PREDICTION IN SOCIAL SCIENCE

How big data has – and hasn't – helped

Robert Northcott, Birkbeck College

Big Data in the Social Sciences, Kent, June 2017

INTRODUCTION

- Most of social science is <u>field</u> science
  - it studies uncontrolled phenomena outside the laboratory and therefore cannot run shielded experiments
- Models derived from general theory usually do not predict individual field cases accurately
  - ever-changing mix of sui generis (and thus unmodelled) causes
- But success is achieved sometimes

- What role for big data?

PLAN

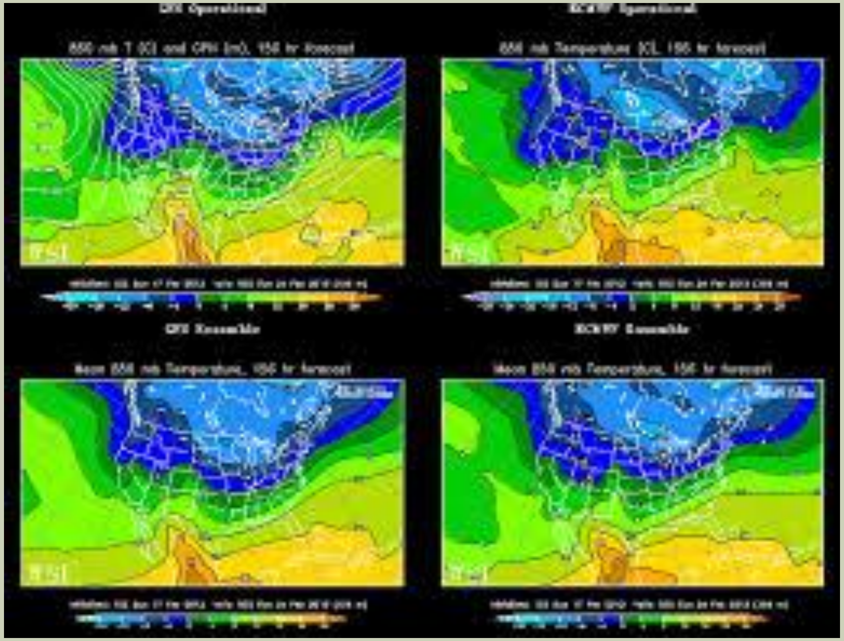
- I will go through a number of case studies of field predictions:
  - weather forecasting (Northcott forthcoming-a)
  - election prediction (Northcott 2015)
  - GDP forecasting (Betz 2006)
  - economic auctions (Guala 2005, Alexandrova 2008)

- Conclusion: big data can help, but only to some extent
- Underlying reason: lack of data is one – but not the only – constraint on predictive success
  - Arguably, similar claims apply across field sciences more widely

- At the end, tentatively: Causal models and other discussion

WEATHER FORECASTING

- Earth's weather system is:
  - chaotic (Lorenz 1969)
  - indefinitely sensitive to model errors too (Frigg et al 2014)
- Yet forecasting accuracy has improved significantly
  - hurricane paths predicted more accurately and further ahead
  - temperature and rainfall predictions are more accurate
- Overall, seven-day forecasts now are as good as three-day forecasts 20 years ago (Bechtold et al 2012)



WEATHER FORECASTING

- What explains this progress? Several factors together:
- 1) Data: huge improvement in quality and quantity since the launch of the first weather satellites in the 1960s
  - Temperature, humidity and other reports of ever greater refinement both horizontally (currently increments of 20km squares) and vertically (currently 91 separate altitude layers)
  - Over 10 million observations per day
- 2) Computing power: hugely increased
  - This enables ever more complex models to be used, ever more simulations to be run, and thus the new data to be exploited

WEATHER FORECASTING

- 3) Analytical methods: e.g. from late 1990s, models featured stochastic terms
- This has enabled the ensemble method: multiple simulations are run, generating probabilistic forecasts
- This overcomes the problem of chaos
  - experience has shown that, as in many chaotic systems, errors in individual simulations 'cancel out' over many iterations

