# Historical Analysis of National Wellbeing Using Digitized Text

**Daniel Sgroi**
University of Warwick, CAGE & IZA
daniel.sgroi@warwick.ac.uk

Joint work with Thomas Hills (Warwick), Eugenio Proto (Bristol) and Chanuki Seresinhe (Alan Turing Institute)

2nd April 2009

# The Meaning of Life?

- Going back at least as far as Aristotle and Confucius "happiness" has been an important concept and arguably even the meaning of life itself... although they had a very different understanding of what 'happiness" means.
- Happiness (or "subjective wellbeing") underpins much of economics but it has played a relatively minor role in the development and application of economic policy in the past.
- There is a growing literature on international patterns of subjective wellbeing: especially since Easterlin's famous "paradox" and the intense controversy surrounding it.
- Several nations including the UK, Australia, China, France and Canada now collect subjective wellbeing data to use alongside GDP in national measurement exercises. OECD & UN also active since 2011.

# A Parallel with GDP

- Development of GDP in the 1930s immediately following the Great Depression; Simon Kuznets (early developer) had different ideas about GDP (e.g., shouldn't include military spending or dis-services).
- Problems with GDP as a way to capture wellbeing:
  - Enviromnent: BP Deep Horizons oil spill increased US GDP.
  - Leisure is not included: wealthier people may choose to "buy" leisure but then income "falls".
  - Other issues: exchange rates, goods/output change over time, informal economy, illegal activity.
- Maddison Historical GDP Project rolls back GDP to the early 19th century, Broadberry et al going back much further for Britain and the Netherlands.
- But what about "National Happiness" data?

# Our Approach

- The availability of life satisfaction survey-based data typically dates back to the mid 1970s at best.
- Our primary objective is to produce a workable proxy for subjective wellbeing going back to 1800, which would enable direct comparisons with GDP over that period.
- Our methods rely on the digitization of books and newspapers, available in numerous corpora, such as the Google Books corpus, the British Newspaper project and the COHA corpora.
- We elected to start in 1800 because the number of digitized books and periodicals shrinks considerably before 1800.

# Word Norms or *Valence*

- The approach we take here is a common approach among studies seeking to infer public mood and relies on affective word norms to derive sentiment from text.
- For example, in a study of 17 million blog posts, (Nguyen et al, 2010) found that a simple calculation based on the weighted affective ratings of words was highly effective (70% accuracy) at predicting the mood of blogs compared against the ground truth provided by the bloggers.
- To make progress we need both a corpus of language (a source of text data) and a set of word norms (what individual words tell us about mood).

# Language Corpus Data

- The main language corpora we use is the *Google Books Ngram* Corpus (`https://books.google.com/ngrams`).
- The corpus is based on a digitized database of several million published books, which was developed as part of the Google Books programme.
- We will focus on data for 4 languages, English (British), English (American), German and Italian.
- We have some results for France and Spain but there are issues associating language with nationality.
- There are no word norms available for Chinese, Hebrew and Russian.

# Alternative Corpora

- To ensure we are robust to the specific corpus we also used:
  - "Find My Past" data from the British Library's "British Newspaper Project" which covers 65 million newspaper and periodical articles from the UK across 200 periodicals going back to 1710.
  - The US English COHA Corpora which includes 400 million words from 1810-2000.
  - Two alternative indices of sentiment (a "National Pleasantness Index" and "National Polarity Index") derived from SenticNet data.
- Since our results turn out to be robust to the choice of corpora in much of what follows we will focus on the Google book corpus.

# Word Norms

- In order to assess the valence of individual words, we used the largest available sets of existing word valence rating norms for each language.
- Word valence rating norms generally ask participants to rate each word from a list on how positive or negative they perceive a word to be.
- To allow for comparison across languages, all of our valence norms use a subset of words. There is a list of a thousand words that served as the basis for developing valence ratings for multiple languages through several independent studies.

# Affective Norms for Different Languages

- For English, ANEW contains about 10,000 words, all rated on a 1 to 9 valence scale by a group of subjects.
- For German, we used the Affective norms for German sentiment terms. This is a list of 1003 words, a German translations of the ANEW list. The valence ratings were collected on a -3 to +3 scale. The mean values were adjusted to reflect a 1 to 9 scale.
- For Italian, we used an adaptation of the ANEW norms containing 1121 Italian words, based on the ANEW material on a 1 to 9 scale.
- We also checked that our results are robust to the specific word norm: we replicated or findings using AFINN, another popular word norm used in psychology and linguistics.

