, Department of Mathematics Sumpter, David; Uppsala University, Department of Mathematics Mann, Richard; University of Leeds, Department of Statistics, School of Mathematics; The Alan Turing Institute
Issue Code (this should have already been entered but please contact the Editorial Office if it is not present):	COUPLING
Subject:	Statistics < MATHEMATICS, Mathematical modelling < MATHEMATICS, Differential equations < MATHEMATICS
Keywords:	Bayesian statistics, Coupling functions, Gaussian processes, Radical right-wing parties

Author-supplied statements

Relevant information will appear here if provided.

Ethics

Does your article include research that required ethical approval or permits?:
This article does not present research with ethical considerations

Statement (if applicable):
CUST_IF_YES_ETHICS :No data available.

Data

It is a condition of publication that data, code and materials supporting your paper are made publicly available. Does your paper present new data?:
Yes

Statement (if applicable):
The Swedish National Council for Crime Prevention provided data for crime, and Statistics Sweden provided data for all the other variables. These data are publicly available from these sources. In addition, the code and data used in this paper are also attached as supplementary material.

Conflict of interest

I/We declare we have no competing interests

Statement (if applicable):
CUST_STATE_CONFLICT :No data available.

Authors' contributions

This paper has multiple authors and our individual contributions were as below

Statement (if applicable):
BRHB performed the data analysis. BRHB wrote the softvare, BRHB and RPM conceived the idea of the Gaussian process approach. All authors designed the study. BRHB drafted the manuscript and all authors then helped with the writing. All authors read and approved the manuscript.

rsta.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
Applied mathematics

Keywords:
Gaussian processes, Coupling functions, Radical right-wing parties, Bayesian statistics

Author for correspondence:
Richard P. Mann
e-mail: R.P.Mann@leeds.ac.uk

Inferring the dynamics of rising radical right-wing party support using Gaussian processes

Björn R. H. Blomqvist[1], David J. T. Sumpter[1] and Richard P. Mann[2,3]

[1]Department of Mathematics, Uppsala University, Uppsala, Sweden
[2]Department of Statistics, School of Mathematics, University of Leeds
Leeds, United Kingdom
[3]The Alan Turing Institute, London, United Kingdom

The use of classical regression techniques in social science can prevent the discovery of complex, non-linear mechanisms, and often relies too heavily on both the expertise and prior expectations of the data analyst. In this paper, we present a regression methodology that combines the interpretability of traditional, well used, statistical methods with the full predictability and flexibility of Bayesian statistics techniques. Our modelling approach allow us to find and explain the mechanisms behind the rise of Radical Right-wing Populist parties (RRPs), that we would have been unable to find using traditional methods. Using Swedish municipality level data (2002-2018) we find no evidence that the proportion of foreign-born residents is predictive of increases in RRP support. Instead, education levels and population density are the significant variables that impact the change in support for the RRP, in addition to spatial and temporal control variables. We argue that our methodology, which produces models with considerably better fit of the complexity and non-linearities often found in social systems, provides a better tool for hypothesis testing and exploration of theories about RRPs and other social movements.

1. Introduction

Radical right-wing populism is resurgent across European societies, posing an existential threat to established democratic systems. Radical Right-Wing populist parties (RRPs) share a common emphasis on ethnonationalism, rooted in myths about the past, and their programs are directed towards strengthening the nation by making it more ethnically homogeneous [1,2]. In recent decades, RRPs have grown in electoral support and are now established throughout Europe: e.g. in Poland, 'Law and Justice' (37.6%, 2015); in Hungary, 'Jobbik' (20.22%, 2016); in France, 'National Front', (13.02%, 2017); Austria, 'Freedom Party of Austria', (27%, 2017); and in Germany (2017), a nation without a RRP in the parliament in decades, 'Alternative for Germany' received 13% support. In Sweden, where RRPs have traditionally had very little support, one such party, the 'Sweden Democrats' obtained 17.53% of the votes in the last election in 2018, an increase from 12.86% in 2014. They became the third (out of 8) largest party in the Swedish Parliament, having more than tripled their support from 2010. Understanding the mechanisms behind their growth is not only of high importance to the political process in Sweden, but also crucial for understanding what is happening all across Europe, and also in places like Australia and in the United States, where populism is on the rise [1,3].

The two predominant theories for why RRPs experience increased support are the social marginality theory [4–8], which suggests a stronger RRP support in socially marginalized areas, and ethnic competition theory [5,9–15], which suggests that voters turn to RRPs because they want to reduce competition, both cultural and economic, with immigrants. We introduce and test these theories as the major competing explanations for increasing RRP support, but it is important to note that other mechanisms have been posited that go beyond the scope of this paper. For example, supply chain explanations, such as the political opportunity structure and party how how parties are organized [16]. Studies done across Europe have come to slightly different conclusions about the relative importance of the two major theories. Some examples supporting the social marginality hypothesis are: RRPs have been found to have a negative correlation with level of education in the populous [4]; workers and middle-class voters are over-represented among new supporters of RRPs [4,17,18]; unemployment is positively correlated with new RRP voters [4]; unemployment together with high share (more than 6.3%) of foreign-born residents in the population have a positive interaction effect on support [19]. Even though there is support for this theory, there are also conflicting results. For example, it has been shown that RRPs receive the most support in the mid-educational stratum [6,8]; unemployment levels have been found have non-significant [4,20] or even negative correlations [6,7] with RRP support. Support for ethnic conflict theory has been tested by investigating if RRPs are more prosperous in areas with a large immigrant population [5]. Previous results show a positive correlation between RRP support and the number of foreign-born within a country [4,7]; RRP support correlates positively with both the proportion of immigrants and asylum seekers [20,21]. However, in Rydgren (2008) [22], a positive correlation is found between the number of immigrants within a country and RRP support in the Netherlands and Denmark, but not in Austria, Belgium, France, and Norway; ethnic heterogeneity has also been reported to have a non-significant correlation to RRP support [18].

In 2011, Rydgren and Ruth [5] presented a meticulous study of the current Swedish RRP, in which the authors find support for both the social marginalization and ethnic conflict theory. Specifically, Rydgren & Ruth [5] found a significant negative correlation between both education level and gross regional product (GRP) per capita and a positive correlation with unemployment rates. They also found RRP support to be positively correlated with a high immigrant proportion from EU/EFTA countries, but negatively correlated with the immigrant ratio from both the Nordic countries and non-European countries.

It should be noted that both social marginalisation and ethnic competition may occur on many dimensions and any attempt to find support for each can be influenced by the choice of data and measurements, as well as contexts such as nationality. In some cases individual-level responses to ethnic competition may not be apparent at the aggregated level, for example if foreign-born residents comprise a substantial part of the voting population in a municipality, then their voting patterns may conflict with the response of other voters in ways that cannot be observed in aggregate. Throughout this paper we use data aggregated to the municipality level, and therefore acknowledge that a more complex reality at the individual level may be hidden by this constraint.

Previous studies in this area use standard regression techniques to analyse drivers of RRP support [23–29]. However, while they allow hypothesis testing within a standardized framework, traditional regression models are unlikely to provide the best fit of the data because they assume linearity or a particular chosen, often polynomial, form, thus missing things that do not correspond to this form, and perhaps falsely identify patterns that actually have a different form. They depend heavily on the analyst to pick the functional form and therefore rely on prior ideas about how the system works. Modelling is most useful where it is based on sound theories that underpin the model formulation. However, even where the analyst has a strongly motivated theory for how a variable of interest affects the system, they may not have equally well developed prior expectations for the effects of confounding variables that must also be controlled.

The challenge of best modelling the available data can be addressed using a Gaussian processes (GP) regression, which is today commonly used in machine-learning [30,31], but which has also been applied to problems in biology and social science (e.g [32,33]). GPs model the relationship between covariates non-parametrically allowing lots of flexibility for the approximating functions. However, GP regression lacks much of the interpretability of standard parametric regression. The advantage of parametric methods is that inferred coefficients tell us the importance of different factors, and the fitted models are simple to understand and to use. This presents a dilemma: do we use more flexible models that promise greater model fit, predictive power and a more data-driven approach, or do we prioritize interpretability by using simpler linear models?

