Data-driven Modeling in the Social Sciences - A pragmatic approach for policy-makers

Shyam Ranganathan
Department of Mathematics, Uppsala University
Box 480, Uppsala 75106
shyam@math.uu.se

Viktoria Spaiser
Institute for Futures Studies
Hollandargatan 13, Stockholm
viktoria.spaiser@iffs.se

Stamatios C. Nicolis
Department of Mathematics, Uppsala University
Box 480, Uppsala 75106
snicolis@math.uu.se

Ranjula Bali Swain
Department of Economics, Uppsala University
Box 513, Uppsala 75120
ranjula.bali@nek.uu.se

David J. T. Sumpter
Department of Mathematics, Uppsala University
Box 480, Uppsala 75106
david.sumpter@math.uu.se

Warning the social sciences against "excessive ambitions," Elster [8] talks about the "physics envy" of social scientists who try to come up with the fundamental laws of social interactions. In the last few decades researchers have tried to formulate physics-like fundamental laws that explain macro-level phenomena such as economic growth, ethnic segregation, cultural values etc and have proposed different theories to explain the underlying mechanisms for those macro-level phenomena with varying success [7, 19, 28, 29, 33]. It is now widely thought that understanding at the micro-level is the key to causal mechanisms in the social sciences [5, 13, 26].

The problem of theorising in the social sciences is that there is often no consensus on the underlying mechanisms of macro phenomena. Different micro processes and mechanisms are suggested to explain the same macro behaviour. The dilemma becomes even more aggravated by the fact that data is often provided on the macro-level only. Be it economic growth, education indicators or social data, the current data sets describe regions or countries, rather than individual people. Overall, this makes it difficult for policy makers to find practical solutions to political and social problems and to arrive at a consensus on how to expend resources in a meaningful way based on predictable outcomes.

In this abstract we suggest an approach to understand social systems through macro-level data alone. This is especially useful for political decision-making because the key problem is one of predicting the outcome of specific policy interventions not for single individuals but for a society as a whole. For the policy-maker, a model needs to provide a usable and reasonably accurate prediction, even if it does not capture all the detailed mechanisms of the underlying process. The short-term dynamics and the non-linear interactions between variables are perhaps the most significant aspects of complex social systems for the purposes of predictive accuracy. Identifying such interactions can give an initial insight into macro-level relationships [3, 9, 14].

Econometrics already 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 [10, 18, 27]. 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. This allows us to predict economic growth accurately if the other variables are specified exactly but the problem is that these covariates also have rates of change that are affected by economic growth and by each other. It is precisely these dynamic feedbacks which are of most interest in the social sciences and where reliable statistical approaches are required [7].

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 economics [32, 34], social development [32], political systems [4, 11, 12] and cultural values [35] of different countries and regions. Identifying relationships between these macro-level indicators poses new challenges, but also opens up new opportunities. This applies not only to national level macro dimension but equivalently to other levels of aggregation (for instance neighbourhood, organisation) and may even be applicable to individual panel data [17, 21].

In [24], we proposed a new approach inspired by machine-learning and algorithmic modeling. We use the available socio-economic indicator data to select between a pool of feasible differential equation models of indicator interactions. Adopting a Bayesian model selection approach ensured that the selected models were reliable and robust. We illustrated our approach on the classic problem of determining an interaction between GDP per capita and democracy (Fig. 1), which are known to correlate but do not have a known causal link [1, 2, 6, 16, 20].

We then went on to look at the more general problem of understanding the transition to democracy and its relation to economic and social development and cultural values [30]. We showed how the data seems to indicate interactions between the components of the Human Development Index (HDI) and measures of emancipative values [33] and

KDD Workshop on Data Science for Social Good 2014, New York

democracy. The complex model suggests that HDI, and particularly the interaction of the HDI components GNI per capita and education, has a positive effect on changes in the democracy index. Rising emancipative values on the other hand are a result of the interaction of HDI (and specifically life expectancy) and democracy. Emancipative values on their own contribute to improvement of life expectancy and education (specifically female education). But high levels of emancipative values, which we usually find in economically and democratically highly developed societies, seem to slow down further economic growth, as hight levels of emancipative values indicate a shift from materialistic to post-materialistic values and priorities. (Fig. 2).

This model (Fig. 2, see further details in [30]), that we derived from the data, seems to suggest necessary theoretical changes to a popular political science theory on democratic transition – the Human Development Sequence (HDS) – which suggested that economic security led to changes in emancipative values which then resulted in a democratic transition [15]. [31] further explores this idea in terms of the micro-level interactions and shows a plausible mechanism using simulations by which these transitions may be brought about.

