

Bayesian Dynamical Systems Modelling in the Social Sciences

Shyam Ranganathan^{1*}, Viktoria Spaiser², Richard P. Mann^{1,2}, David J. T. Sumpter^{1,2}

1 Department of Mathematics, Uppsala University, Uppsala, Sweden, **2** Institute for Futures Studies, Stockholm, Sweden

Abstract

Data arising from social systems is often highly complex, involving non-linear relationships between the macro-level variables that characterize these systems. We present a method for analyzing this type of longitudinal or panel data using differential equations. We identify the best non-linear functions that capture interactions between variables, employing Bayes factor to decide how many interaction terms should be included in the model. This method punishes overly complicated models and identifies models with the most explanatory power. We illustrate our approach on the classic example of relating democracy and economic growth, identifying non-linear relationships between these two variables. We show how multiple variables and variable lags can be accounted for and provide a toolbox in R to implement our approach.

Citation: Ranganathan S, Spaiser V, Mann RP, Sumpter DJT (2014) Bayesian Dynamical Systems Modelling in the Social Sciences. PLoS ONE 9(1): e86468. doi:10.1371/journal.pone.0086468

Editor: Fabio Rapallo, University of East Piedmont, Italy

Received: October 23, 2013; **Accepted:** December 11, 2013; **Published:** January 20, 2014

Copyright: © 2014 Ranganathan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was funded by European Research Council grant IDCAB 220/104702003 and the Riksbanken Jubileumsfond (<http://www.rj.se/english>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shyam.ranganathan@math.uu.se

Introduction

Social science usually aims to explain macro-level phenomena, such as stratification, segregation, democratisation, economic development and changes in values. From the vast number of examples studied across sociology, politics and economics, a few examples include: Does inequality decrease or increase ethnic segregation [1]?; Do economic growth and changing cultural values of a society promote democracy [2,3]? What are the causes of economic growth [4]? How does the structure of a social network affect opinion dynamics [5]? In these wide ranging questions, the macro-level variables can concern a variety of scales, from schools and neighbourhoods, up to companies and countries. The questions about them are similar: from observed macro-level patterns, can we work out the relationships that characterize these patterns [6].

While it is widely recognized that understanding at the micro-level is the key to causal mechanisms in sociology [7–9], it is possible to gain some understanding of social systems through macro-level patterns alone. One example of macro-level inference is fitting sigmoidal and saturating curves to describe diffusion of innovations [10,11]. In this case, the type of growth curve is hypothesized to differ depending on whether innovation is driven by endogenous or exogenous factors at the micro-level. Specifically, by fitting the growth curve

$$\frac{dx}{dt} = (a + bx)(1 - x) \quad (1)$$

to the proportion of individuals adopting a particular activity x over time t , we can infer the relative importance of exogenous and endogenous (social) factors from the relative weights of the parameters a and b , respectively. While having certain known

limitations [12,13], this approach has been usefully applied in the context of, for example, medical innovations and radio airplay.

Only a very small subset of social systems are characterized by sigmoidal growth curves. However, non-linear interactions between variables in social systems are common, and using differential equations to give an initial insight into macro-level relationships has a great deal of potential [14–16]. Econometrics provides a starting point for such an analysis. For example, in growth econometrics cross-country data is used to find which factors promote economic growth [17–19]. However, growth econometric analyses usually focus on the rate of change of one variable as a function of many potential factors, rather than dynamic interactions between variables. It is precisely these dynamic feedbacks which are of most interest in the social sciences and where reliable statistical approaches are required [4]. In recent years, detailed data describing long term changes in social systems has become widely available. For example, a variety of indicators now measure changes in the economics [20,21], social development [21], political systems [22–24] and cultural values [25] of different countries and local regions. Identifying relationships between these macro-level indicators poses new challenges, but also opens up new opportunities. These challenges are not unique to between country comparisons, but arise in everything from social movements, workplaces, and neighbourhoods, down to modeling individual panel data on emotion dynamics [26,27].

The approach we take here is inspired by machine-learning and algorithmic modeling [28,29], in that we use the available data to inform model selection from a pool of feasible models rather than testing specific models against data. We use Bayesian model selection as a measure of the reliability and robustness of the differential equation models, basing our selection on the Bayes factor of each model. Our aim is to identify potential relationships in macro-level data, that can be further investigated in terms of

mechanisms at the micro-level. Identifying this link is achieved by fitting polynomial differential equations, because they describe how change in an indicator variable occurs as an explicit function of the state of other indicators. This stands as a contrast to other function-fitting frameworks used in machine-learning, such as artificial neural networks or Gaussian processes, where the relationship between the variables is more opaque. In the preface to their book, Rasmussen and Williams [30] argue that statistics is concerned with understanding data and relationships in terms of models, whereas machine-learning is more focused on making accurate predictions. Mackay [31] (chapter 45) also highlights the limitations of advanced machine-learning methods for identifying ‘features’ in the data. While our methodology is inspired by such algorithmic model-search methods of machine-learning, our goals are aligned with those of statistics and, of course, sociology, to identify explicit and clearly understandable relationships in the data.

We illustrate our approach on the classic problem of determining an interaction between GDP per capita and democracy (Fig. 1), which are known to correlate while the causal link and specific character of their relations remains an issue of debate [32–36]. The analysis has been performed on a set of 74 countries from 1981 to 2006. We implemented the method of Bayesian Dynamical Systems Modelling in an R package that is now openly available (Bayesian Dynamical System Model, *bdynsys*, to be found on CRAN (<http://cran.r-project.org>) and we used this package to do the analysis on the relation between GDP per capita and democracy. Our intention in this paper is not to speculate over why certain relationships exist, but rather to outline the methodology for identifying existing relationships.

