Historical Analysis of National Wellbeing Using Millions of Digitized Books

Thomas Hills[a] **Eugenio Proto**[b] **Daniel Sgroi**[c]

[a]Department of Psychology, University of Warwick

[b]Department of Economics, University of Warwick, CAGE, IZA and CESifo

[c]Department of Economics, University of Warwick and CAGE

October 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 (there are many)
  - Time is not included: JP Morgan Chase, BP Deep Horizons oil spill increased US GDP; spending savings increases GDP temporarily
  - Leisure is not included: Ramadan reduces GDP, but increases wellbeing (Campante and Yanagizawa-Drott, 2013)
  - Other issues: exchange rates, goods change, 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
- Inferring public mood (i.e, sentiment) from text (e.g., Nguyen et al,. 2010)
- We use Google Ngrams and word norms for valence 6 languages English (British), English (American), German, Italian, Spanish, French.
- valence norms has words rated on a scale from least to most positive (about 1000 words for most languages), all languages based on same set of words (ANEW)

Language Average Valence Computation

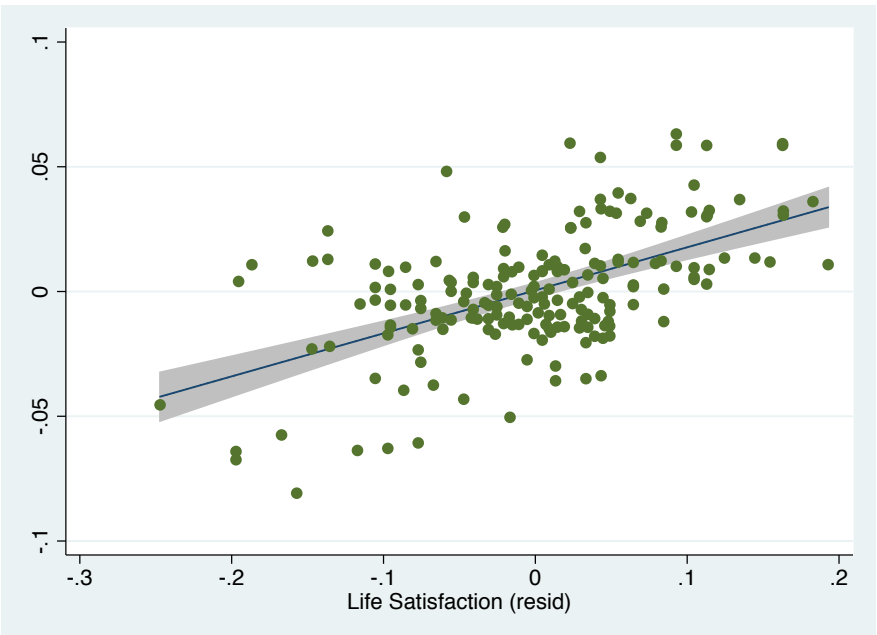
For each language we compute the weighted valence score, $Valence_t$, for each year, t, using the valence, v for each word, j, as follows,

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

where $v_{j,i}$ 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.

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.

