

# Historical Analysis of National Wellbeing Using Millions of Digitized Books

**Thomas Hills Eugenio Proto Daniel Sgroi**

November 16, 2015

# Subjective Wellbeing and Gross Domestic Product

- 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 a very limited history

# GDP, History and Concerns

- 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 disservices)
- Problems with GDP as a way to capture wellbeing:
  - BP Deep Horizons oil spill increased US GDP; spending savings increases GDP temporarily
  - Leisure is not included: wealthier people may choose to “buy” leisure but then income “falls”.
  - Other issues: exchange rates, goods change over time (centuries), informal economies, complexity, sustainability.
- Need to roll back GDP figures (e.g. Maddison Historical GDP Project going back to AD 1)

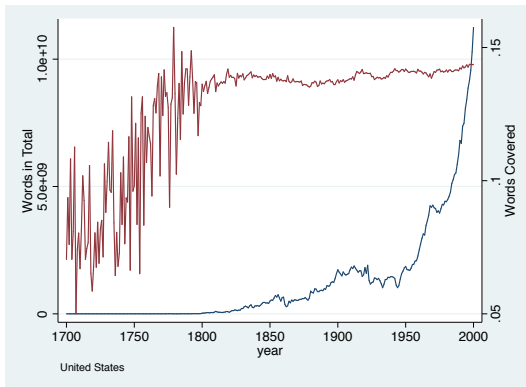
## Our approach

- Our primary objective is to produce a workable proxy for subjective wellbeing going back to 1776 (as a complement to GDP).
- Inferring mood (i.e, sentiment) from text (not new: marketing, psychology, etc.).
- We use Google Ngrams and word norms for valence 6 languages English (British), English (American), German, Italian, Spanish, French.
- Words rated on a scale from least to most positive (about 1000 words for most languages), all languages based on same set of words.

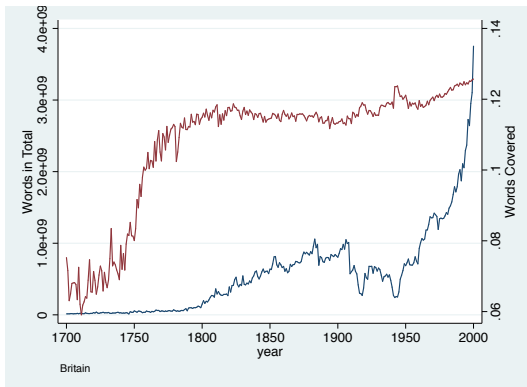
# How to Interpret the Index

- Think about the book market as highly competitive (lots of potential writers and publishers).
- It could be that publishers match book sentiment to the market (happy people want happy books) or not (sad people want happy books).
- Our index is very well correlated with actual survey measures from the 1970s onwards which suggest the former is right.
- So we can "roll it back" before the 1970s.
- Though the further back we go the more we might worry about how accurate comparisons will be (parallel with national income).