Methods

Our basic approach to understanding interactions between indicator variables is to model changes in one variable between times t and $t+1$ as a function of all included model variables at time t . Suppose that we are studying a social system with N indicator variables $X_i, i=1, \dots, N$. Let us assume that we have longitudinal or panel data for the N variables for M entities (such as countries, organisations, individuals etc.) over a length of time

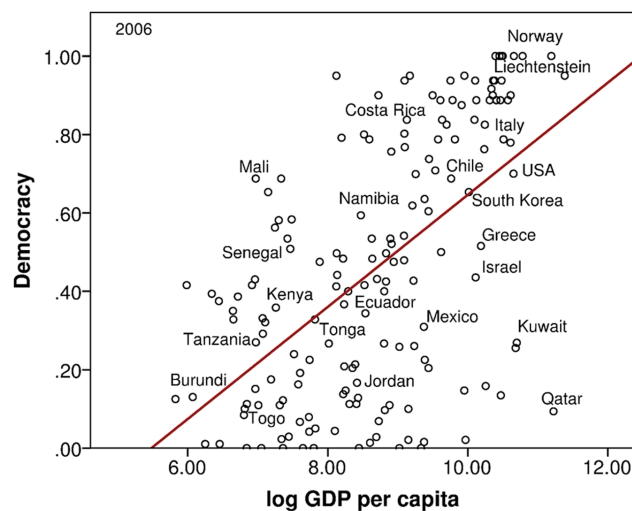


Figure 1. Country-based correlation scatter plot for log GDP per capita and Democracy for the year 2006. The Pearson correlation coefficient between the two variables is 0.571 ($p < 0.01$). doi:10.1371/journal.pone.0086468.g001

T . We denote the data as $x_i(j,t)$ where $j=1, \dots, M$ and $t=1, \dots, T$. The changes in the variables over a time period are denoted as $dx_i(j,t) = x_i(j,t+1) - x_i(j,t)$. We write x_i to denote the $M \times (T-1)$ matrix (the values at time T are not used in the modeling so we drop them from this matrix) whose elements are given by $x_i(j,t)$. Similarly, dx_i is the $M \times (T-1)$ matrix with elements $dx_i(j,t)$. We now outline our approach for two variables and then move on to the general case.

Systems with Two Variables

Our aim is to take these time series and fit an ordinary differential equations model to them. A system of differential equations with two variables can be represented as

$$\frac{dX_1}{dt} = f_1(X_1, X_2) \tag{2}$$

$$\frac{dX_2}{dt} = f_2(X_1, X_2) \tag{3}$$

for some appropriate functions f_1 and f_2 . Note that since we have discrete data, we need to use difference equations instead of differential equations when performing the fitting. However, we express our model in terms of continuous time both for mathematical convenience and because the underlying social process is often continuous.

We can think of the time series of the indicator variables for each entity j as corresponding to different trajectories obtained from the same system. We assume that any individual entity on reaching a certain state, represented by a unique set $\{X_i\}$, will experience the same effect, with distortions due to random, uncorrelated noise. The advantage of this approach is that we can use data arising from different initial conditions in different entities to build up an estimate of the functions $f_i(\cdot)$. The potential disadvantage of this approach arises in cases where there are systematic differences between entities. This limitation should be borne in mind when we consider the results of the fitting.

In order to model as many non-linearities as possible, we take $f_i(\cdot)$ to be polynomial of sufficiently high degree. For computational purposes, we assume that the functions are polynomial in the indicator variables with each term being of degree $-1, 0, 1$ in the variables or a product of such terms. We also allow for terms that are quadratic and cubic in single variables. This keeps the number of evaluated models sufficiently small, while still allowing us to capture complex interactions. Higher order terms are important because they typically result in multi-stable states, which are characteristic of realistic social systems [16].

In our standard implementation of a two variable model, we study models of the form:

$$\begin{aligned} f_1(X_1, X_2) = & a_0 + \frac{a_1}{X_1} + \frac{a_2}{X_2} + a_3 X_1 + a_4 X_2 \\ & + \frac{a_5}{X_1 X_2} + \frac{a_6 X_2}{X_1} + \frac{a_7 X_1}{X_2} + a_8 X_1 X_2 \\ & + a_9 X_1^2 + a_{10} X_2^2 + \frac{a_{11}}{X_1^2} + \frac{a_{12}}{X_2^2} + a_{13} X_1^3 \\ & + a_{14} X_2^3 + \frac{a_{15}}{X_1^3} + \frac{a_{16}}{X_2^3} \end{aligned}$$

A model is defined by a subset of coefficients $\{a_0, \dots, a_{16}\}$ obtained by multiple regression of the change variable with the corresponding polynomial terms. The number of terms m in the model is given by the number of non-zero coefficients. Thus there are 17 models with one term and $\binom{17}{m}$, models with m terms. The objective of the model selection algorithm is to choose among these models the one that best fits the data. We do this in two stages.

In the first stage, we aim to rapidly narrow our search by finding the maximum-likelihood model for each possible number of terms m . We fit the changes in the indicator variables using multiple linear regression over all $2^{17} = 131,072$ possible functions $f_1(X_1, X_2)$. For all models we compute the log-likelihood value

$$\log P(dx_1|x_1, x_2, \phi) \tag{4}$$

where ϕ is the set of parameters that defines each specific model. Assuming that errors are due to additive Gaussian noise, finding the maximum log-likelihood is equivalent to finding the minimum of the sum of squared errors (SSE) scaled by the variance [29], i.e.,

$$\frac{1}{2\sigma_e^2} \sum_{j=1, t=1}^{j=M, t=T-1} (f_1(x_1(j, t), x_2(j, t); \phi) - dx_1(j, t))^2 \tag{5}$$

For each number of terms m we determine the best fit model $f_1^*(X_1, X_2; \phi_1^*(m))$, where the parameter set $\phi_1^*(m)$ maximizes the log-likelihood over all models with m terms.

We assume that the noise variance σ_e^2 is equal to the data variance. If the noise variance is unknown, the posterior distribution will not be Gaussian and hence maximizing the likelihood might not be equivalent to minimizing the SSE. But we can use Monte Carlo methods to compute the log-likelihood value by integrating over all possible noise variances provided we know a prior distribution for σ_e^2 . Note that the sum of squared errors can be used to calculate the coefficient of determination or the R^2 value as

$$R^2 = 1 - \frac{SSE}{M(T-1)\sigma_e^2}$$

Instead of reporting R^2 values for our models, we report the maximum log-likelihood values

$$L_1(m) = \log P(dx_1|x_1, x_2, \phi_1^*(m))$$

In the second stage of our model selection algorithm, we choose the best model among those obtained in the first stage based on their ‘robustness’. Clearly, $L_1(m) \geq L_1(m-1)$ for all m , that is, the maximum log-likelihood value increases monotonically with number of terms, since each term allows an extra degree of freedom on curve fitting (*in extremis* the additional coefficient can be set zero to obtain the same likelihood). For a finite data set, this extra degree of freedom can fit artifactual patterns due to noise. As a result, reliance on $L_1(m)$ alone can lead to overfitting the data by selecting too many terms and thus accepting a model that accurately fits the existing data but that generalises poorly to unseen data and has little predictive power (see [31] Chapter 20).

To address this problem and evaluate the fit of these models, we adopt a Bayesian approach [31,37–40]. We calculate the Bayesian marginal-likelihood [31,41] or the *Bayes factor* $B_1(m)$ for the set of models which have the largest log-likelihood value among models with their respective number of terms. Note that ‘Bayes factor,’ which refers to a *ratio* of model likelihoods is used in Bayesian literature to compare pairs of models [37,42]. We use the same term in this paper to refer to the Bayesian marginal likelihood as defined above, with the understanding that this measure would have the same function as the Bayes factor when comparing between more than two models.