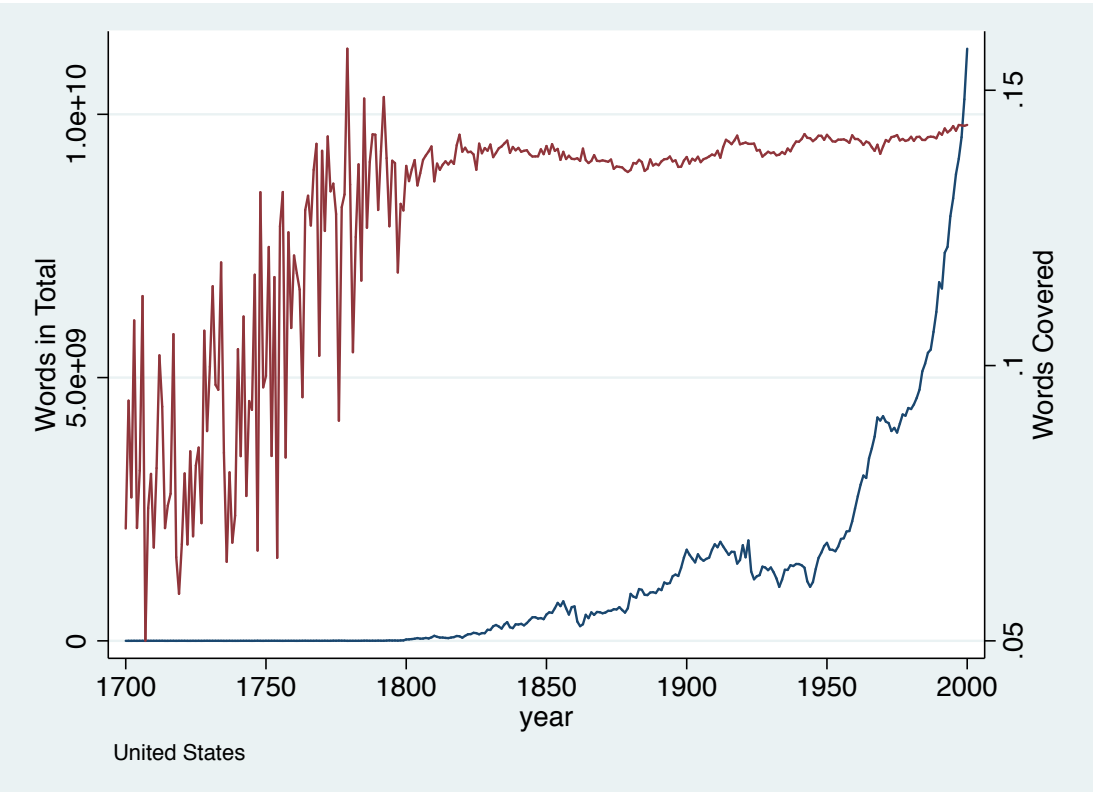


Valence Predicts Aggregate Life Satisfaction

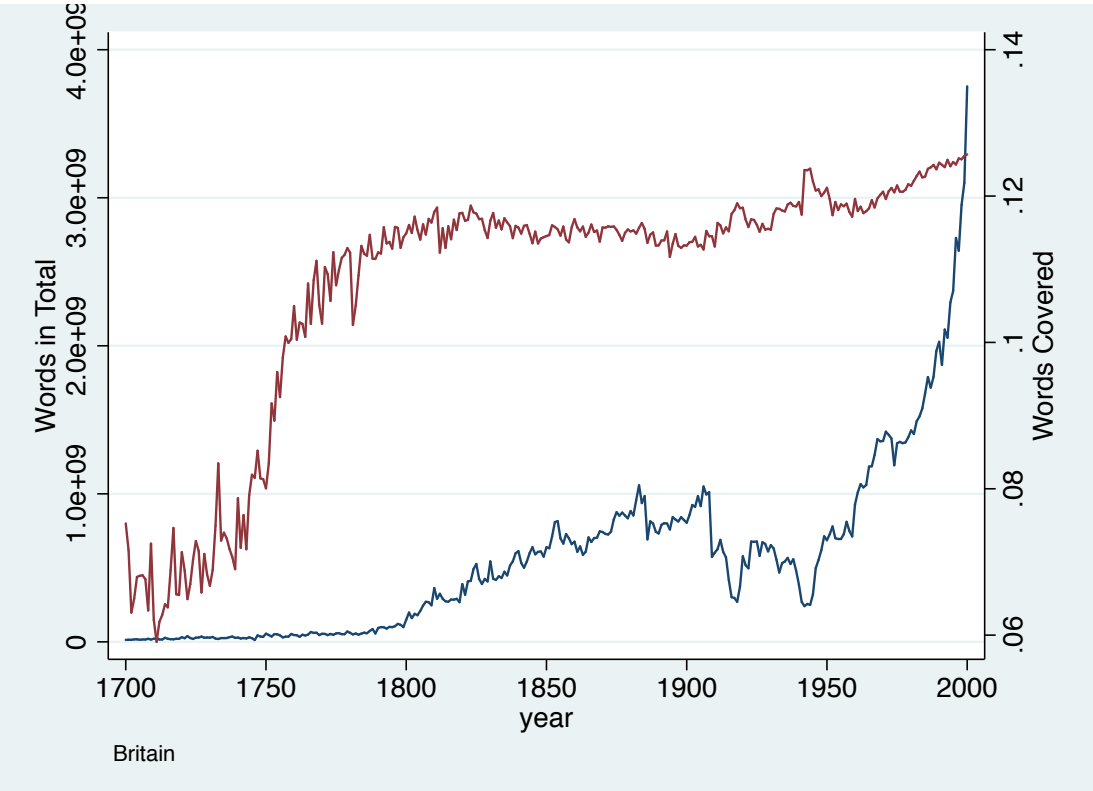
Table: Average life satisfaction per country and year is the dependent variable. Coefficients are in standard deviations.

	1 Year FE b/se	2 with GDP b/se	3 until 2009 b/se	4 W/O Spain and France 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

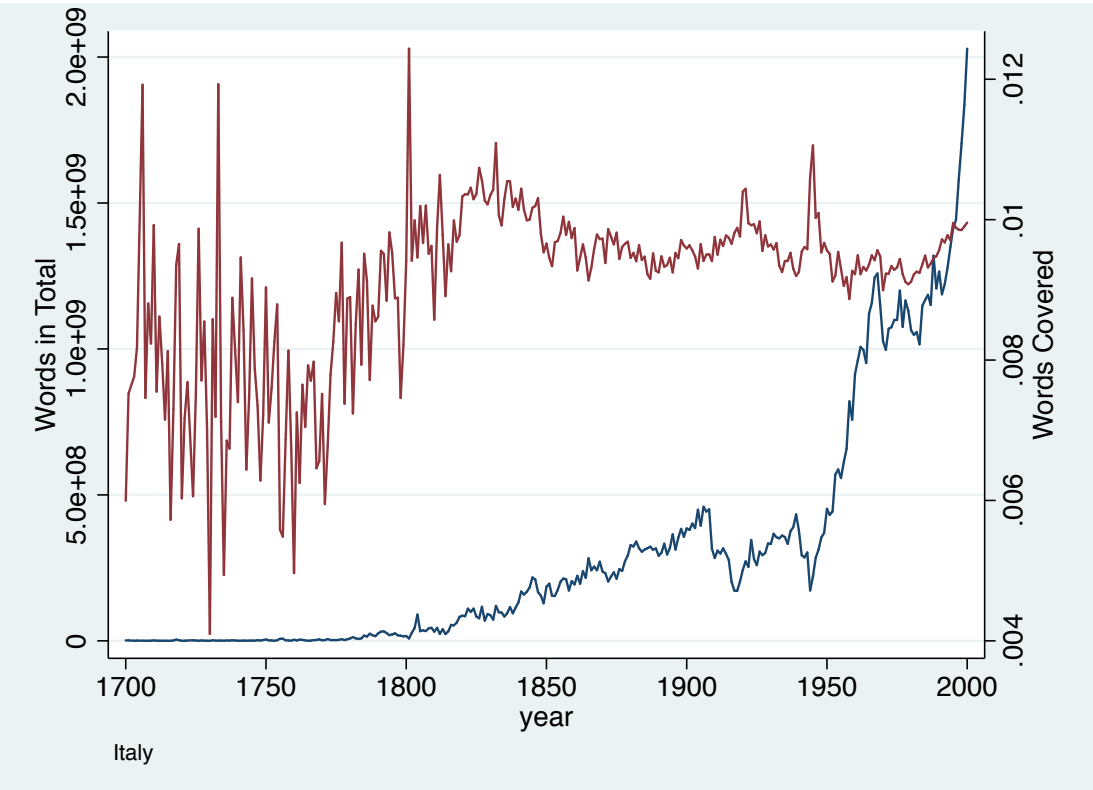
Words and Words Covered.