In this paper, we address precisely this problem. We develop an approach that combines the interpretability of standard regression with the model fitting capabilities of GPs. We choose GPs from among the many possible, flexible approaches inspired by machine-learning (such as neural networks, Random Forests and generalised additive models), because they are (i) intrinsically Bayesian, allowing principled model selection via the marginal likelihood; and (ii) integrate seamlessly with classical linear regression methods when expressed in a Bayesian framework (as we will describe below). Specifically, we use a GP framework to perform Bayesian linear regression to find the best explicit model in the explanatory variables, the variables we wish to investigate, combined with a fully non-parametric statistical control for confounding variables. We present several ways in which we can adequately measure the relevance of different variables and explicitly test theories proposed in political science. We use this approach to model and investigate the rise of the Swedish RRP, the Sweden Democrats, using aggregated municipality-level data, and re-evaluate the predominant theories of RRP support.

We will focus on the dynamics of RRP support (i.e. changes over time), rather than stock values. A dynamical systems approach uses a non-linear differential or difference equation to describe the rate of change of each variable in a social system in terms of itself and other social variables [34–37]. For example, rate of change in RRP support can be fitted as a function of education, unemployment, immigration and so on. One advantage of this approach is that it allows social systems to be described by coupling functions [38], which can then be solved to better understand the dynamics of the social system [39–41]. A similar approach has been adopted in chemistry [42–44], neural science [45] and communications [46].

2. Modelling approaches

We consider data $\mathcal{D}=\{(\mathbf{x}^i,\mathbf{z}^i,dy^i)|i=1,...,N\}$, consisting of N observations of a social system, where $\mathbf{x}=[x_1,x_2,...,x_D]^\top$ denotes a D dimensional input vector of explanatory variables, which we wish to include as predictors in the form of an explicit polynomial function $f(\mathbf{x})$; $\mathbf{z}=[z_1,z_2,...,z_{D_*}]^\top$ denotes a D_* dimensional input vector of confounding variables, which we want to statistically control for, but not model as an explicit polynomial function; and target dy

represents the change over time of the response the variable y, i.e. $dy^t = y^{t+1} - y^t$. We denote the coupling between y and the other variables in the system as

$$dy = h(\mathbf{x}, \mathbf{z}) + \epsilon \tag{2.1}$$

where ϵ is noise. For now, we have not specified the form of the coupling $h(\mathbf{x}, \mathbf{z})$. Indeed, the focus of this paper is on how we can find better fit this coupling functions, both in a predictability and an interpretability sense, using Gaussian processes.

Throughout this paper, we will move the chosen variables between $\mathbf{x}$ and $\mathbf{z}$ depending on what models (variables) we wish to investigate. The original longitudinal data comes from M entities over T time steps. We assume that all the entities' individual time series are just different realizations of the same social system and concatenate the entity level data into input variables $\mathbf{x}, \mathbf{z}$, representing the state of the input variables for the corresponding observation in dy. Hence, the number of observations are $N = M \times (T-1)$. The inputs can be aggregated into a $D \times N$ design matrix X for the explanatory variables, a $D_* \times N$ design matrix Z, for the confounding variables, and to the target vector $\mathbf{dy}$. All input vectors are standardized to unit variance and zero mean, to enable variable comparison.

(i) Bayesian linear regression approach

Regression analysis aims to find a good approximation of the true functional relationship between variables [47]. Standard linear regression [30,48–51] is the simplest statistical model of the form we consider,

$$dy(\mathbf{x}) = \beta_0 + \mathbf{x}\beta + \epsilon \tag{2.2}$$

where β is a vector of regression coefficients and $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ is Gaussian white noise. The ordinary least squares (OLS) method [52–54] can be used to obtain the maximum-likelihood estimates for the parameters in this multivariate linear model.

Often when we test the robustness of a model, of the from given in Eq. 2.2, we wish to control for potentially confounding variables, $\mathbf{z}$ to check if the the results still holds after additional variables are introduced [5,55]. In this case, we fit various forms of,

$$dy(\mathbf{x}, \mathbf{z}) = \beta_0 + \mathbf{x}\beta + \mathbf{z}\gamma + \epsilon, \tag{2.3}$$

where γ is another vector of regression coefficients. We can determine which confounding variables are important and which can be ignored by assessing the degree to which the values of β and the overall explanatory power of the model are affected by the addition of the variable $\mathbf{z}$.

One way of allowing models to explain more complicated relations is to perform regression on a projection onto a higher dimensional feature space, e.g., polynomials of $\mathbf{x}$, using a set of basis functions $\mathbf{b}$. For example, a one-dimensional basis function, for variable x_1, of order three can be set to $\mathbf{b} = [1, x_1, x_1^2]$. Polynomial regression, for example, [29,51,56] allows for nonlinear relations in the explanatory variables in the form,

$$dy(\mathbf{x}) = \beta_0 + f(\mathbf{x}) + \epsilon. \tag{2.4}$$

In this case, β is implicitly defined as the parameters of a linear function $f(\mathbf{x}) = \mathbf{b}(\mathbf{x})^\top \beta$. We further define $B = \begin{bmatrix} \mathbf{b}(\mathbf{x}_1) \\ \vdots \\ \mathbf{b}(\mathbf{x}_n) \end{bmatrix}$ for the evaluation of these basis functions over all data points, where the indices of $\mathbf{x}_i$ refer to the data point identity rather a specific choice of explanatory variable. Polynomial functions are also sometimes used as statistical control for confounding variables in order to see if relations are robust under influence of non-linear relations (see [5]). In this paper, we choose the polynomial form of the basis functions to allow for complex models, but not too complex to retain interpretability.

Throughout this paper we adopt a Bayesian approach to the regressions performed. By adding a prior distribution on β we attain linear models in a Bayesian setting. Specifically, we define $\mathbf{b}(\mathbf{x})$ to be a set of predefined basis functions, consisting of linear and non-linear polynomials of $\mathbf{x}$, and we assume a Gaussian prior distribution on the regression coefficients with mean zero and a covariance matrix V: $\beta \sim \mathcal{N}(0, V)$. In our implementation we further assume a diagonal covariance matrix such that $V = c\mathbf{I}$, where $\mathbf{I}$ is the identity matrix and the constant c is chosen to be the same as the number of observations (1160). This corresponds to ridge regression with an uninformative prior distribution with wide coverage. In addition, it is a conjugate prior that permits analytical evaluation of the posterior distribution. We compare the set of models $\mathcal{M}$, consisting of all possible combinations of the terms in the basis functions using the log marginal likelihood (logML) as measure of model fit [30,34,39,41,57].

(ii) Gaussian process approach

Gaussian processes are a generic method for supervised learning in regression and classification [58]. A GP is defined by [58] as an infinite collection of random variables, any finite number of which have a joint Gaussian distribution. In our setting we consider the random variables that represent the values of the function $g(\mathbf{x})$ we wish to learn about at the locations $\mathbf{x}$. A GP $g(\mathbf{x})$ is fully specified by its mean function $\mu(\mathbf{x}) = \mathbb{E}[g(\mathbf{x})]$ and covariance function $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(g(\mathbf{x}) - \mu(\mathbf{x}))(g(\mathbf{x}') - \mu(\mathbf{x}'))]$, where $\mathbf{x}$, and $\mathbf{x}'$ are any two possible observations [30]. The mean function and covariance function contain all information about the assumptions we have for the function $g(\mathbf{x})$. The mean function is the expected values of the random variable $g(\mathbf{x})$ and the covariance function defines how similar the values of function $g(\mathbf{x})$ are at data points $\mathbf{x}$ and $\mathbf{x}'$. In supervised learning it is assumed that input data $\mathbf{x}$ which are close to each other are likely to have similar outputs $dy(\mathbf{x})$, hence data points close or similar to some test point x_* should be informative about the prediction at that test point and the covariance function specifies what we mean by similarity [58]. The chosen mean and covariance function do not depend on the actual data at this stage, but specify the properties we assume for the functions and are used as a prior for Bayesian inference, and their parameters can be learnt by the data. The model setup we consider has the form,

$$dy(\mathbf{x}) = \beta_0 + g(\mathbf{x}) + \epsilon \tag{2.5}$$

where,

$$g(\mathbf{x}) \sim \quad \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{2.6}$$