The Bayes factor compensates for the increase in the dimensions of the model search space by integrating over all parameter values [31], i.e.,

$$B_i(m) = \int_{\phi_i(m)} P(dx_i|x_1, x_2, \phi_i(m)) \pi(\phi_i(m)) d\phi_i(m) \tag{6}$$

where $i=1,2$. The Bayes factor is thus the likelihood averaged over the parameter space with a prior distribution defined by $\pi(\phi_i(m))$. The prior distribution says how likely we thought a particular parameter value was before we started the fitting procedure [29]. We choose a non-informative prior distribution, such that, $\pi(\phi_i(m))$ is uniform over a range of parameter values. This reflects our lack of knowledge about what parameter values might arise. In many social systems, domain experts may actually have prior knowledge and this can be directly incorporated into $\pi(\phi_i(m))$. We do not do incorporate such information here, but envisage that it could be an important part of combining our data-driven methodology with existing knowledge.

In our implementation the range of values for $\pi(\phi_i(m))$ is chosen to include all feasible values but to be small enough for the integral to be computed using Monte Carlo methods. $B_i(m)$ is computationally expensive to calculate. Therefore we use the two stage algorithm described above, since models of equal complexity (same number of terms in this context) can be more fairly evaluated in terms of their log-likelihood values.

We compute the Bayes factor for the models $f_1^*(X_1, X_2; \phi_1^*(m))$ with different numbers of terms m and obtain the function $B_1(m)$. Using this and other background information, we can choose the best model. For example, in the absence of any prior information, the model with the highest Bayes factor is the optimum choice. However, there may be theoretical considerations for the particular system being studied that makes another model preferable even if it does not have the highest Bayes factor. In this case, we could, in theory, incorporate this information when choosing the prior distribution for the parameters. But where that is not straightforward, we use the Bayes factor plot to serve as a guide in comparing between the models while choosing the most efficient one for the problem being studied.

We repeat the same process to obtain the Bayes factor plots for the $f_2^*(X_1, X_2; \phi_2^*(m))$ models, in order to calculate $B_2(m)$.

In our implementation, and the application given below, we choose to use a uniform prior distribution for $\pi(\phi_i(m))$. The non-informative prior is reasonable when we have no information about the parameter space except its likely range. If we know the parameter distribution is symmetric and the parameter is less likely to be a large value, then a normal distributed prior is a reasonable assumption. In this case, the Bayes factor can be evaluated directly and we do not need to use a Monte Carlo evaluation. Specifically, suppose $\phi_i(m) \sim N(0_m, \lambda^2 I_m)$. The integrand in equation 6 is then the product of normal distributions and the integral is itself a normally distributed random variable with mean $0_{M(T-1)}$ and

covariance matrix given by $\lambda^2 * (X * X') + \sigma_e^2 * I_{M(T-1)}$ evaluated at the observed values of dx_i [42]. Here X is the $M(T-1) \times m$ matrix of the observed values of x_1 and x_2 modified according to the terms chosen in the particular m -term model we are evaluating.

Systems with More than Two Variables

Most social systems have many interacting variables and we can easily extend our methodology to systems with more variables. In this case, we need to get the best possible models $f_i^*(X_1, X_2, \dots, X_N; \phi_i^*(m_i)), i = 1, \dots, N$ using $L_i(m)$ and $B_i(m)$ as described in the two variable case. The main constraint to this approach is that the number of possible models that need to be evaluated grows exponentially with number of variables. For models with three variables, with the polynomial functions that we look at, there are 39 possible models with one term and $\binom{39}{m}$ models with m terms. Due to the computational complexity involved, we only search the model space for models with up to 5 terms in this case.

One approach to solve this problem is to use a model pruning algorithm that will search only a fraction of the entire model space by looking only at models with higher terms that are extensions of the best models with fewer terms. For instance, suppose $f_1^*(X_1, \dots, X_N; \phi_1^*(2))$ was the best two term model and is represented by the non-zero coefficients a_{i1}, a_{i2} . Then, when evaluating three term models we only look at models that have these terms and an additional term a_{i3} . In this manner, we restrict the number of models that we need to evaluate to a very small number. Clearly this might not be optimal in all cases, so instead of extending only the best 2 term model, a better strategy would be to evaluate all 3 term models that are extensions of, say, the 10 best 2 term models.

Lag and Lead Times

In social systems, there is often a lag or lead effect in variables. A change in one variable at any instant t is often not a function of the level of another variable at t but of its level a few timesteps in the past. The approach as we describe above already captures autocorrelation, because we fit $X_i(t+1) - X_i(t)$ as a function of $X_i(t)$, but it does not capture second order effects where, for example

$$\frac{dX_1}{dt} = f_1(X_1(t), X_1(t-1), X_2(t), \dots)$$

To handle this issue, we can extend our approach by including as a new variable the time-lagged variable of interest. For instance, in a two variable system, if there is evidence to suggest that $dx_2(j, t)$ may be a function of both $x_1(j, t-1)$ and $x_1(j, t)$, a simple implementation would be to look at the three variable system comprising x_1, x_2, x_3 , where $x_3(j, t) = x_1(j, t-1), j = 1, \dots, M, t = 2, \dots, T-1$. Now we can compute the best possible models $f_1^*(X_1, X_2, X_3; \phi_1^*(m))$ and $f_2^*(X_1, X_2, X_3; \phi_2^*(m))$ to understand the effects of lagged variables.

Correlated Noise

In the methodology described above, best fit regressions on dx_i are performed assuming that the errors are uncorrelated to obtain the best possible models $f_i^*(X_1, X_2, \dots, X_N; \phi_i^*(m))$ with m terms. But often the distortion or the noise in the socio-economic process occurs due to systematic reasons, for instance due to latent

variables that affect all the model variables simultaneously, and hence is correlated across the variables. We have to account for the correlated noise in finding the best fit models and we handle this using the seemingly unrelated regressions approach [43].

Consider, for simplicity, the two variable system. The residual errors for the models $f_1(X_1, X_2; \phi_1)$ and $f_2(X_1, X_2; \phi_2)$ are given by $e_i = dx_i - f_i(x_1, x_2; \phi_i), i = 1, 2, t = 1, \dots, T-1$ where the e_i are $M \times (T-1)$ matrices. In performing least squares regression, we used the standard assumption that the errors are independent Gaussian processes. Thus, $e_1(j_1, t_1)$ the error in the model for the first variable for entity j_1 at time t_1 is independent of $e_2(j_2, t_2)$ the error in the model for the second variable for entity j_2 at time t_2 for any j_1, j_2, t_1, t_2 .