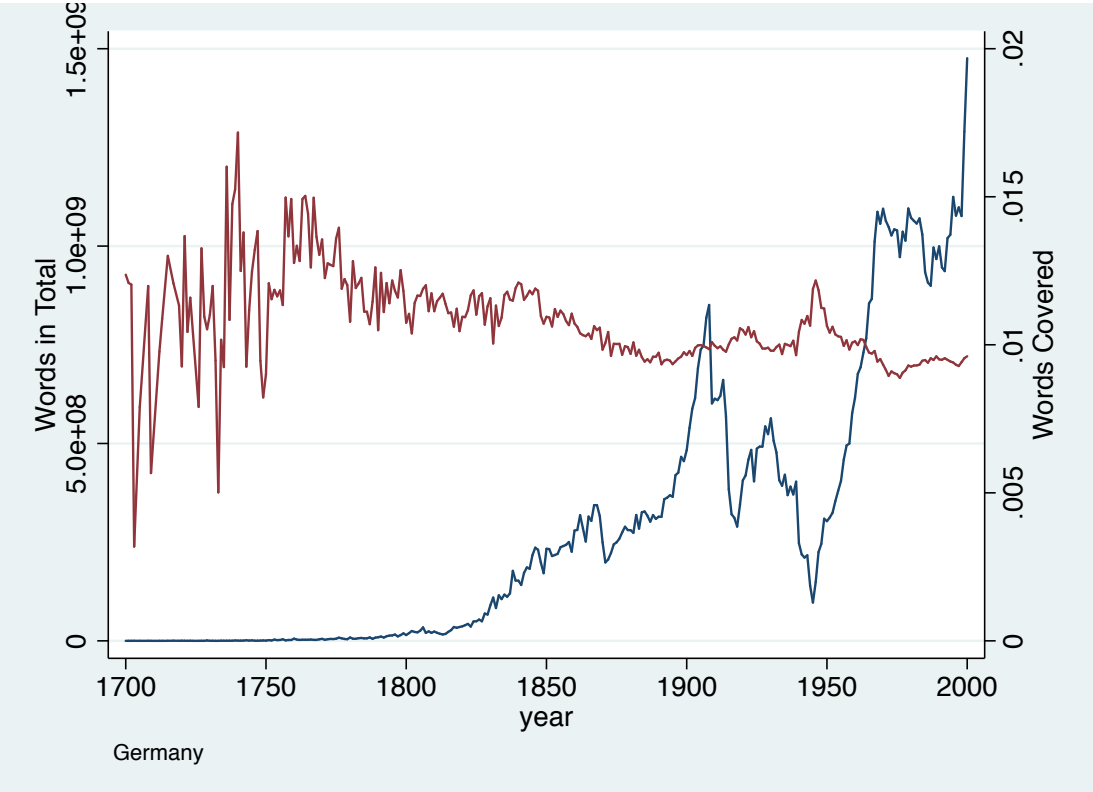
Words and Words Covered.



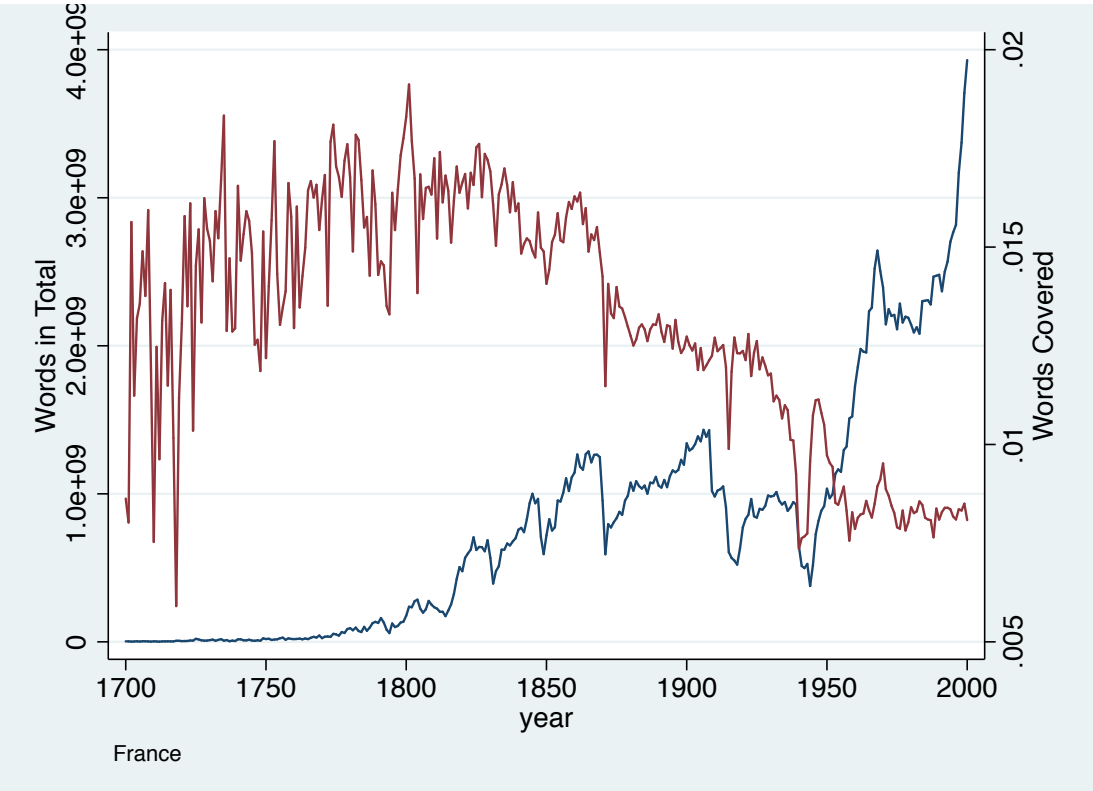
Words and Words Covered.



