Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books

Daniel Sgroi
daniel.sgroi@warwick.ac.uk

-

Joint work with
Thomas Hills and **Eugenio Proto**

January 13, 2017

The Need for an Index of Subjective Wellbeing

- Subjective wellbeing (or "happiness") has played a minor role in the development and application of economic policy in the past.
- Recent call for a dashboard of indicators (Stiglitz Commission, OECD Better Life Index, UN World Happiness Report).
- Many nations now collect subjective wellbeing data to use alongside GDP in national measurement exercises.
- But it's difficult to know how to interpret these, because we have very limited time-series.

Why We Need Long-Run Data

- Understanding what has driven happiness in the past.
- Wars, epidemics, depressions, natural disasters occur infrequently.
- Now is fine, we can just ask people and we have this data available from the 1970s.
- But the past, surely impossible? How can we "go back in time" to ask our great grandparents how happy they are?

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:
  - Environment: 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 (centuries), informal economies.

A Parallel with GDP

- But these issues do not forestall the need to roll back GDP figures to better understand the evolution of national income and its drivers.
- Consider the original motivation for national income accounting.
- Maddison Historical GDP Project rolls back GDP to the early 19th century, Broadberry et al going back much further for Britain and the Netherlands.

Our Approach

- Our primary objective is to produce a workable proxy for subjective wellbeing going back to 1776, which would enable direct comparisons with GDP over that period.
- Our methods rely on the digitization of books, available in the Google Books corpus.
- We elected to start in 1776, for several reasons:
  - 1776 is the date of the American Declaration of Independence, one of the most famous of all historical documents to specifically reference happiness.
  - American Revolutionary War (1775-83) and the French Revolution (1789) as key events denoting the start of the modern era.

Prediction from Written Texts

- Inferring mood from text is commonplace endeavour now: psychiatrists, market researchers, security services...
- Inferring public mood (i.e., sentiment) from large collections of written text represents a growing scientific endeavour:
  - recovering large-scale opinions about political candidates
  - predicting stock market trends,
  - understanding diurnal and seasonal mood variation
  - detecting the social spread of collective emotions,
  - and understanding the impact of events with the potential for large-scale societal impact such as celebrity deaths, earthquakes, and economic bailouts .

Valence

- The approach we take here is a common approach among the studies described above and relies on affective word norms to derive sentiment from text.
- 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.
- Another weighted average technique based on word valence, coined the *Hedonometer*, was created by Dodds and Danforth (2010) and has been used successfully to recover sentiment from songs, blogs, presidential speeches, and temporal patterns of happiness using Tweets.

Language Corpus Data

- The language source we used is the *Google Books Ngram* Corpus `https://books.google.com/ngrams`
- Overall, this data represents about 6% of all books ever published.
- The corpus is based on a digitized database of physically published books, which was developed as part of the Google Books programme.
- We analysed data for 6 languages, English (British), English (American), German, Italian, Spanish, French.
- There are no "word norms" available for Chinese, Hebrew or Russian.

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. This 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.
- 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.
- the French and Spanish norms were also adaptations of the ANEW. These contained 1031 and 1034 words respectively. Both used a 1 to 9 points 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.

Example Words

- 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.