In the case of correlated noise processes, $e_1(j, t)$ and $e_2(j, t)$ are correlated whereas the errors in different entities and at different times are still uncorrelated. Thus in this model, we have the covariance matrix defined by

$$C = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{e2}^2 \end{bmatrix}$$

where the noise variances are σ_{e1}^2 and σ_{e2}^2 , and the noise covariance is given by σ_{12} . In the N variable case this is a symmetric $N \times N$ matrix with the elements defined similarly.

In the seemingly unrelated regressions approach, we obtain the regression coefficients $\phi_1(m_1)$ and $\phi_2(m_2)$ for the models f_1 and f_2 as before. To calculate the regression coefficients in the presence of correlated noise, we use these coefficients and compute the observed error covariance matrix as

$$\hat{C} = \begin{bmatrix} \hat{\sigma}_{e1}^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{e2}^2 \end{bmatrix}$$

where the diagonal terms are computed as

$$\hat{\sigma}_{e1}^2 = \frac{1}{M(T-1)} \sum_{j,t} e_1^2(j,t) - \left(\frac{1}{M(T-1)} \sum_{j,t} e_1(j,t) \right)^2$$

and similarly for $\hat{\sigma}_{e2}^2$. The off-diagonal elements are given by

$$\hat{\sigma}_{12} = \frac{1}{M(T-1)} \sum_{j,t} \left(e_1(j,t) - \frac{1}{M(T-1)} \sum_{j,t} e_1(j,t) \right) * \left(e_2(j,t) - \frac{1}{M(T-1)} \sum_{j,t} e_2(j,t) \right)$$

Using this estimated covariance matrix, we perform multiple regression using the standard methods for regressions in the presence of correlated noise [42] and re-estimate regression coefficients. Using the new coefficients we re-calculate the errors and the estimate of the noise covariance matrix as detailed above and iterate this process until the regression coefficients converge. This typically happens in a few steps, but if convergence is not achieved in a fixed number of steps a stopping criterion can be set.

If the underlying noise covariance matrix is almost diagonal, indicating that error terms are uncorrelated, the parameters estimated by the seemingly unrelated regressions approach will not

differ significantly from the parameters obtained assuming uncorrelated errors.

Results

We now investigate a frequently studied macro-level phenomenon, the relation between democracy and GDP per capita, using our proposed methodology [32–36]. We used data for GDP per capita that is provided by the World Bank (<http://data.worldbank.org>), and a democracy index [44], based on the Freedom House political rights and civil liberties scores [22,23], weighted by the countries’ human-rights-performance, measured by two indices from the Cingranelli/Richards Human Rights data project [24,25]. We modeled the interaction of democracy and GDP per capita for the years 1981–2006 for 74 countries.

We start our analysis with two variable models, in which changes in democracy (dD/dt) are a function of democracy itself (D) and log GDP per capita (G). Figure 2 shows that the maximum log-likelihood $L(m)$ for this model strictly increases with the number of terms m . The Bayes factor $B(m)$ grows significantly when we add a second term, decreases slightly for $m=3$ and $m=4$ terms, before increasing again for $m=5$ terms and decreasing finally when a sixth term is added.

The best two term model for changes in democracy is

$$\frac{dD}{dt} = 0.11G^3 - 0.067\frac{D}{G} \tag{7}$$

In words, this model tells us that democracy grows once GDP per capita has reached a certain threshold, with this threshold being determined by democracy itself. Specifically, democracy grows when $D > 1.64G^4$.

The best five term model, and the best overall model is

$$\frac{dD}{dt} = 0.077G^3 + 1.9D - 0.85\frac{D}{G} - 0.96DG - 0.14D^2 \tag{8}$$

As such complex models are usually more difficult to interpret in words, we might prefer the second best model with just two terms whose Bayes factor is only slightly smaller. Accepting simpler models makes interpretation of the interaction of GDP per capita and democracy more straightforward. On the other hand, the Bayes factor has already taken model complexity into account, and we should look carefully at what the more complex model tells us.

We can investigate the difference between the two and five term models by visualizing the functional form of the two dD models (see Fig. 3). In this figure, the blue colours represent where democracy decreases and the yellow/red colours where democracy increases. The threshold for transition from positive to negative growth is approximately the same in both models (black line in Fig. 3), although the five term model has a point of inflection. A more important difference is that, countries with high GDP but low democracy (bottom right-hand corner of Fig. 3), experience slower growth in democracy in the five term model than in the two term model. The additional flexibility of the five term model thus captures the slow growth in democracy of rich, but undemocratic countries.

Now that we understand better what impact GDP per capita has on democracy, we turn our attention to changes in GDP per capita, dG/dt . The best model (Fig. 4) for dG/dt as a function of GDP itself and democracy is,

$$\frac{dG}{dt} = 0.014 + 0.0064DG - 0.02G \tag{9}$$

The model shows that GDP is primarily growing at a constant rate, but is additionally positively affected by democracy interact-