WEATHER FORECASTING

- 4) <u>Models</u>: are based on Newton's equations of fluid dynamics, but those are not sufficient to generate accurate forecasts
- A whole series of additions have had to be made
  - These additions are under-determined by fundamental theory
  - They are determined instead by trial-and-error
  - (Implications for causal transparency – see later)

- These four sources of progress have interacted with each other:
  - The ensemble method of forecasting was not feasible until sufficient computing power became available
  - Increase in data and computing power have enabled more sophisticated models, although constrained by the need to run the required number of simulations quickly enough to generate timely forecasts
  - Experience of what data improvements most improve the accuracy of the model's predictions, has influenced the gathering of data, such as the choice of instruments on new satellites

WEATHER FORECASTING - SUMMARY

- Improvement, but:
  - 1) not due to data alone
  - 2) only to a limited extent

- What if data was unlimited?
- Still only probabilistic ensemble forecasts would be possible, plus:
  - unknown upper limit on level of accuracy
  - extra data must be collected by new physical instruments

ELECTION PREDICTION

- Two different approaches:
- 1) 'Fundamentals' models
  - Regression of variables such as GDP growth, unemployment etc
- 2) Opinion polling
  - Intelligent aggregation of polls predicts better
  - Good polling aggregation is sophisticated social science
- Is either approach successful?
- What role for data?

ELECTION PREDICTION

- My own main conclusions:
- Polling predicts better than fundamentals models
  - … but still imperfectly – as we have seen recently!
- Polling predicts better than before – more data has helped

- Fundamentals models do potentially offer causal transparency
- … but their lack of predictive success undermines this

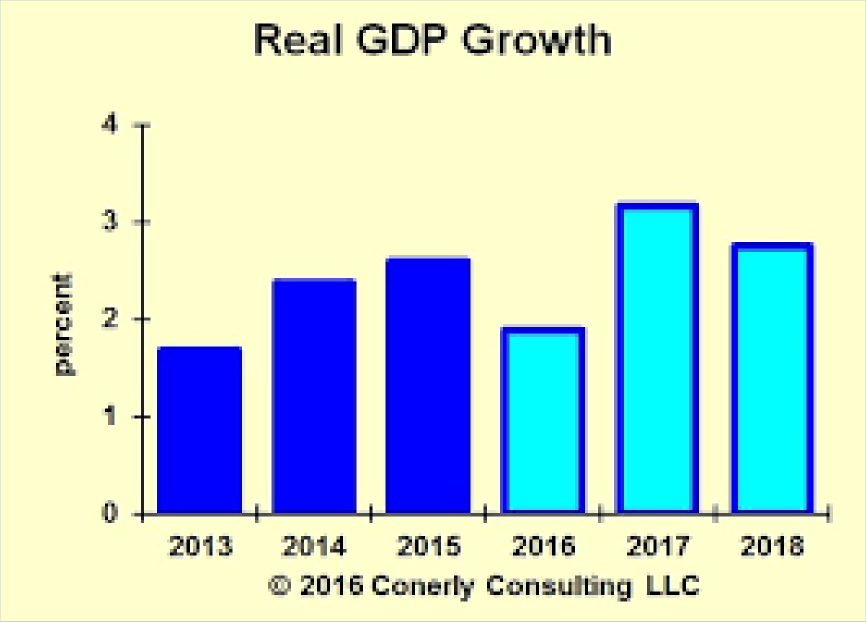
- Summary:
- Polling now offers decent prediction, but no explanation
- Fundamentals models give neither

ELECTION PREDICTION

- What if data was unlimited?
- Even then, predictive paradise would remain elusive:
- 1) Regardless, there is only a finite sample of past elections
  - limited possibility to ‘train’ models
- 2) Methods that predict well in one election don’t always predict well in the next one
  - non-stationary underlying causal process?
- Neither problem can be resolved just by gathering ever more data on voters’ preferences, demographics, consumptions, etc
  - sampling error is not the real problem here