It is particularly worth noting that Bayesian linear regression can be performed through GPs by setting a specific mean function and dot product covariance function $k_{\text{dot}}(\mathbf{x}, \mathbf{x}') = \mathbf{b}(\mathbf{x})^\top V \mathbf{b}(\mathbf{x}')$ evaluated at points $\mathbf{x}$ and $\mathbf{x}'$ [30]. In what follows we, in some cases, use this particular choice, but in other cases, we will choose a more flexible covariance function, the squared exponential (SE) to provide a more pliant model fit (see equation 2.11). Rather than specifying a particular (e.g. linear) functional form for $g(x)$, this instead restricts this function only to be a smoothly varying and differentiable function of x. Gaussian processes with squared exponential covariance functions are non-parametric since we do not assume any parametric form of the function $g(x)$. We will also consider models of the form:

$$dy(\mathbf{x}, \mathbf{z}) = \beta_0 + g_x(\mathbf{x}) + g_z(\mathbf{z}) + \epsilon \tag{2.7}$$

where,

$$\begin{aligned} g_x(\mathbf{x}) \sim & \quad \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \\ g_z(\mathbf{z}) \sim & \quad \mathcal{GP}(\mu(\mathbf{z}), k(\mathbf{z}, \mathbf{z}')). \end{aligned} \tag{2.8}$$

Here we have split up the variables $\mathbf{x}$ and $\mathbf{z}$ into two different functions g_x and g_z. Observe that this model do not allow interactions between the variables $\mathbf{x}$ and $\mathbf{z}$. This model is used as a benchmark for the semi-parametric model explained below.

(iii) Semi-parametric approach

In our proposed semi-parametric approach we combine two GPs to model the change term dy. One GP, $f(\mathbf{x})$, is intended as an interpretable relation in the explanatory variable $\mathbf{x}$, while a non-parametric statistical control GP, $g(\mathbf{z})$ is used for confounding variables $\mathbf{z}$. The model has the form,

$$dy(\mathbf{x},\mathbf{z}) = \beta_0 + f(\mathbf{x}) + g(\mathbf{z}) + \epsilon, \tag{2.9}$$

where

$$\begin{aligned} f(\mathbf{x}) \sim &\quad \mathcal{GP}(0, k_{\text{dot}}(\mathbf{x},\mathbf{x}')) \\ g(\mathbf{z}) \sim &\quad \mathcal{GP}(0, k_{\text{SE}}(\mathbf{z},\mathbf{z}')) \end{aligned} \tag{2.10}$$

and the covariance functions are,

$$\begin{aligned} k_{\text{dot}}(\mathbf{x},\mathbf{x}') = &\quad \mathbf{b}(\mathbf{x})^\top V \mathbf{b}(\mathbf{x}') \\ k_{\text{SE}}(\mathbf{z},\mathbf{z}') = &\quad \sigma_f^2 \exp\left(-\sum_{i=1}^{D_*} \frac{(z_i - z_i')^2}{2l_i^2}\right) \end{aligned} \tag{2.11}$$

evaluated at points $\mathbf{x}$ and $\mathbf{x}'$ and $\mathbf{z}$ and $\mathbf{z}'$. Hyperparameters are a set of parameters for the covariance function, and for SE covariance function, k_{SE} they are the signal variances σ_f^2 and length scales l_i of variable z_i, indicating the relevance of this variable [30]. If the length scale is very long, for a specific variable, the covariance function will be almost completely independent of this variable, and vice versa [30]. This model is inspired by [59] where GPs are used to model residuals, with the conceptual difference to split explanatory variables and confounding variables into different covariance functions. This choice of structure assumes that the group of variables of $\mathbf{x}$ does not interact with the group of variables $\mathbf{z}$. Combining multiple GPs results in a new GP [30],

$$dy(\mathbf{x},\mathbf{z}) \sim \mathcal{GP}(\beta_0, k_{\text{dot}}(\mathbf{x},\mathbf{x}') + k_{\text{SE}}(\mathbf{z},\mathbf{z}') + \sigma_n^2 \mathbf{I}). \tag{2.12}$$

Notice the mean function of the semi-parametric GP is now β_0 and the covariance function is obtained by adding the individual covariance functions in Eq. 2.11 with $\sigma_n^2\mathbf{I}$, the covariance function for the noise ϵ. The choice of how we split data in to explanatory variable $(\mathbf{x})$ and confounding variables $(\mathbf{z})$ depends, like in Eq. 2.3, on how we wish to model dy. Note that setting $g(\mathbf{z}) = 0$ means that Eq. 2.9 and Eq. 2.4 are equivalent. Setting $f(\mathbf{x}) = 0$ and plugging data $[\mathbf{x},\mathbf{z}]$ into Eq. 2.5 is equivalent to model dy using a GP with a SE covariance function.

We test a number of polynomial models $f(\mathbf{x})$ to find the model in $\mathcal{M}$ that best approximates the underlying dependence of the change dy while statistically controlling using non-parametric $g(\mathbf{z})$. We do this by first fitting the model specified by Eq. 2.7, then removing the first non-parametric function $g_x(\mathbf{x})$ and then approximating it with a polynomial $f(\mathbf{x})$. Hence, we estimate the parameters β in $f(\mathbf{x})$ for all models in $\mathcal{M}$ by Bayesian linear regression, maximizing the overall marginal likelihood in the presence of $g(\mathbf{z})$ and picking the model with the highest model evidence. For the optimization of SE covariance functions, we use the automatic relevance determination (ARD) distance measure, and 10 restarts, where variables providing a good (bad) fit are assigned shorter (longer) length scales l_i in the optimization step [30]. A very long length scale of the ith variable (Eq. 2.11) means that the covariance function is almost independent of the ith input, and thereby its contribution to the inference is essentially removed [60]. We use the GPy toolbox [61] provided by the Sheffield Machine Learning group to fit our Gaussian processes.

(iv) Choice of polynomials

For both of the parametric model (Eq. 2.4) and the semi-parametric model (Eq. 2.9) we wish to approximate the relations in the explanatory variables $\mathbf{x}$ using the polynomial function $f(\mathbf{x})$. The functions $f(\mathbf{x})$ consist of linear and non-linear combinations of $\mathbf{x}$. In other words, we project the variables into a feature space and then perform Bayesian linear regression in this new feature space [30]. In order to find the polynomial expression that best approximate the relation between variables $\mathbf{x}$ and the level of change in RRP support (both with (Eq. 2.4) and without (Eq. 2.9) the presence of confounding variables $\mathbf{z}$), we test a predefined set of different model configurations $\mathcal{M}_j$, where $j = [1, ..., N_m]$ is the different model configurations out of N_m possible models. The model configurations consist of all combinations of the available polynomial terms we choose to test for. Hence we assume that $f(\mathbf{x})$ is 'built up' of k basis functions $\mathbf{b}_i$, where $i = [1, 2, ..., k]$ and k the number of terms in the polynomial model we test, and their corresponding coefficients β,

$$f(\mathbf{x}) = \sum_{i=1}^{k} \beta_i \mathbf{b}_i(\mathbf{x}). \tag{2.13}$$

Where the basis functions are a subset of the possible basis functions $\mathbf{b}$, i.e. $\mathbf{b}_i \subseteq \mathbf{b}$ and β_i are the corresponding slope coefficients. In this paper we consider models with all combinations of variables up to order 3. For a constant, one variable $\mathbf{x}_1$ (e.g. unemployment) polynomial models, terms up to order three: a linear x_1 term, a quadratic terms x_1^2 and a cubic term x_1^3, Hence, the basis function we consider in the one dimensional case is,

$$\mathbf{b}(x_1) \quad = [1, x_1, x_1^2, x_1^3] \tag{2.14}$$

and the models we test for consist of all combinations of these four terms. For two variable models (e.g. unemployment and education level), we allow for non-linear relations in the two variables, x_1 and x_2. The basis functions we consider consist of the following combinations variables,

$$\begin{aligned} \mathbf{b}(x_1, x_2) = \quad & [1, x_1, x_2, x_1^2, x_2^2, x_1^3, x_2^3, \\ & x_1x_2, x_1^2x_2, x_1x_2^2] \end{aligned} \tag{2.15}$$

For extra clarity: a two-variable example model of a polynomial approximation of the relation of variables $\mathbf{x} = [x_1, x_2]$ we wish to test in our investigation could be, $f(x_1, x_2) = \beta_1 x_1 + \beta_2 x_1^3 - \beta_3 x_1 x_2$. The user of this approach can choose to include any linear and non-linear combination to the set of tested basis functions $\mathbf{b}$. The included terms can be either more complicated linear and non-linear combinations, but also any special terms the user want to include from some existing theory, in order to test that theory.