Extensions to models on the demographic transition and sustainable development have also been recently undertaken using the same methods [22, 25]. The result of this exploratory data analysis is to show that simple and efficient differential equation models can capture significant amounts of information about the variables of interest while preserving the most important dynamics and mutual feedbacks in the complex system (Fig. 2). While these models do not necessarily lean heavily on existing theory or provide complete and explicit mechanistic explanations, we have shown that theoretical constraints can easily be incorporated where necessary [22]. At the same time, these models provide sufficient predictive accuracy that is essential to policymakers. In [25], for instance, we show that the Millennium Development Goals (MDG), which have come in for criticism from various quarters, were excessively arbitrary and that using the dynamical systems approach that we propose would have led to the setting of systematic and meaningful goals (Fig. 3). In [22] we argue that this should be taken up in the next round of setting of the Sustainable Development Goals (SDG) to provide useful outcomes.

Currently we are working on the role of stochasticity in these systems. Since the models proposed are, in general non-linear, the time dependent behaviour can be complex. Very different trajectories are possible on very different time scales, depending solely on the initial conditions considered. A small random disturbance, accounting for social and/or political uncertainties, may drive the system on very different trajectories. We present results from a three variable model of democracy based on the HDI, emancipative values and the democracy index itself [30] and show how certain initial conditions result in more variability at the same level of noise (Fig. 4).

The methodology that we have described in detail in [24] comes with a toolbox [23] for social scientists with panel or longitudinal data that allows to do exploratory analysis with these data to identify sets of insightful mathematical models for social, economic and political processes. These models can easily include theoretical constraints but they always capture the most important dynamic changes and the non-linearities that are of most significance to policy-makers. The examples cited above show the wide applicability of the approach and we hope that refinements and extensions through further research will provide a complete solution for policy-makers interested in any specific social problem.

1. REFERENCES

[1] R. J. Barro. Determinants of democracy. *Journal of Political Economy*, 107:158–183, 1999.

[2] C. Boix and S. C. Stokes. Endogenous democratization. *World politics*, 55(4):517–549, 2003.

[3] C. Brown. *Differential Equations: A Modeling Approach*. Quantitative Applications in the Social Sciences. SAGE, 2007.

[4] J. Cingranelli and D. L. Richards. Ciri dataset 2008. ciri.binghamton.edu/index.asp, 2010. Accessed September 2013.

[5] J. S. Coleman. Social theory, social research and a theory of action. *American Journal of Sociology*, 91(6):1309–1335, 1986.

[6] L. Diamond and G. Marks. *Reexamining Democracy*. SAGE, 1992.

[7] S. N. Durlauf, P. A. Johnson, and J. R. Temple. Growth econometrics. In P. Aghion and S. N. Durlauf, editors, *Handbook of Economic Growth*, volume 1, pages 555–677. Elsevier, 2005.

[8] J. Elster. Excessive ambitions. *Capitalism and Society*, 4(2), 2009.

[9] J. M. Epstein. *Nonlinear Dynamics, Mathematical Biology, And Social Sciences*. Santa Fe Institute Series. Westview Press, 1997.

[10] C. Fernandez, E. Ley, and M. F. Steel. Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16:563–576, 2001.

[11] FreedomHouse. Freedom in the world. report. www.freedomhouse.org, 2008. Accessed September 2013.

[12] FreedomHouse. Freedom in the world. report. www.freedomhouse.org, 2010. Accessed September 2013.

[13] P. Hedström and R. Swedberg. *Social mechanisms: an analytical approach to social theory*. Cambridge University Press, Cambridge, 1998.

[14] R. R. Huckfeldt, C. W. Kohfeld, and T. W. Likens. *Dynamic Modeling: An Introduction*. Quantitative Applications in the Social Sciences. SAGE, 1982.

[15] R. Inglehart. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton University Press, 1997.

[16] J. Krieckhaus. The regime debate revisited: A sensitive analysis of democracy's economic effect. *British Journal of Political Science*, 34:635–655, 2003.

[17] P. Kuppens, Z. Oravecz, and F. Tuerlinckx. Feelings change: accounting for individual differences in the temporal dynamics of affect. *Journal of personality and social psychology*, 99(6):1042, 2010.

[18] E. Ley and M. F. Steel. On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24:651–674, 2009.

[19] P. Lindenfors, F. Jansson, and M. Sandberg. The

cultural evolution of democracy: Saltational changes in a political regime landscape. *Plos One*, 6:e28270, 2011.

[20] S. M. Lipset. Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review*, 53:69–105, 1959.

[21] T. Lodewyckx, F. Tuerlinckx, P. Kuppens, N. B. Allen, and L. Sheeber. A hierarchical state space approach to affective dynamics. *Journal of mathematical psychology*, 55(1):68–83, 2011.