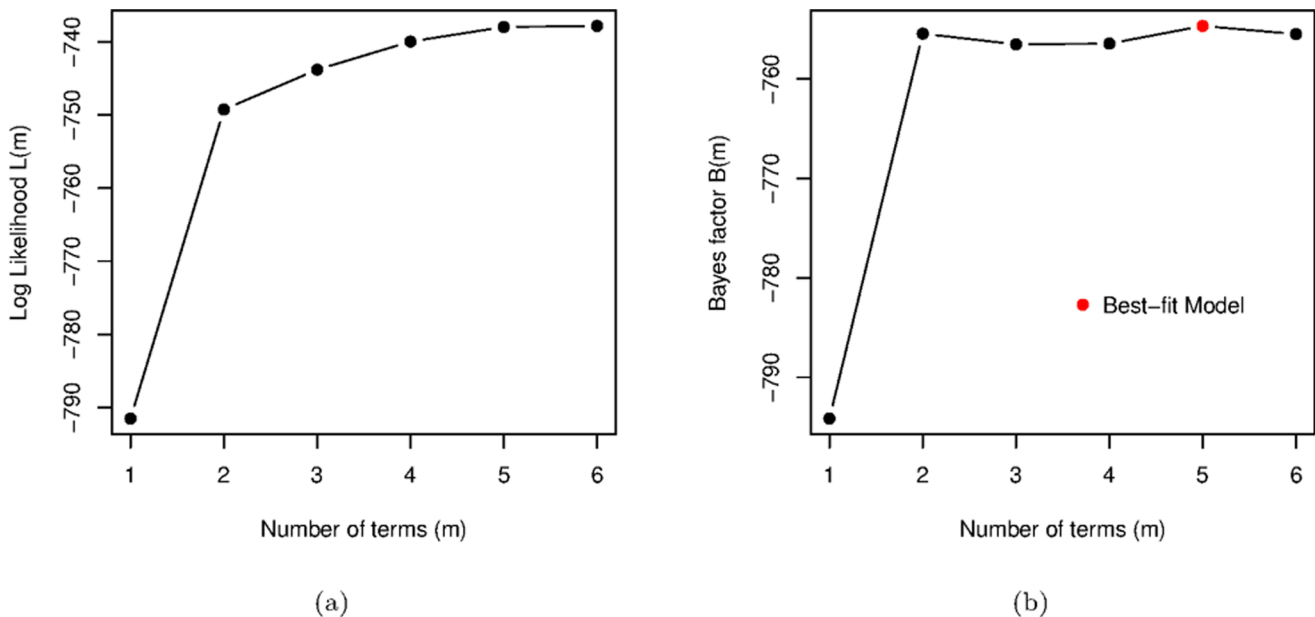


Figure 2. Change in Democracy models. (a) The Log Likelihoods and (b) Bayes Factors for the same models. doi:10.1371/journal.pone.0086468.g002

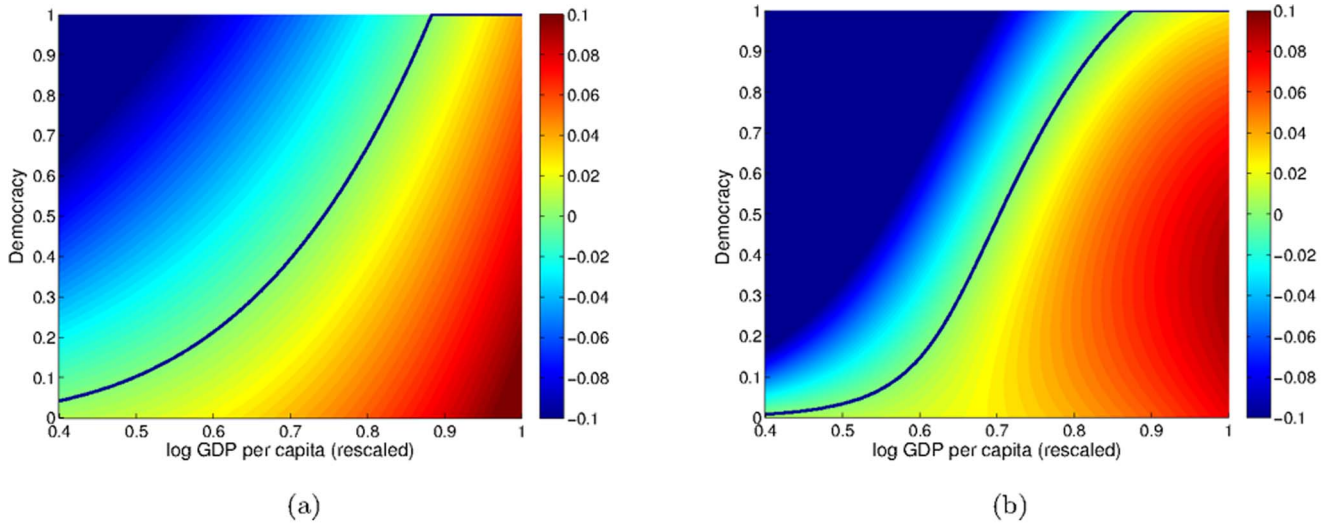


Figure 3. A heatmap of change in democracy models. For (a) the two term model given by equation 7 (b) the five term model given by equation 8. The colour scale gives the rate of change of democracy (i.e. $\frac{dD}{dt}$) as a function of D and G . The black line is the solution $\frac{dD}{dt} = 0$. doi:10.1371/journal.pone.0086468.g003

ing with GDP. Moreover, the growth is self-limiting at high levels of GDP.

With the interactions between democracy and GDP per capita identified, they can be displayed visually using a phase portrait. We provide phase portraits with example trajectories both from the data (Fig. 5a) and the model-predictions made by numerically integrating equations 7 and 9 (Fig. 5b). The model predicts quite well the general trajectories of the countries. For example, increases in democracy for Albania and Argentina and the initial decrease then increase in democracy in Bangladesh. The model is deterministic and therefore does not capture fluctuations seen in the data, but the overall pattern seen for specific individual countries is reproduced.

We now implement the various extensions presented in the methods section. We start by testing for correlated noise in the models. Computing the error correlations for equations 7 and 9

gives an error correlation of 0.113. This correlation is not big, but possibly not negligible either. Accounting for error correlation and re-estimating parameters results in changes of the coefficients:

$$\frac{dD}{dt} = 0.052G^3 - 0.036\frac{D}{G} \tag{10}$$

$$\frac{dG}{dt} = 0.024 + 0.0042DG - 0.018G \tag{11}$$

The re-estimated coefficients are only slightly different than those originally determined and do not make a significant difference to the phase portrait. Generally, it appears that the

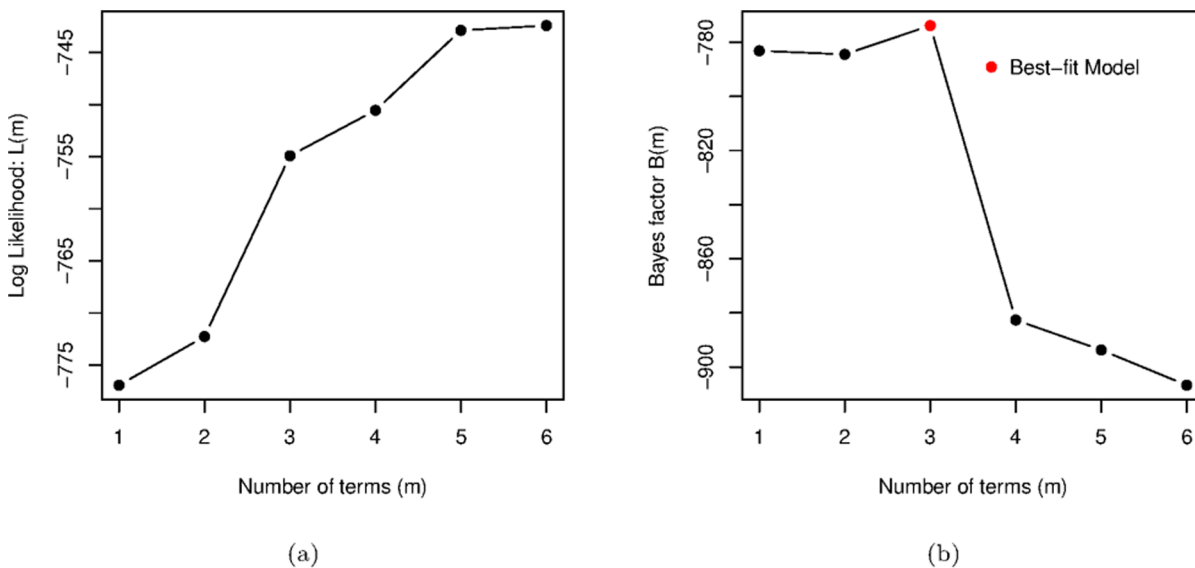


Figure 4. Changes in GDP per capita models. (a) The Log Likelihoods and (b) Bayes Factors for the same models. doi:10.1371/journal.pone.0086468.g004