The number of model configurations we investigate depends on the number of possible basis functions we allow into the model. For k allowed basis functions we get the number $N_m = 2^k - 1$ possible models (where we exclude the case where no input variables are considered). So if we consider the one explanatory variable x_1 setup, using set of basis functions (Eq. 2.14) we have $k = 4$ and thereby 15 different model configurations to investigate. The two variable model Eq. 2.15 then give 1023 models to test, and so on.

(v) Model evidence and slope parameters for the semi-parametric model

After we have performed GP regression for all models $\mathcal{M}_i$, we compare and rank them using marginal likelihood, $p(\mathbf{dy}|X, Z, \mathcal{M}_i)$, which is a measure of the probability of observing the data under the hypothesis that the model configuration $\mathcal{M}_i$ is true. The marginal likelihood is a likelihood function where the hyperparameters in the model have been marginalized. Using an SE covariance function together with a dot product covariance function (our semi-parametric) yields the following expression for the logarithm of the marginal likelihood, also referred to as the model evidence,

$$\begin{aligned} \log p(\mathbf{dy}|X, Z, \theta) = \quad & \frac{1}{2}\mathbf{dy}^\top (B^\top V B + K_{\text{SE}}(Z, Z) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{dy} \\ & -\frac{1}{2} \log |B^\top V B + K_{\text{SE}}(Z, Z) + \sigma_n^2 \mathbf{I}| \\ & -\frac{n}{2} \log(2\pi) \end{aligned} \tag{2.16}$$

where $\theta = [\beta, \sigma_f^2, l]$ is the set of all hyperparameters in the model $\mathcal{M}$. This marginal likelihood punishes over-complicated models and hinders overfitting by penalizing models with too much structure, in terms of the number of parameters & the numbers that they can take [62]. In Eq. 2.16, the quadratic term in $\mathbf{dy}$ gives a positive contribution depending on model fit, the negative log determinant punishes over-structured models and the term proportional to n is for normalization. For expressions of the model evidence for non-parametric and standard polynomial models are found in [30].

The final estimation of the parameters β in $f(\mathbf{x})$ maximizing for the model evidence given a model in $\mathcal{M}$ then becomes,

$$\hat{\beta} = (V^{-1} + B(K_{\text{SE}}(Z,Z) + \sigma_n^2 \mathbf{I})^{-1} B^\top)^{-1} B(K_{\text{SE}}(Z,Z) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{dy} \tag{2.17}$$

Note that the covariance structure from $g(\mathbf{z})$ affects the estimates of β.

3. Data

We use aggregated data on Sweden's 290 municipalities from the last five Swedish election years $(2002 - 2018)$. Based on the choices in previous studies, we include the following variables to check the social marginality theory: education, the proportion of the population with three or more years in post-secondary education; median income; and unemployment. We include foreign born density to test the ethnic conflict theory. We also include population density, mean age, crime, the total number of reported crimes per 100 000 inhabitants. These variables are commonly used to test the 'ethnic conflict' and social marginality theories (e.g. [5]). In addition we use the latitude and longitude of municipality capitals, i.e. we assume that the entire population is located in the 290 different regional capitals, to capture spatial dependence to account for spatial differences, i.e. a spatial regression model [63,64]; time (year); and support of the RRP in the national election. The time variable is included to capture the baseline overall increase of RRP support, like the constant (intercept) in a linear regression between time t and $t+1$. All variables except latitude, longitude and time are logarithm-transformed, using $\log(x_i + 1)$ to make the variables more approximately normally distributed and the '$+1$' to allow zero values in x_i. The Swedish National Council for Crime Prevention [65] provided data for crime, the Swedish Public Employment Service [66] provided data for the unemployment , and Statistics Sweden [67] provided data for all the other variables. All data and Python code can be found in supplementary materials.

4. Results

We compare all three regression approaches; standard regression, GP regression, and the proposed semi-parametric approach, by applying them in turn to investigate variable relations to changes in support for the Swedish RRP (ΔRRP). Model evidence (LogML) and R^2 values are measures of model fit and are presented in Table 1 for 8 different models. Models 1-3 are standard regression models, models 4-7 are non-parametric GP models, and model 8 is the proposed semi-parametric approach models. A good model fit is important because (1) it means that the hypothesized mechanism is more likely to have generated the data and (2) because models with better fit will produce more accurate predictions of the future of the dynamical system. For all models (1-8) the constant intercept parameter is $\beta_0 = 0.70$, i.e a constant increase in RRP support. Polynomial model expressions and slope parameters for models are presented in Table 2.

(vi) Standard regression models

We used Model 1, a one variable version of simple linear regression approach (Eq. 2.2), to investigate the relations between all of the individual variables and the change in support for the RRP. The variables found to have the highest model evidence were RRP support $[-183.77]$, median income $[-221.00]$, and education $[-235.17]$. The slope coefficients β_i for the simple linear regressions shows a negative correlation in RRP support ($\beta = -0.14$), a negative correlation for median income, and also a negative correlation in education ($\beta = -0.08$). The variable with the highest R^2 was RRP support explaining about 20% of the variability in the data. It should be noted here and in respect to later results that the explanatory power of current RRP support for predicting changes in RRP support may represent regression to the mean in a stochastic process, rather than a causal mechanism.

We then applied Model 2, the generalized linear regression approach (Eq. 2.2) to check the robustness of Model 1. We used all 11 variables and found a decrease in model evidence [-939.05] and R^2 value 45%. The reason why this model has a very low model evidence is because it includes all variables, even the ones with very low explanatory power. The slope coefficient of time is then considerable steeper and is now positive ($\beta = 0.24$) indicating a strong time dependence. Median income now have a small positive slope ($\beta = 0.01$). We find the slope coefficient for the level of RRP support is now a strong negative correlated variable ($\beta = -0.51$).

In Model 3 we use a polynomial regression approach (Eq. 2.4), using only single variables, to check if the nonlinear polynomial terms will increase the model fit and understanding. Just two variables, time and level of RRP support, have polynomial models with better model fit. The model with time has a considerably higher model evidence [-78.68] and R^2 value of 54%.

(vii) Gaussian process models

We test four non-parametric GP regression models (Eq. 2.5) with SE (ARD) covariance functions, where the only difference is in the number and identity of explanatory variables included in each model. In Model 4 we use a GP with a single z_i input variable to see how good the estimations can be using only one variable. The one variable that best described the change in RRP support was time with model evidence [171.19] and R^2 describing 54% of the variability, followed by RRP support and median income.

Another way of testing the relevance of a variable in making estimations is to omit it. In Model 5 we used only confounding variable $\mathbf{z}$ as input to the model to see how well a model fits without the exploratory variables z_i. The model with lowest model evidence (i.e. largest difference between Model 5 and Model 6), indicating the highest relevance in the variable z_i is the level of RRP support, having model evidence [405.80] and R^2 value (0.84) was closely followed by time, latitude and education level.

In Model 6 we use all variables as input $\mathbf{z}$ for the GP. This model could be considered as an 'upper bound' of the model evidence and R^2 since we utilize all available data in an entirely non-parametric model. However, as we will see in the next section, more specialized covariance function set-ups might give us a higher model evidence. The difference in model evidence between Model 5 and Model 6 is the room of improvement that explanatory variable x_i can fill,

Table 1: Comparison of model fit for models predicting change in RRP support in national elections. The values are the model log marginal likelihoods and (R^2 values). Amongst the semi-parametric models we indicate (*, bold text) the models in which the inclusion of the variable of interest improves upon an equivalent model excluding that variable (model 5).