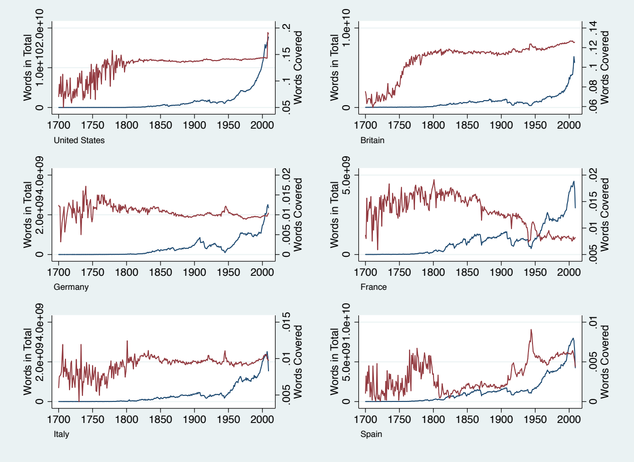
Language Average Valence Computation

- 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,

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

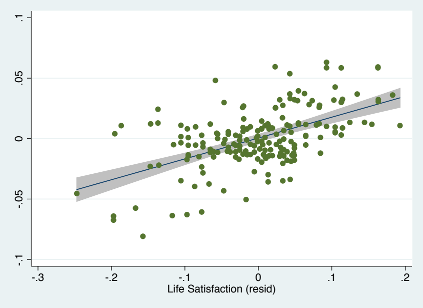
- Note that $v_{j,i}$ is the valence for word j as found in the appropriate valence norms for language i, and $p_{j,i,t}$ is the proportion of word j in year t for the language i.

Words and Words Covered



Valence and Aggregate Life Satisfaction.

Figure: Residual of the average Life satisfaction and of the Valence for the period 1972-2009 for France, Germany, Italy, Spain, UK. The residuals are calculated after regressing valence and life satisfaction against the country dummies.



Valence Predicts Aggregate Life Satisfaction

Table: Average life satisfaction per country and year is the dependent variable. Coefficients are in standard deviations.

	1	2	3	4
	Year FE	with GDP	until 2009	W/O Spain and France
	b/se	b/se	b/se	b/se
Valence	1.4646***	1.3795***	1.3892***	2.1837***
	(0.3535)	(0.3847)	(0.2483)	(0.3453)
Log GDP		0.1747	0.2186	0.5076
		(0.3102)	(0.2327)	(0.3624)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Words Covered	Yes	Yes	Yes	Yes
r2	0.903	0.903	0.904	0.953
N	119	119	163	78

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 do sad people demand happy books?
- Given the comparison with survey measures, the former seems right.

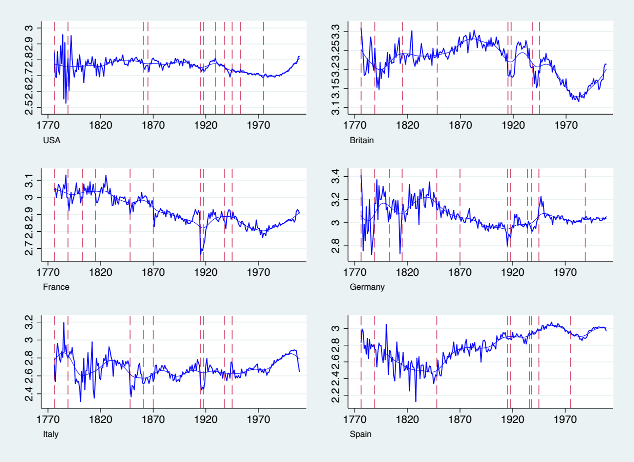
Data Concerns

- 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. So we cannot go back too far, plus we should (and do) control for education (and democracy).
- Freedom of the press (note France in WWII: we can use the index to pick up state control via cross-country comparison).

