SOCIAL PREDICTIONS: FROM CONCEPTS AND DATA TO RULES

Péter Érdi
perdi@kzoo.edu

Henry R. Luce Professor
Center for Complex Systems Studies
Kalamazoo College
http://people.kzoo.edu/perdi/

and

Institute for Particle and Nuclear Physics, Wigner Research Centre, Hungarian Academy of Sciences, Budapest
http://cneuro.rmki.kfki.hu/
SOCIAL PREDICTIONS: FROM CONCEPTS AND DATA TO RULES

- Data, Rules, Predictions: Lessons from Tycho de Brahe, Kepler, Newton, Le Verrier and Adams
- Towards a Predictive Theory of the US National Budget
- Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network.
- Towards a computational social science: from data mining via social simulations to prediction
The world of Tycho Brahe: **DATA COLLECTION** (induction starts here)
Kepler: **MATHEMATICAL** not locally but globally predictive: integral laws.

Newton’s laws: **PREDICTIVE gravitation + differential laws.** When an external force acts on a body of constant mass then the acceleration produced is directly proportional to the force.

The discovery of Neptun: Adams vs Le Verrie. Galle
Towards a Predictive Theory of the US National Budget

Incrementalist model

• (concept driven) decision making model (bounded rationality)
• players: Requesters and Appropriators
• $R_n$ : request in year $n$
• $B_n$ : budgetary allocation in year $n$
• external factors are neglected ("closed system")
• tested on data 1947 – 1963 (non-defense agencies)

\begin{align*}
R_n &= \beta B_{n-1} + \xi_n \quad (1) \\
B_n &= \gamma R_n + \zeta_n \quad (2) \\
B_n &= \delta B_{n-1} + \eta_n \quad (3)
\end{align*}

$$\delta = \gamma \beta$$

$$\eta_n = \gamma \xi_n + \zeta_n$$
Towards a Predictive Theory of the US National Budget

- Exponential Incrementalist model
- self-reinforcing mechanism
- linear positive feedback
- external factors are neglected ("closed system")

\[ \ln B_n = \ln B_0 + \lambda n = A_0 + \lambda n \quad (4) \]
Towards a Predictive Theory of the US National Budget

- Disrupted Exponential Incrementalist model
- Trendline: an idea from Ferenc Jánossy
- External factors can disrupt the internally-dominated, closed incremental system
- Critical junctures

Ferenc Jánossy was the most important Hungarian pioneer of surveys on long time series. In the 1960s he devised the famous theory of trendlines, which allowed him to forecast the great world economic recession of the 1970s a decade in advance.
Towards a Predictive Theory of the US National Budget

• Back to Data!

• Treasury Department, 1791 – 1970

• Office of Management and Budget, 1940 – 2010

• Some inconsistencies: two separate synthetic series, OMB series is more reliable

• The top panel depicts the full historical period, while the bottom panel depicts the post-1950 period. The growth path for US expenditures is exponential, but major deviations occur. Especially noteworthy are the three abrupt ratchets and the distinct curvature after 1980.

• Changes in $A_0$ and $\lambda$
Towards a Predictive Theory of the US National Budget

Statistical Approaches

Method 1
is a smoothing technique applied to the budget series by taking the cumulative sum of the budget values—roughly the numerical integration of these values—allowing us to focus on the main trends in the data. We may think of this as kind of bird’s-eye view of the budget process.

Method 2
is an examination of rates of change instead of budgetary levels, again seeking deviations from the hypothesized exponential path. More specifically, we analyze the logarithm of year-to-year change ratio, \( \log(B(t)/B(t - 1)) \); the derivative of the logarithm of the budget.
Towards a Predictive Theory of the US National Budget

- Method 1 for **total budget**

- larger plotted line: the entire historical period,

- inset graph: isolated periods

- four distinct historical periods, separated by major events.

- Exponential "equilibrium" (more precisely constant acceleration) is indicated for three of them, but not for the period between the First and Second World Wars.
Towards a Predictive Theory of the US National Budget

When a department or agency actually spends money: draws money from the Treasury and gives it to someone - that’s a budget outlay. Federal budget is just a spending plan, outlays represent the real cost.