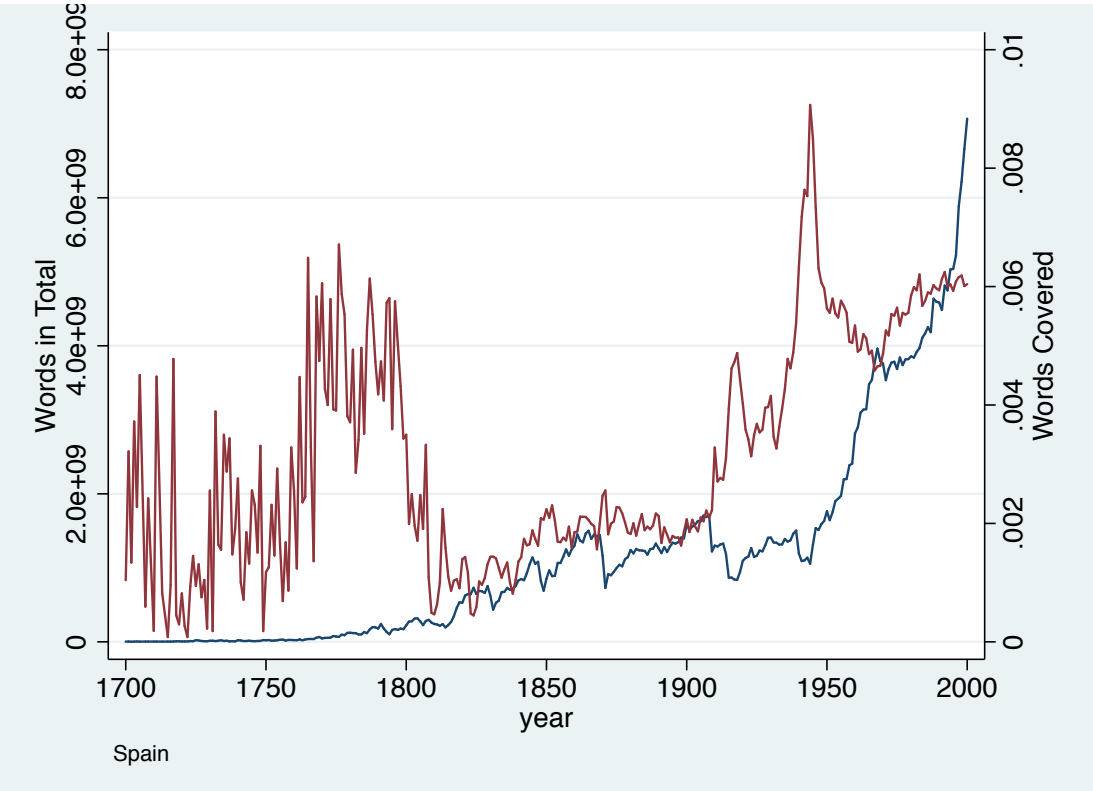
Words and Words Covered.



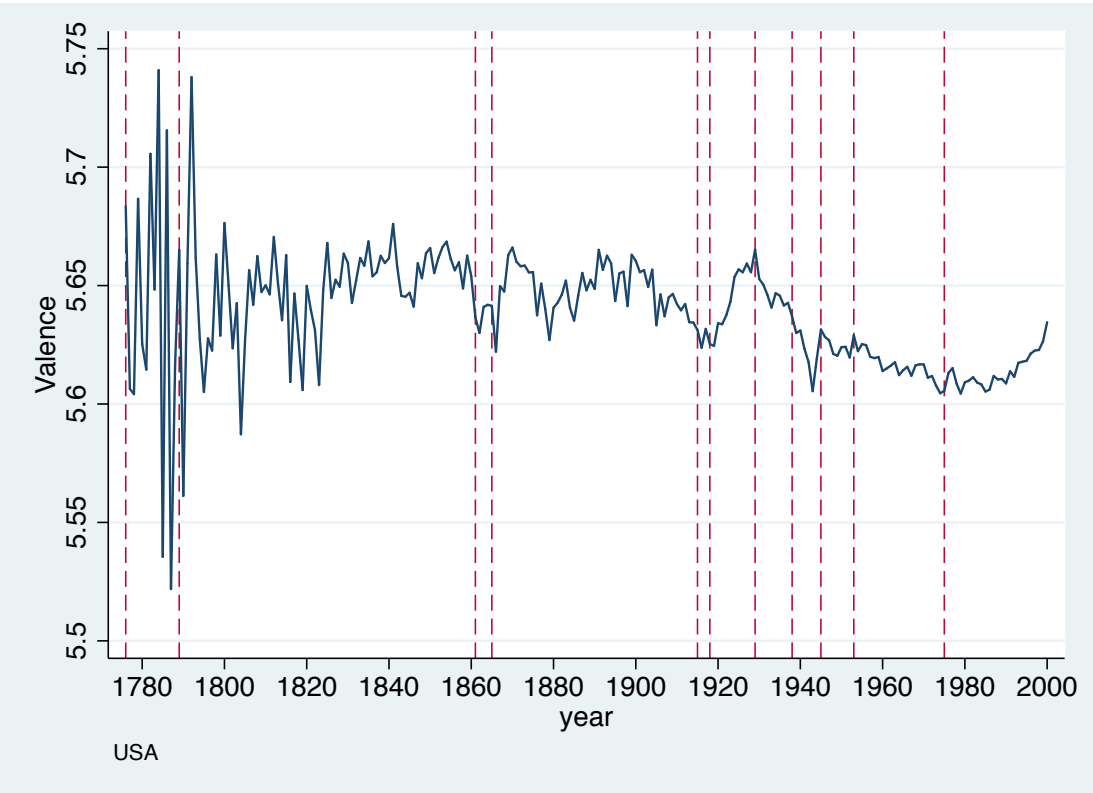
Words and Words Covered.