## Valence and Words in Different Languages

- High end: Happiness 8.53, Enjoyment 8.37, Vacation 8.53, Joy 8.21, Relaxing 8.19, Peaceful 8, Lovemaking 7.95, Celebrate 7.84.
- Low end: Murder 1.48, Abuse 1.53, Die 1.67, Disease 1.68, Starvation 1.72, Stress 1.79, Unhappy 1.84, Hateful 1.9.
- Middle: Neutral 5.5, Converse 5.37, Eight 5.37, Century 5.36, Machinery 4.65, Platoon 4.65.

# Language Average Valence Computation

- For each language we compute the weighted valence score, $Valence_t$, for each year, $t$, using the valence, $v$ for each word, $j$, as follows,

$$Val_{i,t} = \sum_{j=1}^{n} v_{j,i} p_{j,i,t};$$

- Note that $v_{j,i}$ is the valence for word $j$ as found in the appropriate valence norms for language $i$, and $p_{j,i,t}$ is the proportion of word $j$ in year $t$ for the language $i$.

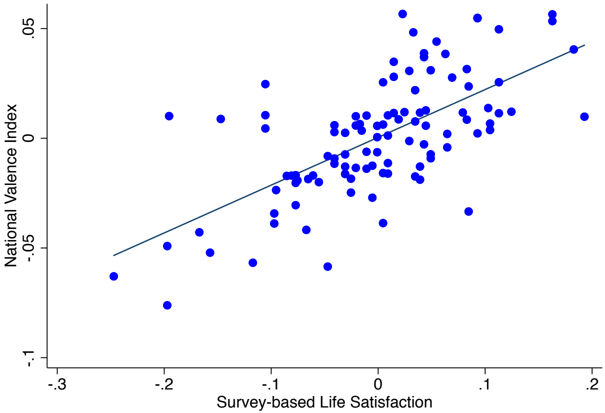
# The Evolution of Language

- Words have changed meanings over time: bad has meant good, dig has meant understand, etc.
- To control for this we constructed versions of our index that include only high stability words.
- The method boils down to looking at the neighbourhood of words: the argument being that when words change meaning they start to be used together with different words. High stability words keep the same neighbours.
- It turns out that our results are robust to using the full set of words, the top 25% or the top 50% and to variations on the stability method we use. This is likely to be because while some words do change meaning we use a large enough pool that this effect is small overall.

# How to Interpret the Index

- Think about the book market as highly competitive (lots of potential writers and publishers): publishers "match" books to demand.
- It could be that publishers match happy people to "happy books" or happy people to "sad books"?
- It could be that writers are inspired by the age in which they live: for instance a happy period inspires "happy books"?
- We will try to answer this question by comparing the available data on life satisfaction with our word-valence based index.
- The analysis in the paper involves lags (which makes sense for books), though we also duplicate everything for newspapers (and find similar results).

# Valence and Life Satisfaction Survey Data




correlation = 0.5508 (pvalue<0.01)

# Comparing the Data

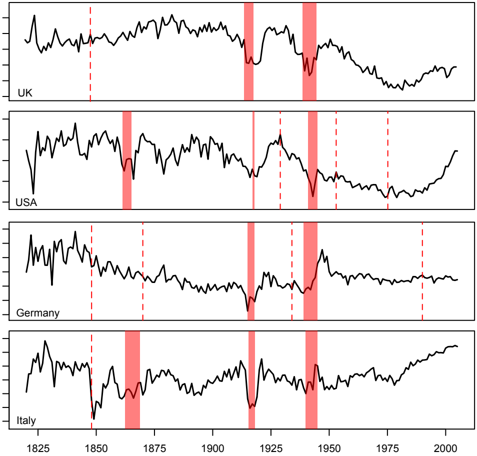
- The plot compares the Eurobarometer measure of life satisfaction with the word valence-based index for the period when the overlap (1973-2009) for the UK, Germany and Italy.
- Both variables (the National Valence Index and Eurobarometer Life Satisfaction measures) are expressed in the form of residuals after controlling for country fixed-effects, so that values represent variations around the averages for each of the three countries.
- A similar plot is generated if we compare our index with US life satisfaction data taken from the World Database of Happiness.
- It looks like life satisfaction and the content of books are positively correlated: so happy books go hand-in-hand with happy periods of time.