- Method 1 for Defense Outlay
  - Larger plotted line: the entire historical period,
  - Inset graph: isolated periods
  - Five distinct historical periods, separated by major events ...
  - President Theodore Roosevelt’s military expansion in around 1900, post-Vietnam withdrawal of military expenditures.
Method 1 for Domestic Outlay

- larger plotted line: the entire historical period,

- inset graph: isolated periods

There are periods in which exponential growth seems not to be exponentially stable for the domestic budget.
Fluctuations around the trend

skewness is a measure of the asymmetry of the probability distribution
kurtosis is any measure of the "peakedness" of the probability distribution

Figure 9: Histogram of the Logarithm of Budget Changes for the Full Data Series (1791-2010)
for Defense, Domestic, and the Total Budget
- **Residuals** from Fittings to the Derivative of Log Budget of Stable Periods
- Are critical junctions part of the general theory?
- If critical junctures are just part of a broader budgetary dynamics that characterizes the whole budgetary series
- then we expect frequency distributions examining only the distributions for the stable periods to resemble these distributions
- Analysis of the **residuals** in the stable periods.
- residuals are the random adjustments to the general trends
- Compared to the full series analysis presented in Figure 9, the histograms have fewer cases in the tails, and the kurtosis, which assesses punctuations in change data series, is reduced.
- Need for incorporating critical junctures into the theory
- But for some periods the kurtosis remains large, indicating budget punctuations within the stable period
Towards a Predictive Theory of the US National Budget

Fluctuations around the trend

Is the adjustment process stationary? This is true for the middle and late periods for the total budget, but not for the first period. But these deviations do not seem to have effects on the total budget path.
Towards a Predictive Theory of the US National Budget

- basic driver of budget change: self-reinforcing incremental system
- three major periods of budget stability
- critical junctures: changes in the intercept and slope
- after WWII: two stages: from the War to the late 1970s large slope
- followed by deceleration: the last period does not fit well to the general theory
- more general theory is needed to find a social Neptune
Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network

Patent Universe

Databases: USPTO and NBER

- 4 million patents
- 44 million citations

1790 First Patent (1975)

Hierarchical Organization

- **Individual**: Patents
- **Mesoscopic**: 450 classes – 120,000 subclasses
- **Macroscopic**: 6 categories – 36 sub-categories

EXTRACTION of GOVERNING RULES of NETWORK DEVELOPMENT and CLUSTER FORMATION

Microscopic Dynamics

Attractiveness depends on in-degree and age

Double Pareto distribution

linear preferential attachment

Mesoscopic – Macroscopic Analysis

Citation Vector

Non-assortative citations
Mesoscopic-macroscopic analysis

there ARE clusters!

Citation vector: predictor

Minoxidil - Citation distribution over years
Prediction of Emerging Technologies Based on Analysis of the U.S. Patent Citation Network

Emerging technologies can be identified as changes in the co-citation clusters.

Hot-spot detection
Extending the Definition of the Citation Vector

- the LENGTH of the vector tells, how important – well cited – a patent is
- the DIRECTION of the citation vector enables to compare the effect of external patent classes for two patents
Alternative Mechanisms for Technological Evolution

Mechanism identification - Birth

Motivation:
- growth
- contraction
- merging
- splitting
- death


Mechanism identification: Recombination

The rotation of the Citation Vector can be observed over time: it shows which external classes have the greatest effect on the Citation Vector over time.

Alternative Mechanisms for Technological Evolution – Example of Merging (Class 347)
Towards a computational social science: from data mining via social simulations to prediction

Computational social science is the interdisciplinary studies of complex social systems by combining

• social theories
• data mining
• dynamical modeling techniques

Conclusions

• Newton’ dynamical model: predictions led to the discovery of Neptune and Pluto
• What to do with this dynamical approach to social sciences?
• Social sciences: the overwhelming majority of data has been evaluated and interpreted in static or equilibrium perspective
• Something should be done with social data deluge
• The a priori knowledge of the rules, which basically govern the dynamic evolution of the systems, is difficult;
• By combining inductive and deductive strategies and using high-performance computing: we hope to predict our social Neptunes and Plutos
Bryan D. Jones is the J.J. ”Jake” Pickle Regent’s Chair in Congressional Studies in the Department of Government at the University of Texas at Austin. He is the Director of the NSF-Funded Policy Agendas Project, a web-based data system that monitors policy change in the US Federal Government since the Second World War.

László Zalányi (Wigner Res. Centre for Physics, Hungarian Acad. Sci. Budapest) is involved in almost every project.
Prediction of Emerging Technologies Using Patent Citation Network Analysis

Péter Érdi, Péter Bruck, István Réthy, Jan Tobochnik, Katherine Strandburg, Gábor Csárdi, Kinga Makovi, Péter Volf

Center for Complex Systems Studies, Kalamazoo College, Kalamazoo, MI 49006, USA
Complex Systems and Computational Neuroscience Group, Wigner Research Centre for Physics, Hungarian Academy of Sciences, Budapest, Hungary
Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, Hungary
Network and Subscriber Data Management, Nokia Siemens Networks, Budapest, Hungary
Department of Sociology, Columbia University, New York, NY, USA
New York University School of Law, New York, NY, USA