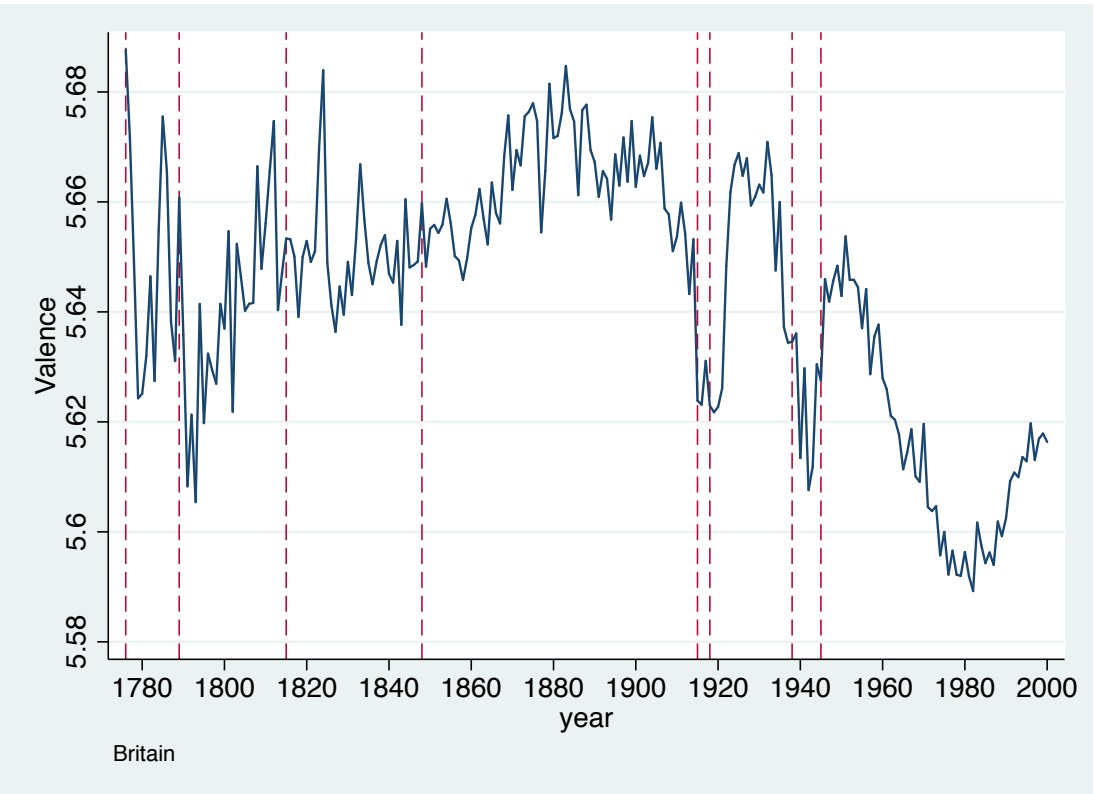
Words and Words Covered.