# Valence Predicts Aggregate Life Satisfaction

Table: Average life satisfaction per country and year is the dependent variable.

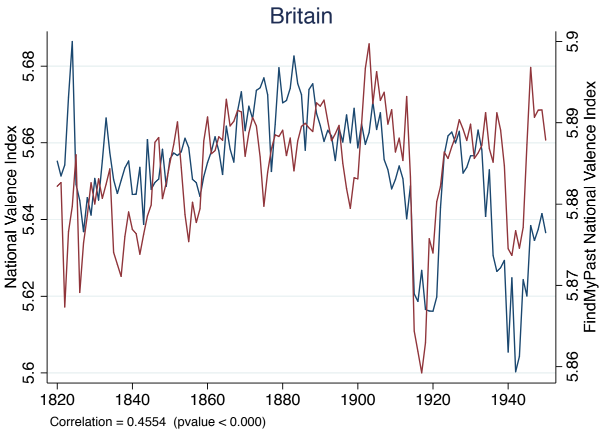
| | 1 | 2 |
|---|---|---|
| | Year FE | CS trends |
| | b/se | b/se |
| National Valence Index | 2.8551*** | 1.6596** |
| | (0.2867) | (0.2246) |
| GDP | Yes | Yes |
| Country Specific Trend | No | Yes |
| Year FE | Yes | No |
| | | |
| r2 | 0.730 | 0.588 |
| N | 104 | 104 |

# A Time-Series Plot of the NVI, 1820-2009



# Comparing Newspapers and Books for the UK, 1820-2009

Note: Blue is the book-based NVI, red is based on newspapers.



# Econometrics

- Next up we see what has mattered in the determination of the NVI in the past.
- First however, we need to note some issues:
    - Long-run biases might emerge from country-specific factors such as culture, language, religion and demographics (immigration, population age structure). We can control these to some extent through country fixed effects.
    - Literacy was lower in the past, Language different. We control for education, trends, year fixed effect.
    - To help with freedom of the press, we control for democracy.

# Historical Determinants of the Valence Index, 1820-2009

Table: The countries included are Germany, Italy, the UK and the United States

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| | Year FE | Year FE | Year FE | CS Trends |
| | b/se | b/se | b/se | b/se |
| (log) GDP(t-5) | 0.0826*** | | 0.0698*** | 0.0550** |
| | (0.0090) | | (0.0106) | (0.0130) |
| Life Expectancy(t-1) | | 0.0048** | 0.0030 | 0.0016 |
| | | (0.0013) | (0.0014) | (0.0013) |
| Internal Conflict(t-1) | | | | –0.0184** |
| | | | | (0.0040) |
| Words Covered(t) | Yes | Yes | Yes | Yes |
| Democracy(t) | Yes | Yes | Yes | Yes |
| Education Inequality(t) | Yes | Yes | Yes | Yes |
| Year FE | Yes | Yes | Yes | No |
| Country-Specific Trends | No | No | No | Yes |
| r2 | 0.752 | 0.705 | 0.774 | 0.571 |
| N | 412 | 412 | 412 | 412 |

# Summary of the Main Findings

- Our index based on average word valence of a language predicts country aggregate subjective wellbeing for several countries.
- But more than that it can go back much further than existing measures.
- Our index correlates positively with life expectancy, GDP (mildly) and negatively with conflict.
- Our findings are robust to different corpora (books, newspapers) and word norms.
- Our findings are also robust to the stability of word meanings over time.

# Why this Matters

- The concept of "National Happiness" is important but there is a paucity of historical data: we help to fix that problem.
- More generally, the methods discussed in this talk can be applied much more widely than in the happiness economics literature (and we are working on more right now).
- This also represents a combination of "Big Data", increases in computational power and interdisciplinarity (Economics, Psychology and Computational Linguistics are represented in the team) which is perhaps a foretaste of one future direction for economics as a discipline.