	Standard regression models			Non-parametric models				Semi-parametric models
	Model 1:	Model 2:	Model 3:	Model 4:	Model 5:	Model 6:	Model 7:	Model 8:
Variable (i)	$\beta_i x_i$	$\beta_i x_i + \Sigma_{j\neq i}\gamma_j z_j$	$f(x_i)$	$g(x_i)$	$g(\mathbf{z}_{j\neq i})$	$g(\mathbf{z})$	$g(x_i) + g(\mathbf{z}_{j\neq i})$	$f(x_i) + g(\mathbf{z}_{j\neq i})$
Latitude	−262.60(0.03)	−939.05(0.45)	−262.60(0.03)	−255.37(0.04)	568.06(0.82)	651.69(0.88)	623.85(0.84)	**573.90 (0.82)***
Time	−407.19(0.08)	−939.05(0.45)	151.29(0.54)	171.19(0.54)	442.85(0.84)	651.69(0.88)	633.48(0.86)	**618.99 (0.86)***
RRP support	−183.77(0.20)	−939.05(0.45)	−78.62(0.32)	−29.89(0.43)	405.80(0.81)	651.69(0.88)	658.20(0.86)	**648.83 (0.86)***
Longitude	−273.13(0.01)	−939.05(0.45)	−273.13(0.01)	−264.95(0.02)	630.26(0.85)	651.69(0.88)	633.02(0.85)	625.16(0.85)
Education	−235.17(0.07)	−939.05(0.45)	−235.17(0.07)	−229.09(0.08)	593.54(0.87)	651.69(0.88)	650.77(0.88)	**642.75 (0.87)***
Population density	−244.93(0.05)	−939.05(0.45)	−244.93(0.06)	−242.15(0.06)	649.47(0.88)	651.69(0.88)	655.40(0.88)	**649.61 (0.88)***
Foreign born	−244.89(0.06)	−939.05(0.45)	−244.89(0.06)	−237.84(0.06)	649.01(0.88)	651.69(0.88)	650.58(0.88)	643.24(0.88)
Mean age	−268.74(0.02)	−939.05(0.45)	−268.74(0.02)	−262.19(0.02)	650.22(0.88)	651.69(0.88)	650.22(0.88)	644.73(0.88)
Unemployment	−264.09(0.02)	−939.05(0.45)	−264.09(0.02)	−259.04(0.03)	651.14(0.88)	651.69(0.88)	651.33(0.88)	645.79(0.88)
Median income	−221.00(0.10)	−939.05(0.45)	−221.00(0.10)	−183.47(0.16)	651.72(0.88)	651.69(0.88)	654.52(0.88)	645.73(0.88)
Crime	−268.09(0.02)	−939.05(0.45)	−268.09(0.02)	−264.97(0.02)	651.589(0.88)	651.69(0.88)	652.08(0.88)	646.14(0.88)

Table 2: Parameter values for models of change in RRP support.

	Standard regression models			Semi-parametric model
Variable (i)	Model 1: $\beta_i x_i$	Model 2: $\beta_i x_i + \Sigma_{j \neq i} z_j$	Model 3: $f(x_i)$	Model 8: $f(x_i) + g(\mathbf{z}_{j \neq i})$
Latitude	0.05	-0.16	$0.05x_i$	$-0.07x_i$
Time	-0.09	0.24	$0.19 + 0.40x_i - 0.19x_i^2 - 0.29x_i^3$	$0.75x_i - 0.20x_i^2 - 0.25x_i u^3$
RRP Support	-0.14	-0.51	$0.10 - 0.10x_i^2. - 0.06x_i^3$	$-0.42x_i + 0.06x_i^2$
Longitude	0.03	0.001	$0.03x_i$	-0.17
Education	-0.08	-0.01	$-0.08x_i$	$-0.09x_i$
Population density	-0.07	-0.08	$-0.07x_i$	$-0.04x_i$
Foreign born	-0.07	-0.01	$-0.07x_i$	-0.07
Mean age	0.04	0.01	$0.04x_i$	-0.08
Unemployment	0.05	0.01	$0.05x_i$	-0.10
Median Income	-0.09	0.01	$-0.09x_i$	-0.09
Crime	-0.04	0.01	$-0.04x_i$	-0.09

Table 3: Relevance of variables to predict changes in RRP support in national elections using squared exponential (ARD) covariance function. Low length scale indicates more relevance. All variables are normalized to unit variance to enable variable comparison.

			Cumulative	
Ranking	**Variable**	Length scale	LogML	R^2
1	Latitude	0.62	−255.37	0.04
2	Time	0.77	260.69	0.65
3	RRP support	1.30	419.00	0.76
4	Longitude	1.83	517.81	0.84
5	Education	4.04	639.26	0.87
6	Population density	10.48	644.57	0.88
7	Foreign born	12.48	649.79	0.88
8	Mean age	14.16	650.97	0.88
9	Unemployment	22.46	651.59	0.88
10	Median income	36.53	651.59	0.88
11	Crime	48.06	651.69	0.88

indicating that longitude, population density, crime, foreign-born, mean age, unemployment, and median income are variables with little potential to improve the models.

In Model 7 we use all variables, but we split them up into the sum of two different non-parametric functions, one for the explanatory variables x_i and the other for the confounding variables $\mathbf{z}$. This makes interactions between the split variables impossible, and the effects are instead additive. Model 7 are later going to be used to benchmark (in Model 8) how well a polynomial function $f(x_i)$ can approximate the $g(\mathbf{z})$ part of eq. 2.7.

(viii) Variable relevance

Including all variables into an ARD model (Model 6) and optimizing the hyperparameters, we implicitly obtain the relevance of the different variables by telling us how sensitive the response dy is to changes in the corresponding variable (shown in Table 3). A short length-scale indicates that a variable is more relevant, and long length-scale can be used to remove irrelevant variables [68]. However, this is not always true and the length-scale should be used as a guide to the variable's likely effect on the model evidence, not as a definitive measure of the size of the effect. The most relevant variables to model change in SD support are, in order, latitude ($l = 0.62$), time ($l = 0.77$), RRP support ($l = 1.30$), longitude (West/East) ($l = 1.83$) and education ($l = 4.04$).

In Table 3 we also present cumulative log marginal likelihood and R^2, where one variable after the others is included, in order of shortest length scales. The cumulative log marginal likelihoods and R^2 values indicates that almost all of the explanation power are obtained after the five variables (up to education) with the shortest length scales.

(ix) Semi-parametric model

Lastly, we look at a semi-parametric model where an explicit expression gives the relation in the explanatory variable x_i and the confounding variables $\mathbf{z}$ are modelled using a SE covariance function (Eq. 2.9). In Model 8 we allow the configuration to take the polynomial form with the highest model evidence. By comparing the semi-parametric Model 8 to Model 5, we can see how much the approximating polynomial expression using x_i improves the model evidence. Only the variables: RRP level, time, latitude, education, and population density improved upon the model evidence of Model 5. The variable where the model evidence increased the most was in the variable for the level of support for the RRP.

(x) Higher dimensional models

By using two or more variables in the polynomial function f in (Eq. 2.9) we can gain more understanding of the linear or non-linear relation between these explicit variables. Using the variables: support for RRP (S) and time (t) we get the resulting best polynomial model, $f(S,t) = -0.56S + 0.89t - 0.14t^2 - 0.25t^3$ with model evidence [618.38] and R^2 value (0.88). Notice the regression to the mean effect of RRP support, and that the model captures the decrease in growth of RRP support before the last election in 2018. (Fig. 1A). The best model using education (E) and population density (P) is $f(E,P) = -0.04E - 0.08P$ with model evidence [639.56] and R^2 (0.87) (Fig. 1B). Both the variables latitude and longitude have short length scales, but only latitude contribute positively to the model evidence in Table. 1. Investigating the impact of these relevant variables, according to length scales, we find that the best semi-parametric model is: $f(\text{latitude}, \text{longitude}) = -0.07 \cdot \text{latitude}$; only the latitude gives a contribution to the change in RRP support. This result is shown in Fig. 1C.