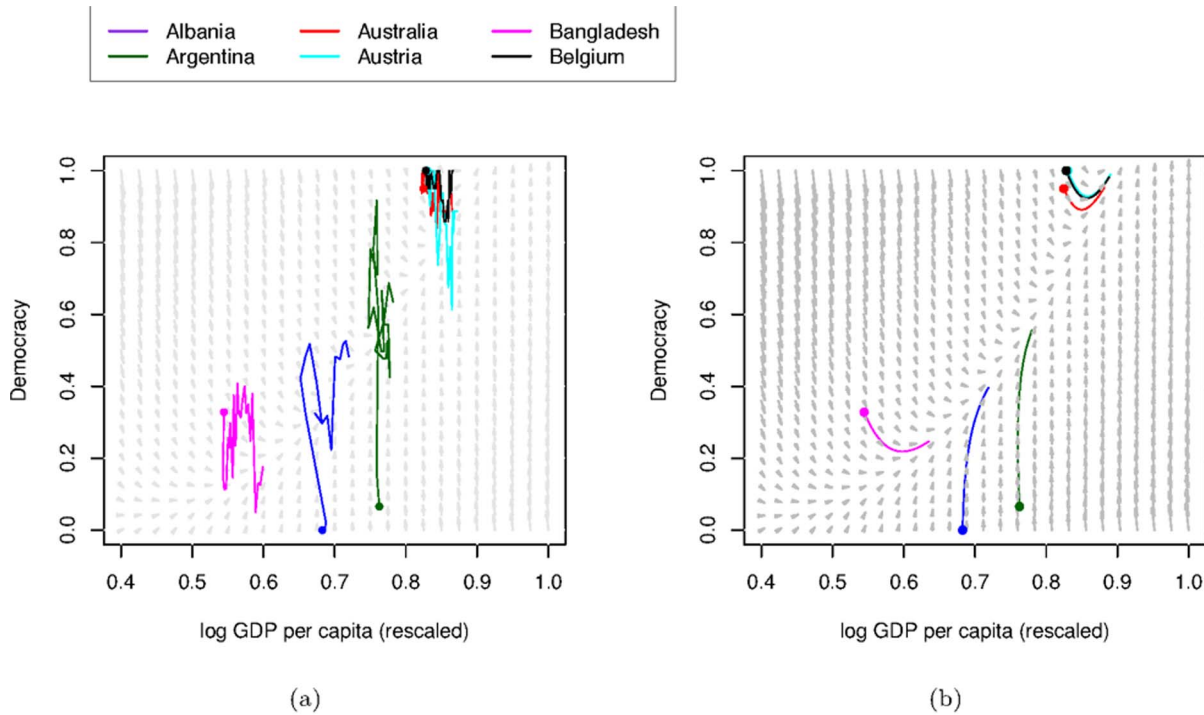


Figure 5. Phase portraits with trajectories. From (a) country by country data and (b) model predictions made by integrating equations 7 and 9, with coloured circles representing the initial conditions. The arrows in the phase portrait indicate the direction and magnitude of the changes dG and dD as a function of the two variables D and G themselves. Specifically the arrows are the vector (dG, dD) given by equations 7 and 9. These arrows are the same in both sub-figures.
doi:10.1371/journal.pone.0086468.g005

effects of both democracy on GDP per capita and of GDP on democracy are slightly overestimated using the original method.

A natural extension to three variables in this case is the inclusion of a lag. In particular, we might expect earlier levels of GDP to influence future changes in democracy. We therefore look at changes in democracy $dD = D_{t+1} - D_t$ as a function of itself (D_t), GDP per capita (G_t) and lagged GDP per capita (G_{t-1}). We now use the additional subscript to make clear the lags involved. Using G_{t-1} as an additional variable results in the following best model:

$$\frac{dD}{dt} = 0.085G_{t-1}^3 + 0.34\frac{D_t}{G_{t-1}} - 0.24\frac{D_t}{G_t G_{t-1}} - 0.14D^2 \quad (12)$$

The Bayes factor for both three variable lagged and two variable non-lagged models are shown in Fig. 6. While this model does have a better Bayes factor than both the two and five term non-lagged models, the difference is marginal. Moreover, equation 12 resembles the best five-term model for changes in democracy in equation 8. G_{t-1} has a positive non-linear effect on democracy, and democratic growth remains self-limiting. These observations lead us to conclude that the additional complexity added by the lagged term outweigh the marginal improvement in goodness of fit.

The best model (see Fig. 6) for changes in GDP per capita when fitting a three term model with G_{t-1} is

$$\frac{dG}{dt} = 9.7G_t^2 + 8.8G_{t-1}^2 - 18G_t G_{t-1} \quad (13)$$

Here, democracy is no longer a predictor for GDP per capita, which is now solely predicted by itself at different points in time. This is now a significantly different model than the one given by equation 9. Moreover, the model in equation 13 fits the data significantly better than the model given by equation 9 (see Fig. 6). It is possible that in the two variable models democracy, to an extent, played the role of the lag variable, rather than acting as a proper predictor of change in GDP. We then further note that factoring equation 13, we get

$$\frac{dG}{dt} \approx 9(G_t - G_{t-1})^2 \quad (14)$$

suggesting that the growth in G between times $t-1$ and t is the best predictor of changes between times t and $t+1$. This possibility is confirmed if we use $G_t - G_{t-1}$ as the third indicator variable. Now, fitting dG as a function of D_t, G_t and $G_t - G_{t-1}$ we find that the best model for changes in GDP per capita has only a single term,

$$\frac{dG}{dt} = 0.7G_t(G_t - G_{t-1}) \quad (15)$$

This models turns out to be the overall best model for changes in GDP per capita (Fig. 7).