[22] S. Ranganathan, R. Bali Swain, and D. J. Sumpter. A dynamical systems approach to modeling human development. *in preparation*.

[23] S. Ranganathan, V. Spaiser, R. P. Mann, and D. J. Sumpter. Package bdynsys. http://cran.r-project.org/web/packages/bdynsys/index.html, 2013. Updated 2014.

[24] S. Ranganathan, V. Spaiser, R. P. Mann, and D. J. Sumpter. Bayesian dynamical systems modelling in the social sciences. *PloS one*, 9(1):e86468, 2014.

[25] S. Ranganathan and R. B. Swain. Analysing mechanisms for meeting global emissions target-a dynamical systems approach. *Stanford Center for International Development Working Paper 491*.

[26] W. Raub, V. Buskens, and M. V. Assen. Micro-macro links and microfoundations in sociology. *Journal of Mathematical Sociology*, 35:1–25, 2011.

[27] X. Sala-i Martin. I just ran four million regressions. NBER Working Papers, 1997.

[28] T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[29] R. Sethi and R. Somanathan. Inequality and segregation. *Journal of Political Economy*, 112:1296–1321, 2004.

[30] V. Spaiser, S. Ranganathan, R. P. Mann, and D. J. Sumpter. The dynamics of democracy, development and cultural values. *PloS one*, 9(6):e97856, 2014.

[31] D. J. Sumpter and V. Spaiser. Non-linearity, causality and the micro-macro link. *in preparation*.

[32] UNDP. http://www.undp.org, 2011. Accessed September 2013.

[33] C. Welzel and R. Inglehart. Liberalism, postmaterialism, and the growth of freedom. *International Review of Sociology*, 15(1):81–108, 2005.

[34] WorldBank. World development indicators. http://data.worldbank.org, 2009. Accessed September 2013.

[35] WorldValueSurvey. World value survey. http://www.wvsevsdb.com/wvs/WVSData.jsp, 2012. Accessed September 2013.

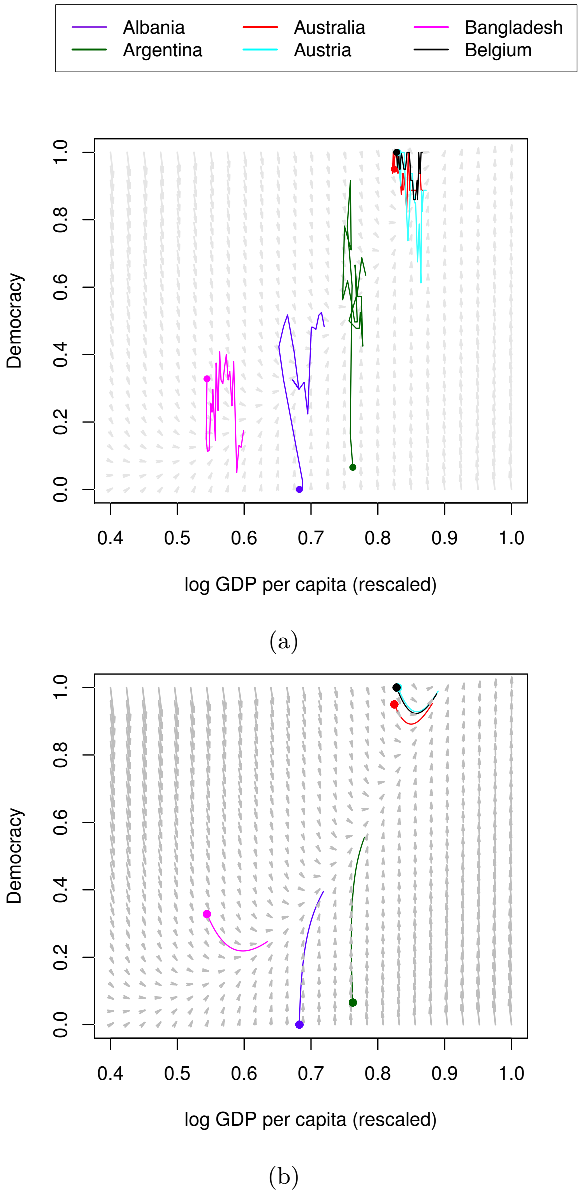

Figure 1: **Phase portraits with trajectories from (a) country by country data and (b) model predictions made by integrating differential equations, that we obtained using the Bayesian Dynamical Systems approach [24], with coloured circles representing the initial conditions. The arrows in the phase portrait indicate the direction and magnitude of the changes in GDP (dG) per capita and democracy (dD) as a function of the two variables democracy and GDP per capita themselves. Specifically the arrows are the vector (dG, dD) given by the obtained, best-fit differential equations [24]**