In Fig. 2 we can see the relationship between education level and RRP support using the two standard 1-variable models, Model 1 (Fig. 2A); our standard 11 variables Model 2 (Fig. 2B); our best polynomial model, Model 3 (Fig. 2C); and our semi-parametric approach using Model 8 (Fig. 2D). As we clearly can see, the results are more detailed for our semi-parametric model (Model 8). If we put all of the most significant variables– except latitude and longitude– time (t), RRP support (S), and education (E) into one polynomial model with all others as confounding $\mathbf{z}$ we get the best model,

$$\Delta\text{RRP} = -0.57S - 0.10E - 0.02E^2 + 0.92t - 0.14t^2 - 0.25t^3$$

with model evidence [610.19] and R^2 (89%), indicating that there are no clear interactions between the explanatory variables.

5. Discussion

In this paper, we have investigated the driving factors behind the rise of RRPs. The variables that we found to best explain the changes in support of the Swedish RRP were, in order of relevance: latitude, time, RRP support, longitude, education level and population density. Out of these variables, the only one we found to support the social marginality theory is education, unlike in [5], where GRP, reported crime and unemployment were also found to be significant. Based only on proportion of foreign-born residents, we found no support for the ethnic competition theory as the proportion of foreign born was found to have very little contribution to the model fit. However, we acknowledge that this is a relatively crude test of ethnic competition theory, and individual-level processes may exist that are not observed on the level of municipality-aggregated data. The conflict between our results and earlier studies suggests that an individual-level approach is needed to test this theory further.

Our new modelling approach has allowed us to obtain a model fit with R^2-values capturing over 88% of the variance in the data. This offers considerably better predictability when compared to the best model for change in Swedish RRP support in Rydgren & Ruth [5] where they could explain 9% of the variability in the data. Compared to [5], we used more variables (e.g. latitude, longitude, population density, domestic migration, median income, and time) and longer time series, (2002-2018) instead of (2006, 2010), but it is also the updated methodology that allow us to capture so much of the dynamics — we only can fit 45% of the variability in the data with Model 2, corresponding to the approach of [5] (see Table. 1).

The identified strong effect of time suggests an intrinsic acceleration in support for the RRP. However, time itself cannot causally influence RRP support. Instead, this is likely to be the result of one or multiple underlying causes for increased RRP support not picked up due to our choice of variables. For example, time could capture: increased media coverage, social media usage or a transition of the demand side (voter support) to meet a renewed supply of political ideas by the RRP [69]. This suggests that there is still substantial scope for identifying the underlying causal factors influencing RRP support in Sweden and beyond.

Our results have important implications for policy makers who want to understand the development of RRP support over the last two decades. The conclusions we can draw from our investigation are that (1) there are large regional differences, with the increase in RRP support being concentrated in southern Sweden to an extant that is not explained solely by differences in the other variables we have measured ; (2) RRP ideas have spread faster in rural regions with low population densities (controlling for the overall spatial pattern); and (3) that a high education level has made the population less susceptible to the recent overall increase in RRP support.

Our results pertain to changes in RRP support as opposed to stock values. We consistently find that current levels of RRP support are negatively correlated with future increases. This may in part be a regression to the mean amid stochasticity, but could also represent a process whereby regions with high existing levels of RRP support have limited potential for further growth in that support. Our goal has been to explain the large change in RRP support that has been a feature of Swedish and European politics in this century, but we should bear in mind that in some regions support for RRPs was already high before this period, which may influence the dynamical models we have inferred. For example, if support for RRPs was already very high in regions with high immigrant populations in 2002, we may observe no further increase as a result of this factor during our study period.

In contemporary social science, classical statistical techniques have been favoured over more complex models in part because of their ease of interpretation. Our methodology substantially improves the trade-off between model accuracy and interpretability, and thus allows us to discover important effects within a complex, non-linear and multivariate system. Here we have focussed on modelling the change of one variable, the rise of an RRP. It is however straightforward to use our method to simultaneously fit coupled equations, as in [34,39], with each fit made independently [70]. Our method thus provides a powerful way for identifying and explaining the complex, coupled relationships found within social systems. The methods we present can also be applied in a wide variety of fields to explicitly study variables of interest contributing to the dynamics, while controlling for additional variables in a flexible way.

Data Accessibility. The Swedish National Council for Crime Prevention [65] provided data for crime, the Swedish Public Employment Service [66] provided data for the unemployment , and Statistics Sweden [67] provided data for all the other variables. All data and Python code are also attached as supplementary materials.

Authors' Contributions. BRHB performed the data analysis. BRHB wrote the softvare, BRHB and RPM conceived the idea of the Gaussian process approach. All authors designed the study. BRHB drafted the manuscript and all authors then helped with the writing. All authors read and approved the manuscript.

Competing Interests. The authors declare that they have no competing interests.

Acknowledgements. The authors would like to acknowledge Viktoria Spaiser, Ranjula Bali Swain, Alex Szorkovszky and Ian Vernon for all the valuable feedback on the manuscript.

References

1. Jens Rydgren.
The Oxford Handbook of the Radical Right.
Oxford University Press, 2018.
2. Jens Rydgren.
The sociology of the radical right.
Annu. Rev. Sociol., 33:241–262, 2007.
3. Ronald Inglehart and Pippa Norris.
Trump, brexit, and the rise of populism: Economic have-nots and cultural backlash.
2016.
4. Marcel Lubbers, Mérove Gijsberts, and Peer Scheepers.
Extreme right-wing voting in western europe.
European Journal of Political Research, 41(3):345–378, 2002.
5. Jens Rydgren and Patrick Ruth.
Voting for the radical right in swedish municipalities: Social marginality and ethnic competition?
Scandinavian Political Studies, 34(3):202–225, 2011.
6. Kai Arzheimer and Elisabeth Carter.
Political opportunity structures and right-wing extremist party success.
European Journal of Political Research, 45(3):419–443, 2006.
7. Pia Knigge.
The ecological correlates of right–wing extremism in western europe.
European Journal of Political Research, 34(2):249–279, 1998.
8. Jocelyn AJ Evans.
The dynamics of social change in radical right-wing populist party support.
Comparative European Politics, 3(1):76–101, 2005.
9. Ruud Koopmans.
Contested citizenship: Immigration and cultural diversity in Europe, volume 25.
U of Minnesota Press, 2005.
10. Hanspeter Kriesi.
Movements of the left, movements of the right: Putting the mobilization of two new types of social movements into political context.
Continuity and change in contemporary capitalism, pages 398–423, 1999.
11. Hubert M Blalock.
Per cent non-white and discrimination in the south.
American Sociological Review, 22(6):677–682, 1957.
12. Hubert M Blalock.
Toward a theory of minority-group relations.
1967.
13. Susan Olzak.
The dynamics of ethnic competition and conflict.
Stanford University Press, 1994.
14. Thomas F Pettigrew.
Demographic correlates of border-state desegregation.
American Sociological Review, 22(6):683–689, 1957.
15. Silke L Schneider.
Anti-immigrant attitudes in europe: Outgroup size and perceived ethnic threat.
European Sociological Review, 24(1):53–67, 2008.
16. Matt Golder.
Far right parties in europe.
Annual Review of Political Science, 19:477–497, 2016.
17. Elisabeth Ivarsflaten.
The vulnerable populist right parties: No economic realignment fuelling their electoral success.