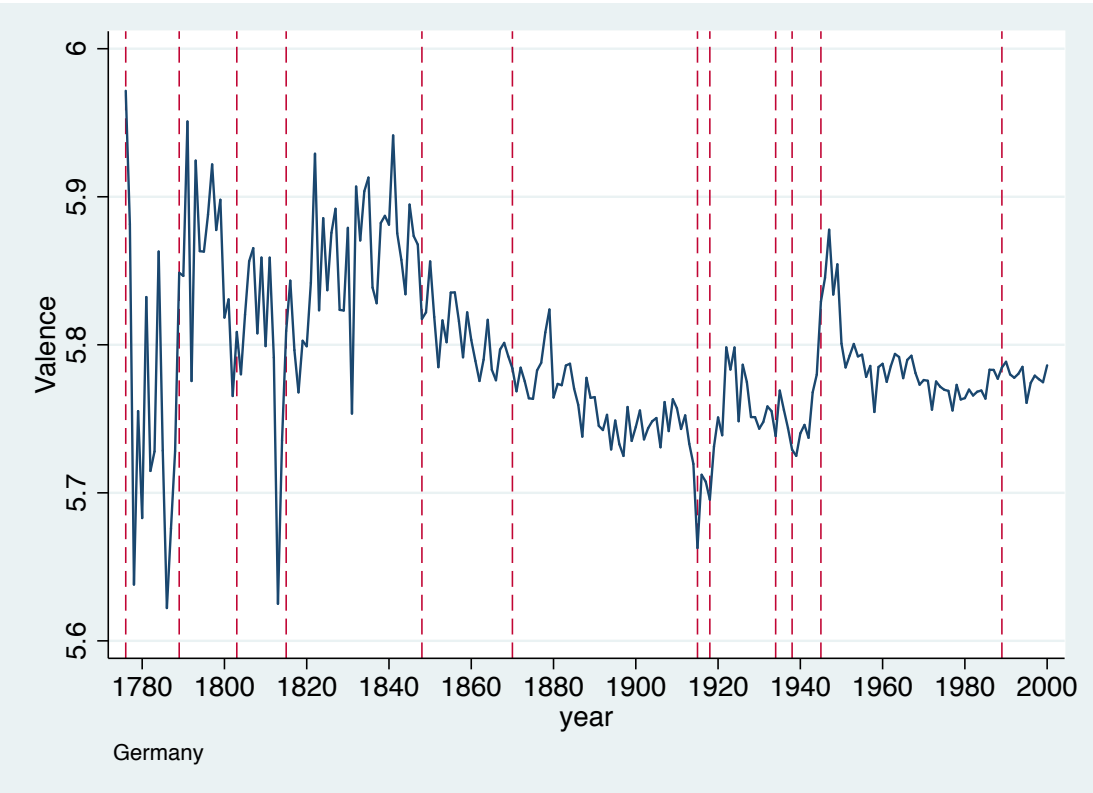
Evolution of the Valence in US English



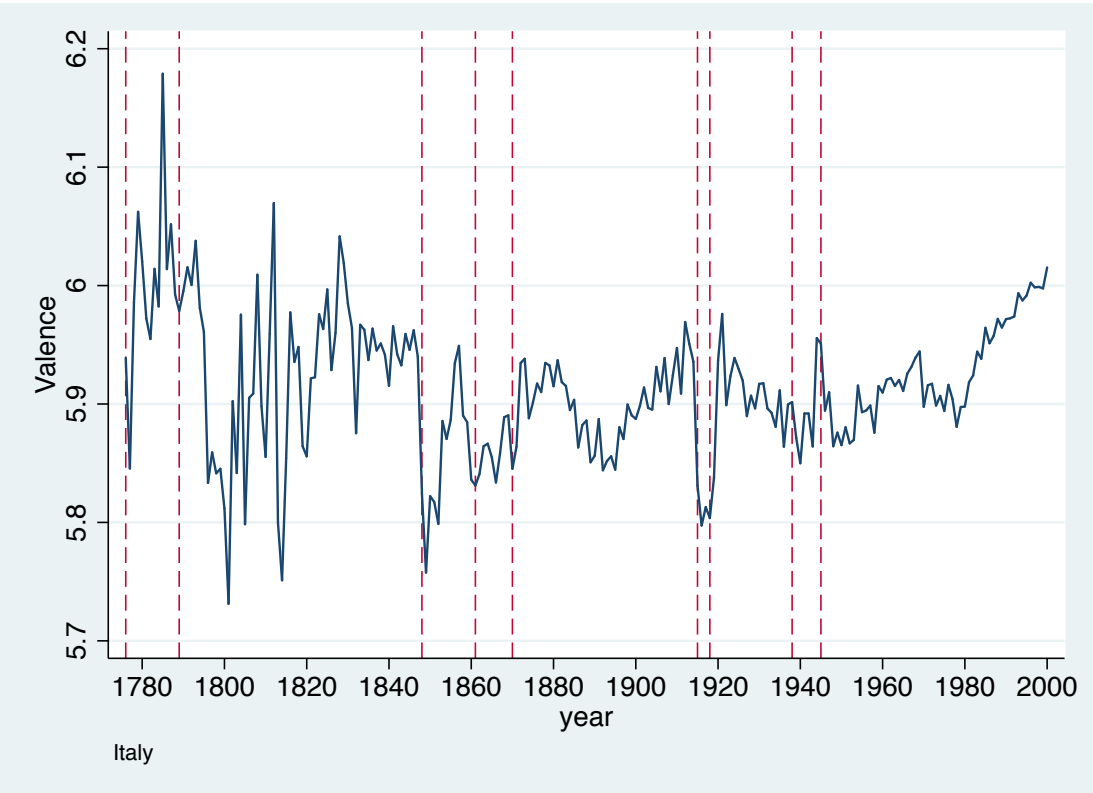
Evolution of the Valence in British English