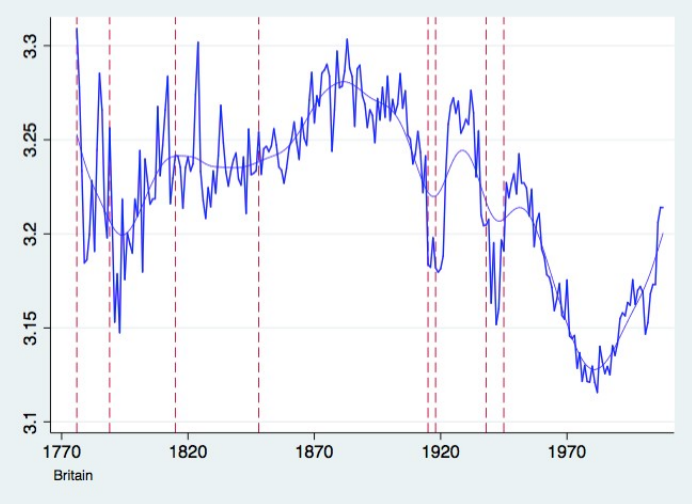
The Evolution of Literature

- What about the content of books changing? Year fixed effects helps if content evolved in a similar way across different countries, otherwise country-specific trends can help if content at least changed in a linear way (but differently across countries).
- The countries we have selected are similar in terms of economic development and literary evolution, note the advent of "Literary Realism":
  - George Eliot (1819 - 1880) in UK;
  - Mark Twain (1835 - 1910) in US;
  - Honore de Balzac (1799 - 1850) in France;
  - Theodor Fontaine (1819 - 1898) in Germany;
  - Benito Perez Galdos (1843 - 1920) in Spain;
  - Alessandro Manzoni (1785 - 1873) in Italy.

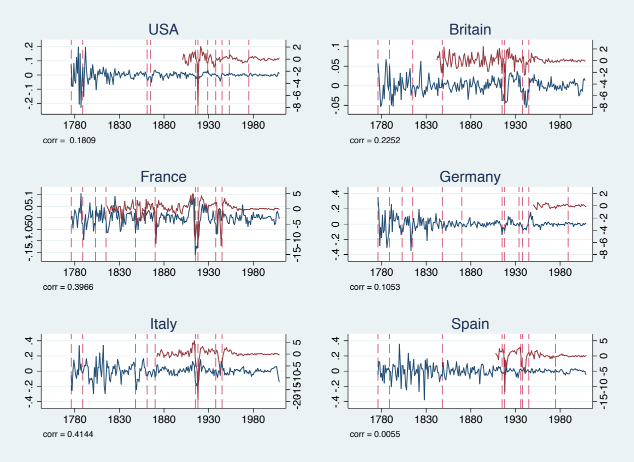
The HPS Index over Time



The UK



HP-filtered HPS Index and Life Expectancy



Historical Determinants of Estimated Subjective Wellbeing, with Detrended Data

	1	2	3	4
	Baseline	Controls	Infant Mort.	Conflicts
	b/se	b/se	b/se	b/se
Life Expectancy t-1	0.0057***	0.0065***	0.0062***	0.0039***
	(0.0012)	(0.0014)	(0.0014)	(0.0009)
GDP (log) t-1	−0.0586	−0.0657	−0.0724	−0.0525
	(0.0378)	(0.0437)	(0.0387)	(0.0263)
Infant Mortality t-1			−0.0004**	−0.0002
			(0.0001)	(0.0002)
Internal Conflict^{-} t-1				−0.0012
				(0.0015)
External Conflict$^{-}t-1$				0.0018
				(0.0025)
WW1 t-1				−0.0746**
				(0.0227)
WW2 t-1				0.0077
				(0.0136)
Democracy		−0.0026**	−0.0027**	−0.0021**
		(0.0010)	(0.0009)	(0.0007)
Education Inequality		0.0005	−0.0003	−0.0000
		(0.0010)	(0.0015)	(0.0012)
Words Covered		0.6402	0.7196	3.1986
		(0.6591)	(0.7141)	(4.0215)
Country FE	Yes	Yes	Yes	Yes
r2	0.126	0.166	0.175	0.296
N	765	648	633	586

Conclusions

- Average Word Valence (the HPS index) of a language predicts country aggregate subjective wellbeing of the corresponding country well when we have both.
- Average word valence positively correlates with life expectancy, not with income.
- Big shock events (Great Depression, World Wars) have a huge impact but recoveries are rapid.
- We stress caution when making super-long-run comparisons for many reasons, not least aspirations and the fact that happiness tends to be relative.
- There is a lot still to do: methodological advances, new applications.