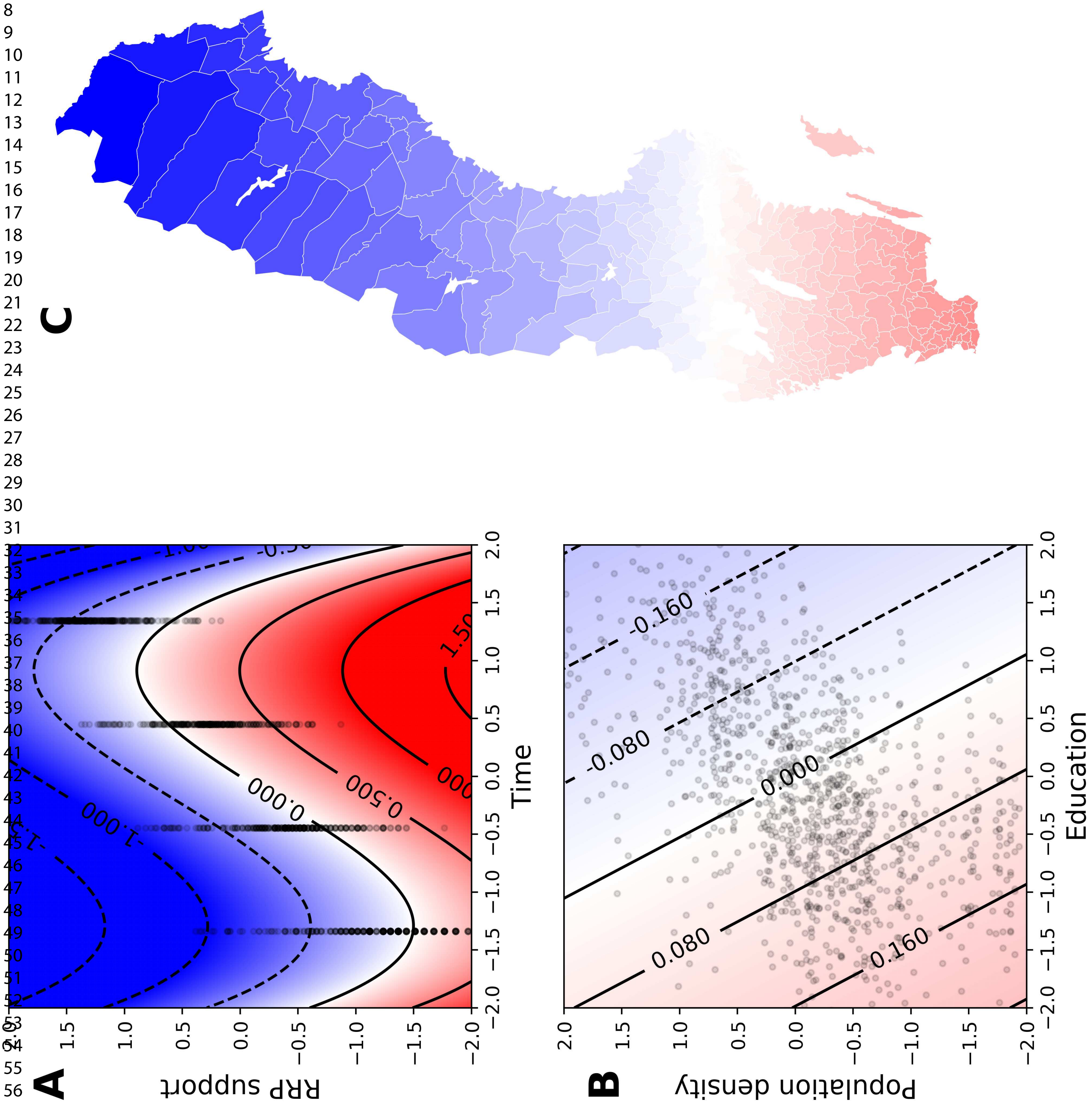
European Journal of Political Research, 44(3):465–492, 2005.
18. Pippa Norris.
Radical right: Voters and parties in the electoral market.
Cambridge University Press, 2005.
19. Matt Golder.
Explaining variation in the success of extreme right parties in western europe.
Comparative Political Studies, 36(4):432–466, 2003.
20. Duane Swank and Hans-Georg Betz.
Globalization, the welfare state and right-wing populism in western europe.
Socio-Economic Review, 1(2):215–245, 2003.
21. Wouter Van der Brug, Meindert Fennema, and Jean Tillie.
Why some anti-immigrant parties fail and others succeed: A two-step model of aggregate electoral support.
Comparative Political Studies, 38(5):537–573, 2005.
22. Jens Rydgren.
Immigration sceptics, xenophobes or racists? radical right-wing voting in six west european countries.
European Journal of Political Research, 47(6):737–765, 2008.
23. Jeremy Miles and Mark Shevlin.
Applying regression and correlation: A guide for students and researchers.
Sage, 2001.
24. Rachel A Gordon.
Regression analysis for the social sciences.
Routledge, 2015.
25. Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken.
Applied multiple regression/correlation analysis for the behavioral sciences.
Routledge, 2013.
26. David Kaplan.
Bayesian statistics for the social sciences.
Guilford Publications, 2014.
27. James Jaccard and Robert Turrisi.
Interaction effects in multiple regression.
Number 72. Sage, 2003.
28. Leona S Aiken, Stephen G West, and Raymond R Reno.
Multiple regression: Testing and interpreting interactions.
Sage, 1991.
29. James P Stevens.
Applied multivariate statistics for the social sciences.
Routledge, 2012.
30. Carl Edward Rasmussen and Christopher KI Williams.
Gaussian processes for machine learning.
MIT press Cambridge, 2006.
31. Kevin P Murphy.
Machine learning: a probabilistic perspective.
MIT press, 2012.
32. Richard Mann, Robin Freeman, Michael Osborne, Roman Garnett, Chris Armstrong, Jessica Meade, Dora Biro, Tim Guilford, and Stephen Roberts.
Objectively identifying landmark use and predicting flight trajectories of the homing pigeon using gaussian processes.
Journal of The Royal Society Interface, 8(55):210–219, 2011.
33. Richard P Mann, Viktoria Spaiser, Lina Hedman, and David JT Sumpter.
Choice modelling with gaussian processes in the social sciences: A case study of neighbourhood choice in stockholm.
PloS one, 13(11):e0206687, 2018.
34. Shyam Ranganathan, Viktoria Spaiser, Richard P Mann, and David JT Sumpter.
Bayesian dynamical systems modelling in the social sciences.
PloS one, 9(1):e86468, 2014.
35. Viktoria Spaiser, Peter Hedström, Shyam Ranganathan, Kim Jansson, Monica K Nordvik, and David JT Sumpter.