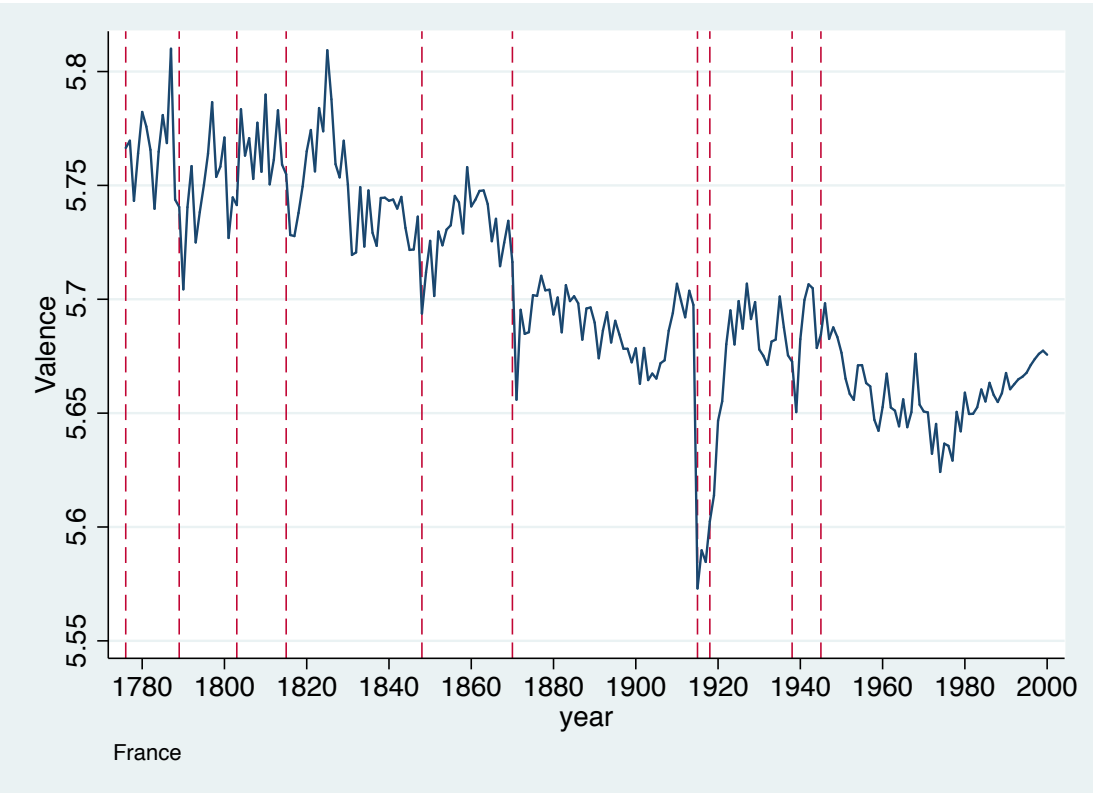
Evolution of the Valence in German



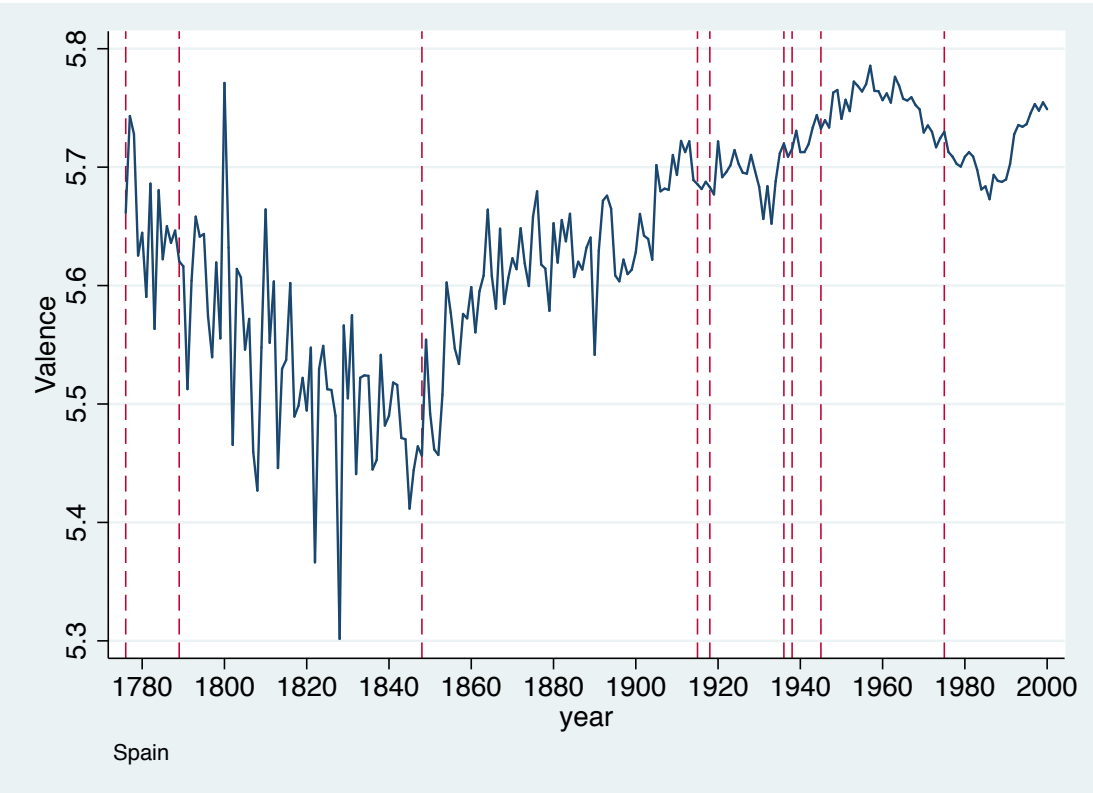
Evolution of the Valence in Italian



Evolution of the Valence in French



Evolution of the Valence in Spanish



Country Specific Factors

- Long-Run Biases (e.g., religion): Control for Country fixed effect
- Books not entirely representative: Education and democracy
- Evolution of literature: year and country-trend effects

Econometric Models

- predict, $\hat{Sat}_{i,t}$, from the simple model:

$$Sat_{i,t} = a_i + bVal_{i,t}, \tag{1}$$

1. Year Fixed Effect:

$$\hat{Sat}_{i,t} = \sum_{z=1}^{Z} \beta_z x_{z,i,t} + \gamma wc_{i,t} + \alpha_i + \eta_t + u_{i,t}; \tag{2}$$

2. Country-Specific Trends:

$$\hat{Sat}_{i,t} = \sum_{z=1}^{Z} \beta_z x_{z,i,t} + \gamma wc_{i,t} + \alpha_i + \delta_i t + u_{i,t}; \tag{3}$$