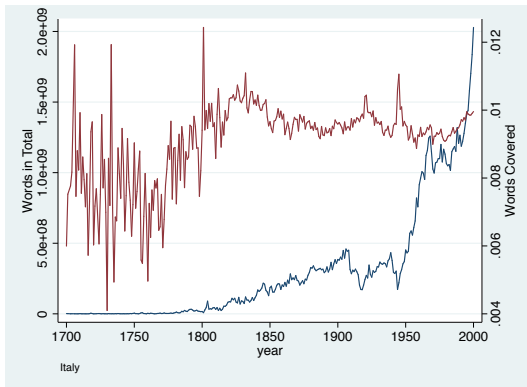
# Words and Words Covered: USA



# Words and Words Covered: UK

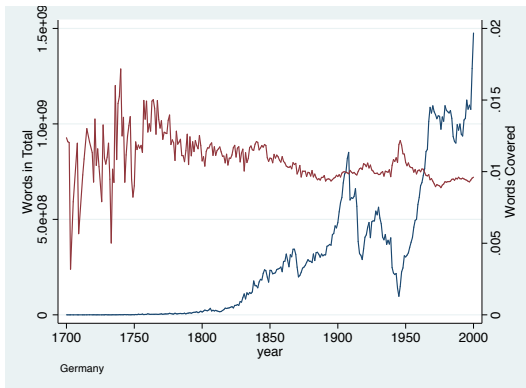


# Words and Words Covered: Italy

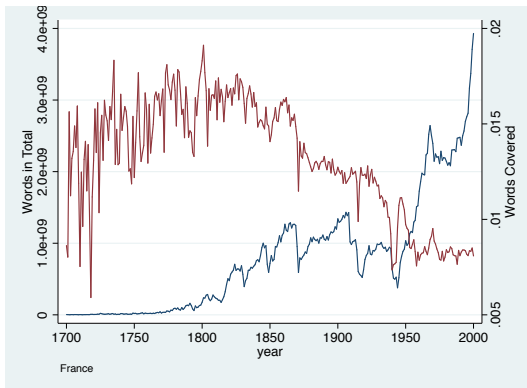




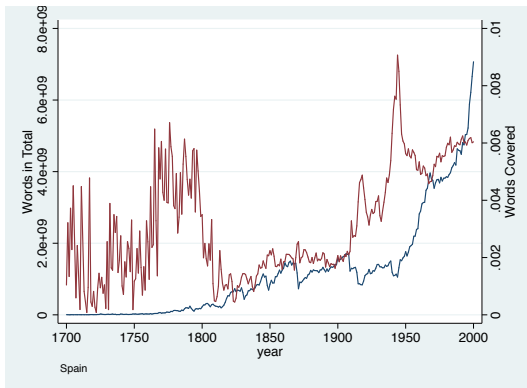
# Words and Words Covered: Germany



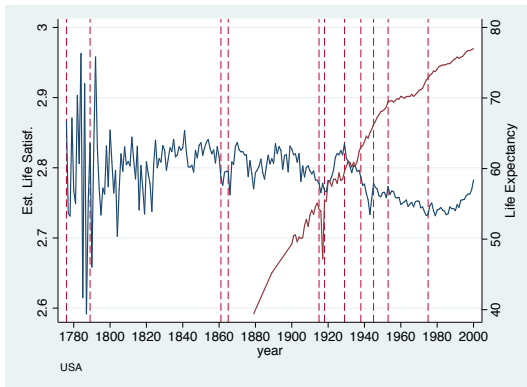
# Words and Words Covered: France



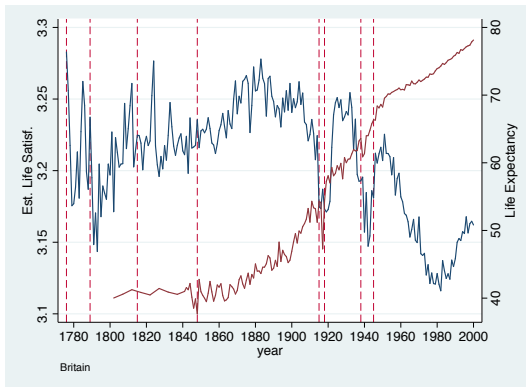
# Words and Words Covered: Spain



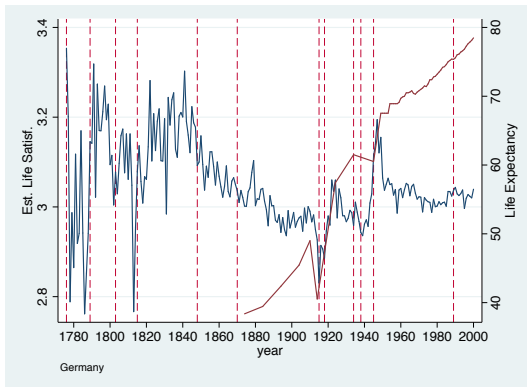
# Evolution of Valence in US English



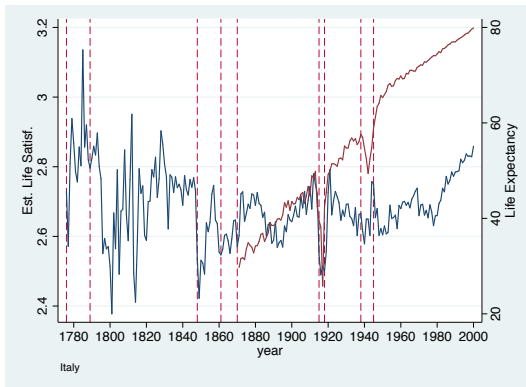
# Evolution of Valence in British English



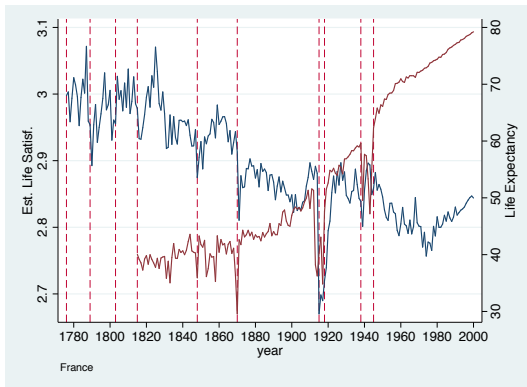
# Evolution of Valence in German



# Evolution of Valence in Italian

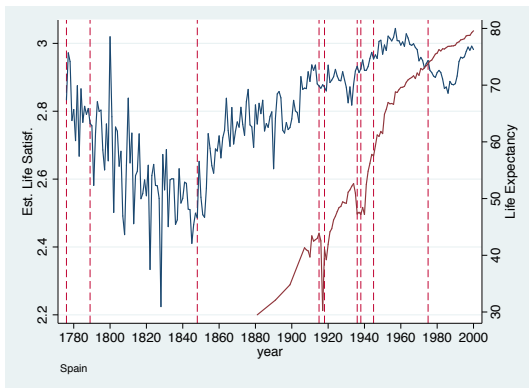


# Evolution of Valence in French

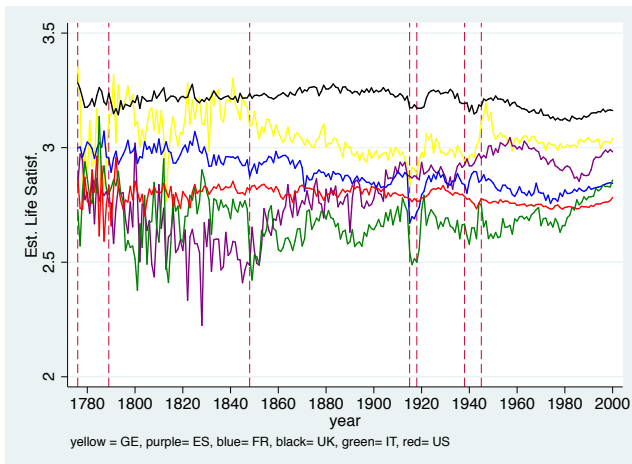




# Evolution of Valence in Spanish



# All Countries



# Summary

- Average Word Valence of a language predict country aggregate Subjective Wellbeing of the corresponding country
- Average Word Valence positively correlate with GDP and Life Expectancy
- an increase of 1% life expectancy is equivalent to more than 5% increase in yearly GDP
- One year of internal conflict costs the equivalent of a 50% drop in GDP per year
- Public debt has a short-run positive effect