GDP PREDICTION

- Naïve benchmark: assume that GDP growth will be the same this year as last year
  - Forecasts for 12 months ahead barely outperform this
  - Forecasts 18 months ahead don't outperform it at all
- Forecasts fail to predict turning points, i.e. when GDP growth changes sign
  - E.g. in 60 cases of negative growth, the consensus forecast was for negative growth on only three of those occasions
- (Even worse with exchange rates, stock prices, etc)



GDP PREDICTION

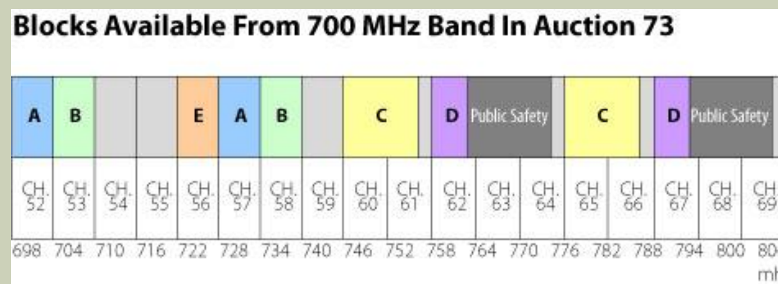
- Little or no sustained difference between different forecasters or different methods
  - purely numerical extrapolations, informal and formal
  - non-theory-based economic correlations, informal (indicators and surveys) and formal (multivariate time series)
  - theory-based econometric models, which sometimes feature hundreds or even thousands of equations
- Forecasting record has not improved over the last 50 years
  - despite more theory, data and computing power
- The induction: more data won't solve this, unlike weather case
  - Again, a non-stationary underlying causal process?
  - Other explanations: open system, reflexivity, chaos, bad theory, measurement errors – data alone won't solve these either

ECONOMIC AUCTIONS

- An example of a successful intervention/ creation of an artefact
- The 1994-6 US spectrum auctions raised huge sums, a triumph for its creators
  - 2000-1 UK ones as well
- A success in many ways
- Contrast with other such auctions:
  - New Zealand 1990
  - Australia 1993
  - Switzerland 2000

ECONOMIC AUCTIONS

- The particular auction design used was extremely intricate
  - Open vs sealed bid
  - Simultaneous vs sequential
  - Package vs individual
- As well as theoretical issues, there were many practical ones too
  - Bubbles
  - Interactive effects
  - Software and detailed procedure



Block	Area Type	Licenses	Frequencies (MHz)
A	Economic Area	176	698-704, 728-734
B	Cellular Market Area	734	704-710, 734-740
C	Regional Economic Area Grouping	12	746-757, 776-787
D	Nationwide	1*	758-763, 788-793
E	Economic Area	176	722-728

* Subject to conditions respecting a public-private partnership.

ECONOMIC AUCTIONS

- It was impossible to predict a particular design's efficacy from auction theory alone
  - or to predict an individual rule's impact either
- Rather, many experiments and *ad hoc* adjustments were required to fine-tune
- The design was constructed and tested:
  1) as a whole
  2) by trial-and-error

ECONOMIC AUCTIONS

- Progress came not from new theory, but rather from the new extra-theoretical work
  - That was the difference between the successful and unsuccessful cases

- It did not come from new data about bidders
  - Rather, the relevant new data was <u>experimental</u>

SUMMARY SO FAR

- 1) Weather: Prediction has improved, although still limited
  - Data has helped. More data may help more, although not unlimitedly
- 2) Elections: Prediction has improved, although still limited
  - Data has helped. More data likely will not help a lot more
- 3) GDP: Prediction has not improved, is very limited
  - More data is not helping at the moment
- 4) Auctions: Intervention has improved
  - Experimental data has helped

- So the overall picture is mixed:
- More data does help sometimes (of course!)
- But it is not obviously a panacea in any of the cases

- … Next, what of causal models?