Data

Table: Main Variables

Variable	Mean	Std. Dev.	Min.	Max.	N
Valence	5.72	0.114	5.302	6.07	1259
Life Satisfaction	2.984	0.175	2.52	3.248	190
per capita GDP (Maddison)	6771.196	6362.951	1007.867	31357	984
per capita GDP (Penn)	25064.164	6553.946	13069.197	43511.594	232
Life Expectancy	59.771	14.774	25.81	82.400	798
External Conflict	0.427	0.495	0	1	1206
Internal Conflict	0.111	0.314	0	1	1206
Democracy	3.983	6.548	-10	10	1079
Govern. Debt (in % of GDP)	68.793	52.352	0.003	261.759	885
Inflation	4.212	18.43	-67.605	344.569	1202
Education Inequality	26.964	19.559	6.111	98.935	784
Words Covered	0.049	0.057	0	0.191	1259

Determinants of the Estimated Subjective Wellbeing

	GDP and Life Expect.		Confl. and Ineq.		W/O Spain and France	
	Year FE	Trends	Year FE	Trends	Year FE	Trends
	b/se	b/se	b/se	b/se	b/se	b/se
Life Expectancy	0.2591***	0.1186***	0.2500***	0.0978**	0.3463***	0.1795***
	(0.0395)	(0.0362)	(0.0410)	(0.0475)	(0.0548)	(0.0644)
GDP	0.0470***	0.0231**	0.0338*	0.0067	0.1199***	0.0759***
	(0.0168)	(0.0101)	(0.0185)	(0.0124)	(0.0247)	(0.0175)
Internal Conflict			–0.0200***	–0.0190***	–0.0185**	–0.0300***
			(0.0066)	(0.0051)	(0.0081)	(0.0039)
External Conflict			–0.0066	–0.0083**	–0.0023	–0.0118**
			(0.0048)	(0.0042)	(0.0066)	(0.0051)
Education Inequality			0.0005	0.0102	0.0180	0.0066
			(0.0055)	(0.0075)	(0.0118)	(0.0088)
Democracy	–0.0137***	–0.0184***	–0.0097	–0.0147***	0.0253***	0.0014
	(0.0052)	(0.0041)	(0.0061)	(0.0049)	(0.0072)	(0.0043)
Trend USA		–0.0017***		–0.0013***		–0.0032***
		(0.0003)		(0.0004)		(0.0006)
Trend Britain		–0.0016***		–0.0014***		–0.0033***
		(0.0003)		(0.0004)		(0.0006)
Trend Germany		–0.0006		0.0001		–0.0016**
		(0.0004)		(0.0005)		(0.0007)
Trend Italy		–0.0003		0.0002		–0.0021***
		(0.0004)		(0.0006)		(0.0008)
Trend France		–0.0013***		–0.0009**		
		(0.0002)		(0.0004)		
Trend Spain		–0.0014***		–0.0009		
		(0.0005)		(0.0007)		
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No	Yes	No
Words Covered	Yes	Yes	Yes	Yes	Yes	Yes
r2	0.973	0.966	0.973	0.969	0.989	0.982
N	692	692	605	605	377	377

The Effect of Public Debt and Inflation on Estimated Subjective Wellbeing

	Debt		with Inflation		W/O Spain and France	
	Year FE	Trends	Year FE	Trends	Year FE	Trends
	b/se	b/se	b/se	b/se	b/se	b/se
Life Expectancy	0.3292***	0.1138**	0.3615***	0.0859	0.3292***	0.1138**
	(0.0501)	(0.0562)	(0.0548)	(0.0538)	(0.0501)	(0.0562)
GDP	0.0396*	0.0258	0.0350*	0.0165	0.0396*	0.0258
	(0.0212)	(0.0160)	(0.0191)	(0.0161)	(0.0212)	(0.0160)
Internal Conflict	−0.0272***	−0.0227***	−0.0277***	−0.0230***	−0.0272***	−0.0227***
	(0.0061)	(0.0049)	(0.0061)	(0.0050)	(0.0061)	(0.0049)
External Conflict	−0.0035	−0.0081*	−0.0083	−0.0086**	−0.0035	−0.0081*
	(0.0051)	(0.0044)	(0.0053)	(0.0043)	(0.0051)	(0.0044)
Education Inequality	−0.0002	0.0087	0.0012	0.0090	−0.0002	0.0087
	(0.0056)	(0.0073)	(0.0059)	(0.0073)	(0.0056)	(0.0073)
Govern. Debt	0.0042	0.0113***	0.0079**	0.0104***	0.0042	0.0113***
	(0.0032)	(0.0028)	(0.0034)	(0.0029)	(0.0032)	(0.0028)
Inflation			0.0625***	−0.0356		
			(0.0176)	(0.0300)		
Democracy	−0.0078	−0.0128**	−0.0106*	−0.0122**	−0.0078	−0.0128**
	(0.0059)	(0.0054)	(0.0057)	(0.0056)	(0.0059)	(0.0054)
Trend USA		−0.0019***		−0.0015***		−0.0019***
		(0.0005)		(0.0005)		(0.0005)
Trend Britain		−0.0018***		−0.0015***		−0.0018***
		(0.0005)		(0.0005)		(0.0005)
Trend Germany		−0.0005		−0.0002		−0.0005
		(0.0006)		(0.0005)		(0.0006)
Trend Italy		−0.0003		0.0001		−0.0003
		(0.0007)		(0.0007)		(0.0007)
Trend France		−0.0012**		−0.0009*		−0.0012**
		(0.0005)		(0.0005)		(0.0005)
Trend Spain		−0.0014		−0.0009		−0.0014
		(0.0009)		(0.0008)		(0.0009)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No	Yes	No
Words Covered	Yes	Yes	Yes	Yes	Yes	Yes
r2	0.978	0.973	0.978	0.973	0.978	0.973
N	565	565	548	548	565	565

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