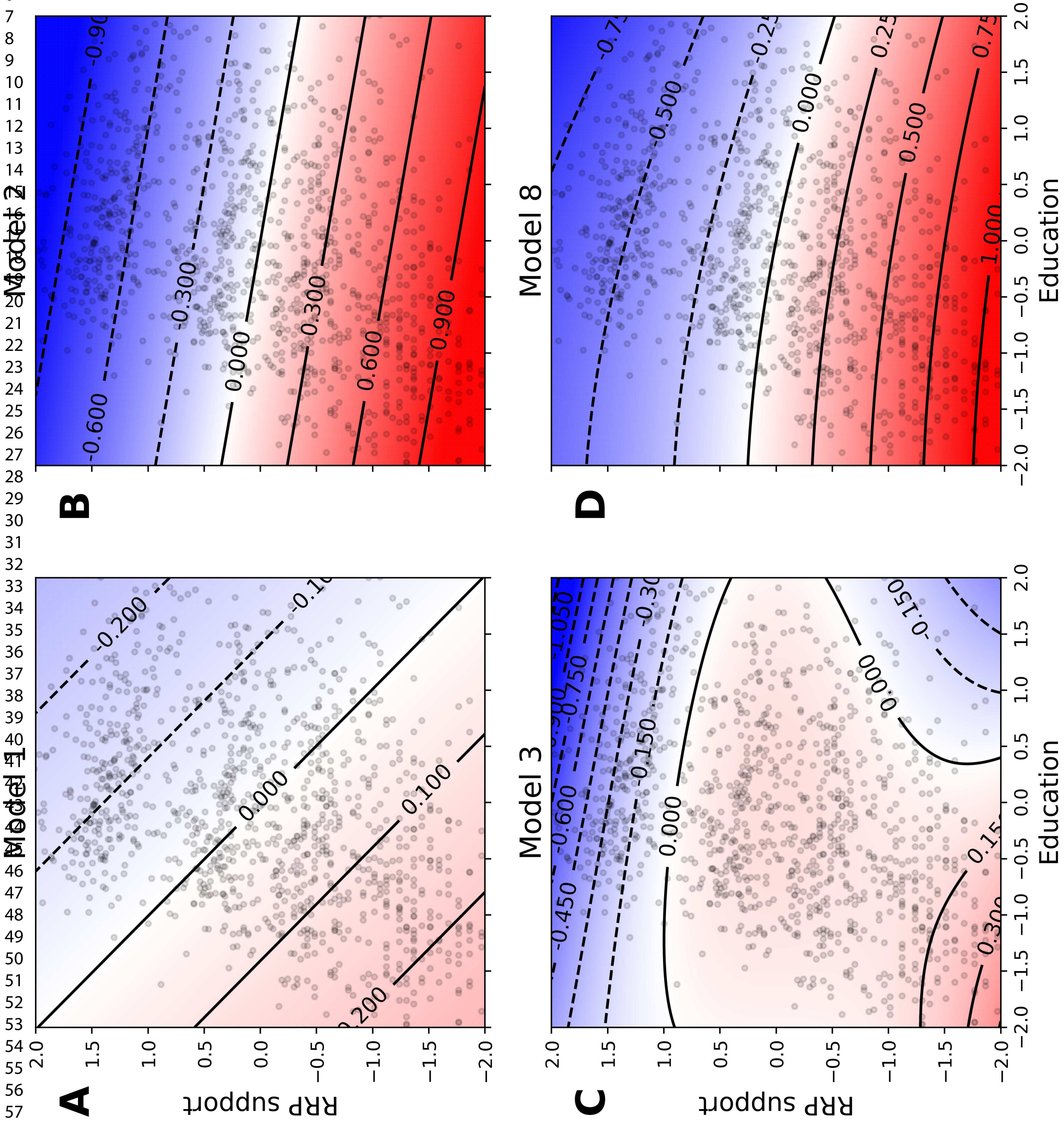
Identifying complex dynamics in social systems: A new methodological approach applied to study school segregation.
Sociological Methods & Research, page 0049124116626174, 2016.
36. S Ranganathan, R Bali Swain, and David JT Sumpter.
The demographic transition and economic growth: A dynamical systems model.
Palgrave Communications, 1, 2015.
37. BRH Blomqvist, RP Mann, and DJT Sumpter.
Using bayesian dynamical systems, model averaging and neural networks to determine interactions between socio-economic indicators.
PLoS ONE, 2018.
38. Tomislav Stankovski, Tiago Pereira, Peter VE McClintock, and Aneta Stefanovska.
Coupling functions: Universal insights into dynamical interaction mechanisms.
arXiv preprint arXiv:1706.01810, 2017.
39. Viktoria Spaiser, Shyam Ranganathan, Richard P Mann, and David JT Sumpter.
The dynamics of democracy, development and cultural values.
PloS one, 9(6):e97856, 2014.
40. Shyam Ranganathan, Stamatios C Nicolis, Viktoria Spaiser, and David JT Sumpter.
Understanding democracy and development traps using a data-driven approach.
Big data, 3(1):22–33, 2015.
41. Viktoria Spaiser, Shyam Ranganathan, Ranjula Bali Swain, and David JT Sumpter.
The sustainable development oxymoron: quantifying and modelling the incompatibility of sustainable development goals.
International Journal of Sustainable Development & World Ecology, 24(6):457–470, 2017.
42. Isao T Tokuda, Swati Jain, István Z Kiss, and John L Hudson.
Inferring phase equations from multivariate time series.
Physical review letters, 99(6):064101, 2007.
43. István Z Kiss, Craig G Rusin, Hiroshi Kori, and John L Hudson.
Engineering complex dynamical structures: sequential patterns and desynchronization.
Science, 316(5833):1886–1889, 2007.
44. Jun Miyazaki and Shuichi Kinoshita.
Determination of a coupling function in multicoupled oscillators.
Physical review letters, 96(19):194101, 2006.
45. Tomislav Stankovski, Valentina Ticcinelli, Peter VE McClintock, and Aneta Stefanovska.
Neural cross-frequency coupling functions.
Frontiers in systems neuroscience, 11:33, 2017.
46. Tomislav Stankovski, Peter VE McClintock, and Aneta Stefanovska.
Coupling functions enable secure communications.
Physical Review X, 4(1):011026, 2014.
47. David GT Denison.
Bayesian methods for nonlinear classification and regression, volume 386.
John Wiley & Sons, 2002.
48. John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman.
Applied linear statistical models, volume 4.
Irwin Chicago, 1996.
49. George AF Seber and Alan J Lee.
Linear regression analysis, volume 936.
John Wiley & Sons, 2012.
50. Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining.
Introduction to linear regression analysis, volume 821.
John Wiley & Sons, 2012.
51. John O Rawlings, Sastry G Pantula, and David A Dickey.
Applied regression analysis: a research tool.
Springer Science & Business Media, 2001.
52. Alexander Von Eye and Christof Schuster.
Regression analysis for social sciences.
Academic Press, 1998.
53. Simo Puntanen and George PH Styan.
The equality of the ordinary least squares estimator and the best linear unbiased estimator.
The American Statistician, 43(3):153–161, 1989.

54. Arthur P Dempster, Martin Schatzoff, and Nanny Wermuth. A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, 72(357):77–91, 1977.
55. Hubert M Blalock. *Causal models in the social sciences*. Transaction Publishers, 1985.
56. Norman R Draper and Harry Smith. *Applied regression analysis*. John Wiley & Sons, 2014.
57. Björn R H Blomqvist, Richard P Mann, and David J T Sumpter. Using bayesian dynamical systems, model averaging and neural networks to determine interactions between socio-economic indicators. submitted, 2017.
58. Carl Edward Rasmussen and Christopher KI Williams. Gaussian processes in machine learning. *Lecture notes in computer science*, 3176:63–71, 2004.
59. BJN Blight and L Ott. A bayesian approach to model inadequacy for polynomial regression. *Biometrika*, 62(1):79–88, 1975.
60. Sunho Park and Seungjin Choi. Gaussian process regression for voice activity detection and speech enhancement. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 2879–2882. IEEE, 2008.
61. GPy. GPy: A gaussian process framework in python. http://github.com/SheffieldML/GPy, since 2012.
62. Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.
63. Jean-Paul Chiles and Pierre Delfiner. *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley & Sons, 2009.
64. Luc Anselin. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural economics*, 27(3):247–267, 2002.
65. Swedish national council for crime prevention, statistical database. http://statistik.bra.se/solwebb/action/index, 2019. [Online; accessed 8-March-2019].
66. Swedish public employment service. https://arbetsformedlingen.se/, 2019. [Online; accessed 27 February- 9 March, 2019].
67. Statistics sweden, statistical database. http://www.statistikdatabasen.scb.se/pxweb/en/ssd/, 2019. [Online; accessed 27 February- 9 March, 2019].
68. Carl Edward Rasmussen. *Evaluation of Gaussian processes and other methods for non-linear regression*. PhD, Thesis, University of Toronto, 1996.
69. Cas Mudde. *Populist radical right parties in Europe*, volume 22. Cambridge University Press Cambridge, 2007.
70. Björn Rune Helmer Blomqvist. *Gaussian process models of social change*. PhD thesis, Department of Mathematics, Uppsala University, 2018.

Figure 1: Impact of time, education and latitude & longitude on RRP support. (A) Impact of time and RRP support on ΔRRP support. (B) Impact of education and population density on ΔRRP support. (C) Impact of latitude & longitude on ΔRRP support. The points in A & B are data observations. Red regions correspond to positive contributions on ΔRRP support, and blue regions to negative contributions, with numerical values indicated by contour lines. Observe that we show logarithmic values on the black axes (bottom and left) of A & B

Figure 2: Four different models of change in RRP support dependent on education level and current RRP support. (A) Model 1, (B) Model 2, (C) Model 3, and (D) Model 8. Panel A shows a visualisation of the two, one variable, linear models (Model 1). Panel B shows the linear model with all variables included (Model 2), Panel C shows the best two variable polynomial model ($f(S, E)$). Panel D shows the semi-parametric model (Model 8) with the highest model evidence $[633.26]$ and R^2 of 87%. The points in each panel are data observations. Red regions correspond to positive contributions on RRP support, and blue regions to negative contributions, with numerical values indicated by contour lines. Observe that we show logarithmic values on the black axes (bottom and left)


C
A
RRP support
Time
B
Population density
Education


A
Model 1
B
Model 2
C
Model 3
D
Model 8
RRP support
Education