For completeness, we also checked whether changes in democracy improve with $G_t - G_{t-1}$ as the third indicator. In this case, the best model is again equation 8, and the best two term model is equation 7, indicating that lags in GDP are not important in modeling changes in democracy.

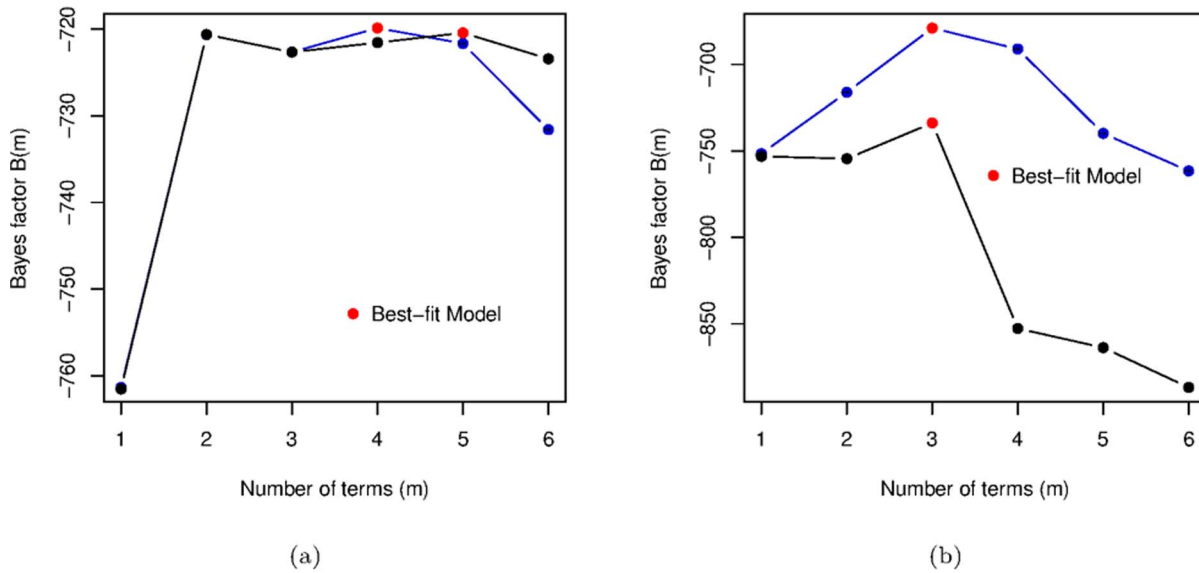


Figure 6. Choosing the best models. (a) Bayes factors for the two-variable dD models (black) and for the three-variable dD models (violet). The two red marked models refer to equation 8 and equation 12. (b) Bayes factors for the two-variable dG models (black) and for the three-variable dG models (violet). The two red marked models refer to equation 9 and equation 13. In both cases, the Bayes factor for the two and three variable models are calculated over the same subset of data, which is (necessarily) one year shorter than that used in Fig. 2 and Fig. 4. doi:10.1371/journal.pone.0086468.g006

Discussion

The approach to studying social systems we present here emphasizes exploratory model fitting. Exploratory approaches help to identify new and unexpected patterns and explanations [46–48]. Such an approach is not completely atheoretical, since we need to think about which variables to investigate. However, instead of defining how the variables should interact and then testing this pre-defined relation in the data, we allow the data to inform us about the mathematical linear or non-linear relation-

ships between the indicator variables. This does not imply that our approach is only applicable in cases where there is no existing theory. Indeed, if a theory can be expressed in terms of relations in a differential equation then it can also be compared, via Bayes factor, to other alternative models.

Our methodology can be applied to any social system which has reasonable amounts of longitudinal or panel data, that is data with repeated measurement over time for a number of independent entities. On the macro-level the method can be used to study cross-national development dynamics, for instance, the relationship between a country’s gross domestic product, child mortality and education levels. If regional or city district data is available it is possible to use the method to study for instance neighbourhood segregation processes. On a meso-level the researched entities could be organisations, companies or schools, to study, for instance, dynamic female employment patterns of companies.

In studying social systems there is seldom one single unique best fit model that fully explains the data. As we saw when comparing the Bayes factor of equations 7, 8 and 12, several models provided a robust fit to data. The advantage of using Bayes factor is that complicated models are automatically punished. Including more terms does not necessarily improve fit, since all of the extra parameters are included in the stochastic integration in equation 6. For example, the complex non-linear interaction (given by equation 8 and visualized in in Fig. 3) whereby rich, undemocratic countries showed slower democratic growth than some poorer and more democratic countries gives only a slightly more robust fit to data than the simpler two-term interaction (given by equation 7 and visualized in Fig. 3). At this stage, we should accept both as plausible models.

We do not attempt here to give a political or sociological interpretation of the best fit models. However, the dynamic nature of the fitted models provides a starting point for thinking about causal mechanisms [7,8]. For example, we see that the best model for explaining change in democracy as a function of economic development involves a threshold. When GDP is sufficiently high, democracy typically increases. This observation certainly provides

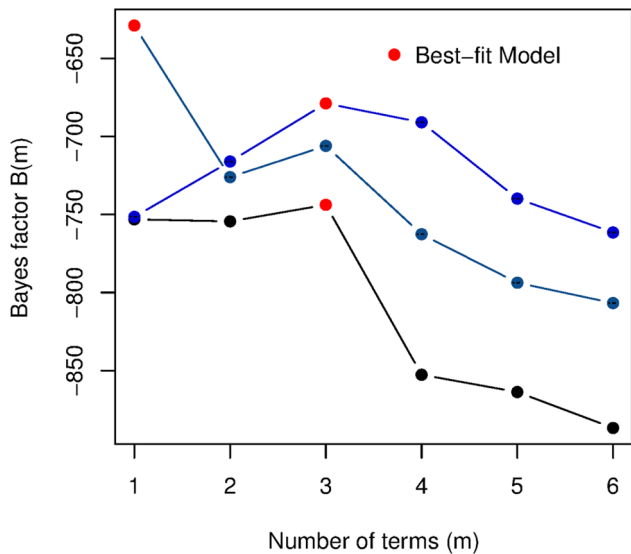


Figure 7. Bayes factor plots. For the two-variable dG models (black), the three-variable dG models with lag GDP per capita (G_{t-1}) (violet) and the three-variable dG models with $G_t - G_{t-1}$ as the third indicator variable (blue). doi:10.1371/journal.pone.0086468.g007

more understanding than the correlation presented in figure 1, but it does not fully open up the black box relationship between the two indicators. In order to open up this black box, we need to investigate both the role of other indicator variables in linking democracy and economic growth, as well as proposing possible micro-level mechanisms which can explain the transitions to democracy.