CAUSAL TRANSPARENCY

- Often a <u>less</u> physically realistic weather model has been preferred, purely because it is more accurate predictively
  - Commercial imperative has focused minds methodologically
- Weather models are tested holistically
- Ubiquity of interactive effects means that the effect of a given tweak is not stable; it may alter once other parts are altered
  - "It is very difficult to understand how exactly changes in model formulation affect the climate of the model" (Jung et al 2010)
- As a result, the weather model is not causally transparent

- A similar holistic story with the auction design
  - No causal transparency there either

CAUSAL TRANSPARENCY

- In our cases, it's not generalizable causal models that predict successfully:
  - Weather – holistic ad hoc adjustments beyond theory
  - Elections – causal models out-predicted by polling
  - GDP – causal models no more successful than other methods
  - Auctions – a holistically built mechanism beyond theory
- Successful models typically local, i.e. context-specific
  - Tetlock 2015: predictive success typically is not generalizable
- So far, more data has not begun to mitigate this

CAUSAL TRANSPARENCY

- Two partial caveats:
- 1) Limited extrapolation in auction and election cases
- Lessons from one case did help in new cases:
  - UK auction 2000-1
  - Later US presidential elections
- But not infallibly so:
  - Switzerland auction 2000
  - US midterm elections
- Need new models each time
- 2) Some modularity in weather case
  - Occasionally possible to test if changes compose non-linearly

IS NATURE KIND?

- Do there exist cross-contextual causal regularities in field sciences, available to exploit?
  - If yes, big data can help discover them
  - If no, big data won't magically create them
- The evidence of our case studies is pessimistic about this
- Predictive success was only achieved by models that are:
  - Not causally transparent
  - Of limited generalizability

IS NATURE KIND?

- Arguably, this pessimistic pattern is common in field sciences (Reiss 2008, Northcott forthcoming-b):
- Generalizable causal models fail to predict accurately
- Causal relations are <u>fragile</u>, i.e. do not generalize much

WHAT ROLE FOR BIG DATA?

- When is big data epistemically helpful? Answer (Pietsch 2015):
  - Stable background conditions ('stationarity')
  - Sufficient relevant data
  - Correct causal vocabulary

- Well known big data success stories satisfy these conditions:
  - CRISPR biology
  - Facebook and Google internet experiments
  - Natural language translation
  - Many other examples (Mayer-Schoenberger & Cukier 2013)
- In these cases: stationarity plus lots of relevant data

- When stationarity and the other conditions are satisfied, the hype may be justified
  - big data can indeed make a big difference

WHAT ROLE FOR BIG DATA?

- Extrapolation <u>across</u> contexts: need for <u>causal</u> models:
  - Theory is required because predictive analytics will break down with non-stationary processes
  - ‘Look under the hood’
- I agree that the ‘death of theory’ hype breaks down here
- In field sciences, Pietsch’s conditions often are not satisfied
  - as in our case studies, and perhaps usually

- On the other hand: if nature is unkind, correct causal models will be hard to come by using any method, big data or not
  - Purely predictive models may then be the only way forward
  - A different version of ‘death of theory’ (Northcott forthcoming-a)

REFERENCES

- Alexandrova, A. (2008). ‘Making models count’, *Philosophy of Science* 75, 383-404.
- Bechtold P. et al (2012). ‘Progress in predicting tropical systems: the role of convection’, ECMWF Research Department Technical Memorandum no 686.
- Betz, G. (2006). *Prediction or Prophecy?* (Deutscher Universitaets-Verlag)
- Frigg, R., S. Bradley, H. Du, and L. Smith (2014). ‘Laplace’s Demon and the Adventures of His Apprentices’, *Philosophy of Science* 81, 31-59.
- Guala, F. (2005). *Methodology of Experimental Economics*. (Cambridge)
- Jung, T. et al (2010). ‘The ECMWF model climate: recent progress through improved physical parametrizations’, *Q. J. R. Meteorol. Soc.*, 136, 1145–1160.
- Lorenz, E. (1969). ‘Three Approaches to Atmospheric Predictability’, *Bulletin of the American Meteorological Society 50: 345–349.*
- Mayer-Schoenberger, V. and K. Cukier (2013). *Big Data*. (John Murray)
- Northcott, R. (2015). ‘Opinion polling and election predictions’, *Philosophy of Science* 82, 1260-1271.
- Northcott, R. (forthcoming-a). ‘When are purely predictive models best?’, *Disputatio*.
- Northcott, R. (forthcoming-b). ‘Economic Theory and Empirical Science’, in C. Heilmann and J. Reiss (eds) __
- Northcott, R. and A. Alexandrova (2015). ‘Prisoner’s Dilemma doesn’t explain much’, in M. Peterson (ed) *The Prisoner’s Dilemma* (Cambridge), pp64-84.
- Pietsch, W. (2015). ‘Aspects of Theory-Ladenness in Data-Intensive Science’, *Philosophy of Science* 82, 905-916.
- Reiss, J. (2008). *Error in Economics*. (Routledge)
- Tetlock, P. & D. Gardner (2015). *Superforecasting*. (Random House)