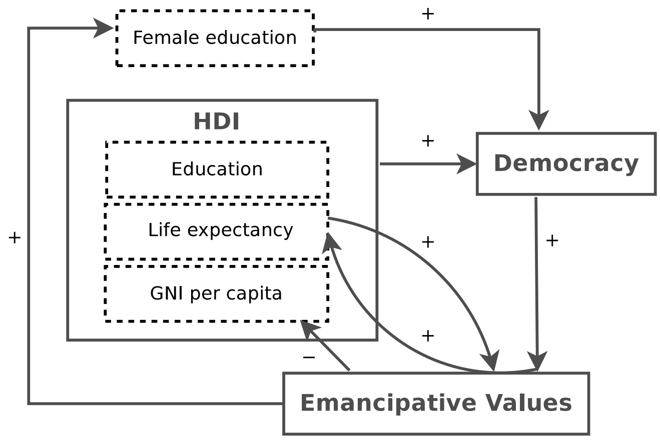

Figure 2: **A data-driven model of human development**

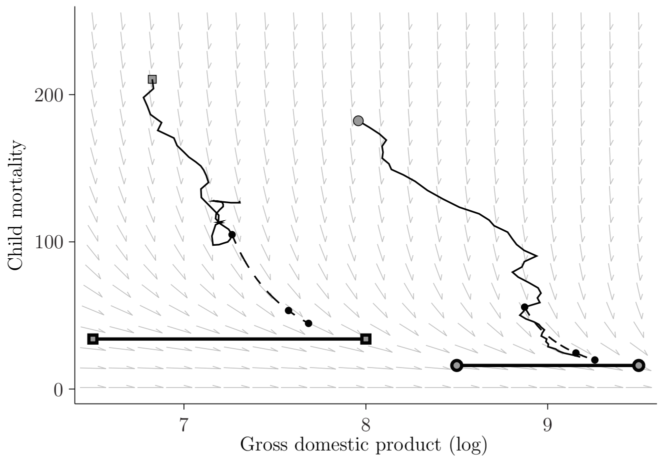

Figure 3: **Development of Kenya (solid diamond) and Brazil (hollow diamond) between 1959 and 2009. The solid lines are the model predictons, based on data of all countries in the world prior to 1990. The prediction trajectory uses 1990 data for the two countries as initial conditions and integrates forward to 2015 (target year for MDG). The MDG target was to reduce by two-thirds the child mortality from 1990 levels by 2015, i.e. down to 34 for Kenya and 16 for Brazil. These are shown as horizontal lines.**

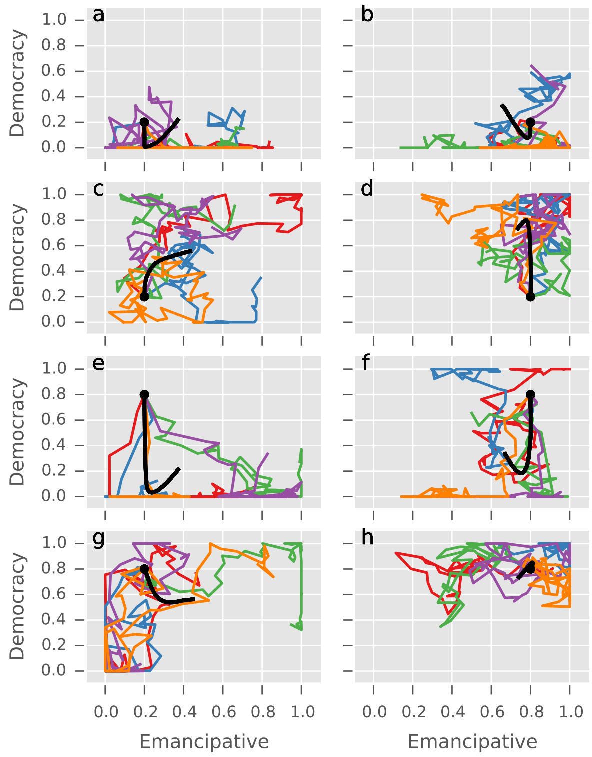

Figure 4: **Trajectories taken by different countries in the plane of Democracy and Emancipative Values in the presence of noise for the non-linear model described in Spaiser et al [30]. The different sub-plots show countries at different initial conditions for the variables D, E and H. D and E are represented in the plot itself and each pair of sub-plots in the same column a and c, b and d, e and g, and f and h represent respectively low H and high H values at the initial condition. The trajectories show the evolution of countries over a period of 50 years with each coloured line showing one particular realisation while the black line shows the mean of 100 such realisations. The sub-plots indicate that certain initial conditions are more sensitive to noise than others and allow for very different historical trajectories**