The suggestion that the terms in the fitted models could relate to plausible causal mechanisms is one argument for adopting an approach based on differential equations. If our aim was only to predict future changes in indicator values then any one of a range of statistical or machine learning techniques, such as Gaussian processes or neural networks, could be employed in model fitting. We could further employ Bayesian model averaging in which we weight models according to their Bayes factor [19]. While these techniques are undoubtedly of use in making predictions,

differential equations lend themselves more readily to interpretation. The use of differential equations has been the strength of earlier work using logistic growth to model the diffusion of innovations [10]. Fitting models has informed debate over micro-level mechanisms [12,13]. The methods outlined here provide a rigorous extension of the differential equation approach to a general set of interactions between multiple indicators.

Acknowledgments

The authors would like to acknowledge Stamatios C. Nicolis for numerous discussions that helped shape the current paper.

Author Contributions

Analyzed the data: SR VS RPM DJTS. Contributed reagents/materials/analysis tools: SR VS RPM DJTS. Wrote the paper: SR VS RPM DJTS.

References

- Sethi R, Somanathan R (2004) Inequality and segregation. *Journal of Political Economy* 112: 1296–1321.
- Welzel C, Inglehart R (2005) Liberalism, postmaterialism, and the growth of freedom. *International Review of Sociology* 15: 81–108.
- Lindensfors P, Jansson F, Sandberg M (2011) The cultural evolution of democracy: Saltational changes in a political regime landscape. *Plos One* 6: e28270.
- Durlauf SN, Johnson PA, Temple JR (2005) Growth econometrics. In: Aghion P, Durlauf SN, editors, *Handbook of Economic Growth*, Elsevier, volume 1. 555–677.
- Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *Journal of Consumer Research* 34: 441–458.
- Opp KD (2011) Modeling micro-macro relationships: Problems and solutions. *The Journal of Mathematical Sociology* 35: 209–234.
- Hedström P, Swedberg R (1998) Social mechanisms: an analytical approach to social theory. Cambridge: Cambridge University Press.
- Coleman JS (1986) Social theory, social research and a theory of action. *American Journal of Sociology* 91(6): 1309–1335.
- Raub W, Buskens V, Assen MV (2011) Micro-macro links and microfoundations in sociology. *Journal of Mathematical Sociology* 35: 1–25.
- Hamblyn RL, Jacobsen RB, Miller JLL (1973) A mathematical theory of social change. John Wiley & Sons.
- Rossmann G, Chiu MM, Mol JM (2008) Modeling diffusion of multiple innovations via multilevel diffusion curves: Payola in pop music radio. *Sociological Methodology* 38: 201–230.
- Friedkin NE (2010) A multilevel event history model of social diffusion: Medical innovation revisited. *Journal of Mathematical Sociology* 34: 146–155.
- Van den Bulte C, Lilien G (2001) Medical innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology* 106: 1409–1435.
- Huckfeldt RR, Kohfeldt CW, Likens TW (1982) *Dynamic Modeling: An Introduction. Quantitative Applications in the Social Sciences*. SAGE.
- Brown C (2007) *Differential Equations: A Modeling Approach. Quantitative Applications in the Social Sciences*. SAGE.
- Epstein JM (1997) *Nonlinear Dynamics, Mathematical Biology, And Social Sciences*. Santa Fe Institute Series. Westview Press.
- Sala-i Martin X (1997) I just ran four million regressions. NBER Working Papers.
- Fernandez C, Ley E, Steel MF (2001) Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16: 563–576.
- Ley E, Steel MF (2009) On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* 24: 651–674.
- WorldBank (2009) World development indicators. Available: <http://data.worldbank.org>. Accessed September 2013.
- UNDP (2011) Available: <http://www.undp.org>. Accessed September 2013.
- FreedomHouse (2008) Freedom in the world. report. Available: www.freedomhouse.org. Accessed September 2013.
- FreedomHouse (2010) Freedom in the world. report. Available: www.freedomhouse.org. Accessed September 2013.
- Cingranelli J, Richards DL (2010) Ciri dataset 2008. Available: ciri.binghamton-n.edu/index.asp. Accessed September 2013.
- WorldValueSurvey (2012) World value survey. Available: <http://www.wvsvsdb.com/wvs/WVSDData.jsp>. Accessed September 2013.
- Kuppens P, Oravecz Z, Tuerlinckx F (2010) Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of personality and social psychology* 99: 1042.
- Lodewyckx T, Tuerlinckx F, Kuppens P, Allen NB, Sheeber L (2011) A hierarchical state space approach to affective dynamics. *Journal of mathematical psychology* 55: 68–83.
- Breiman L (2001) Statistical modeling: The two cultures. *Statistical Science* 16: 199–215.
- Bishop CM (2006) *Pattern recognition and machine learning*. Springer.
- Rasmussen CE, Williams CKI (2006) *Gaussian Processes for Machine Learning*. The M.I.T Press.
- MacKay DJC (2003) *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.
- Lipset SM (1959) Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review* 53: 69–105.
- Diamond L, Marks G (1992) *Reexamining Democracy*. SAGE.
- Barro RJ (1999) Determinants of democracy. *Journal of Political Economy* 107: 158–183.
- Boix C, Stokes SC (2003) Endogenous democratization. *World politics* 55: 517–549.
- Kriekhaus J (2003) The regime debate revisited: A sensitive analysis of democracy's economic effect. *British Journal of Political Science* 34: 635–655.
- Jeffreys H (1939) *Theory of Probability*. Oxford: Oxford University Press.
- Berger J (1985) *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.
- Jaynes E (2003) *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.
- Bernardo JM, Smith AFM (2009) *Bayesian Theory*. Chichester: John Wiley & Sons.
- Skilling J (2006) Nested sampling for general bayesian computation. *Bayesian Analysis* 1: 833–859.
- Robert CP (1994) *The Bayesian Choice: a decision-theoretic motivation*. New York: Springer-Verlag.
- Amemiya T (1985) *Advanced econometrics*. Oxford: Blackwell.
- Welzel C (2013) *Freedom Rising, Human Empowerment and the Quest for Emancipation*. Cambridge University Press.
- Cingranelli J, Richards DL (1999) Respect for human rights after the end of the cold war. *Journal of Peace Research* 36: 511–534.
- Gelman A (2004) Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics* 13 (4): 755–779.
- Stebbins RA (2001) *Exploratory Research in Social Sciences*. London: SAGE.
- Tukey J (1977) *Exploratory Data Analysis*. Addison-Wesley.