PURELY PREDICTIVE MODELS

	Explanation	No explanation
Empirically accurate	Slot 1: Newtonian cannonball	Slot 2: polling aggregation, weather forecasting
Not empirically accurate	Slot 3: (Empty)	Slot 4: fundamentals election prediction, much actual social science?

NOTES

- I'll only discuss epistemic issues, not, e.g., ethical ones
- I take field sci to be the relevant category, not socl sci
- Thus I am istd in fld scis genly r/t just socl scis
  - 'nature' here means social science nature
- I'll use 'prediction' and 'forecast' interchangeably
  - For Q&A: E no uniform anal of these
  - Prediction = in-sample conseqs of mdl; or extrapolation to new subjects; detstc fut earthquake claim; pbstc fut climate claim
  - Forecast = strictly re fut, out-of-sample data; based on past data for known subjects; pbstc fut earthquake claim; detstc fut weather claim
  - Also Projection = cdnl extrapolation ignoring possible fut non-stationarities (IPCC)
  - Scenario = a projection selon one mdl given one set of parameter values (IPCC)

NOTES

- For Q&A:
- … de facto epistemic merits of prediction in fld scis, incl their link to interventions (tho not ctrfctl ones)
  - Big data advocates themselves emph prediction
- … resolution of mdl errors problem in weather eg
  - Basically brute predictive efficacy – E no analytical soln
- … the ‘caveats to causal transparency’ slide
- … I agree w the hypists that the hypothetico-deductive method may not be apt for fld scis
- … big data can itself be used to tackle non-stationarity?
  - e.g. Amazon’s recommendation algorithm incorps depreciatn of data

NOTES F Q&A AND AFTER

- Pietsch presentation:
  - Emphs thl vs phenomenological sci distinction
  - Sees big data inductivism as follg Keynes-style method of predictive analogy

- Pietsch re my elections case:
  - What about the campaigns' use of targeted data?
  - Can such data predict results better than polls can?
  - In reply: but secret ballot means we don't know indivl voters' votes, so this is no route to getting more data pts than aggte results
  - E no public evid yet of better result prediction, malgre evid of campaigning effectiveness
  - Ctr-reply: but you can ask people who they voted for. OK, some may lie etc. But even imperfect accuracy cld be sffct to build an effective election-prediction mdl.
  - Ctr-ctr-reply: but what if E non-stationarity re reln b/w demographics n votes? All turns on the quality of evid motivating the campaigning tactics

- Genl: big data may enable better tradl hypoth-testing
  - An eg: Ogbonnaya's paper
  - Also Ghiara some genl argts for this

NOTES F Q&A AND AFTER

- A rdg ref:
  - Karin Knorr-Cetina, 'Epistemic cultures: how the sciences make knowledge', 1999
  - Apparently this contrasts physics n biology, somewhat along the lines of my th-predictive contrast
- Another rdg ref:
  - Leo Breiman, 'Two cultures of statl mdllg', 2001
  - Is Angrist/Pischke type advocacy re machine learning, by a statistician
- F Plato: Meno's Q – is the goal of sci kn or truth?
- … hmmm
- Genl: aim this paper at Studies HPS? E little novel phil wn it, rather only case studies