

Using data from Google Books to measure wellbeing in the past

WARWICK
THE UNIVERSITY OF WARWICK

Dr Daniel Sgroi
University of Warwick

Scope

- “ Joint work with Eugenio Proto and Thomas Hills An attempt to measure wellbeing in the past (long before survey data existed)
- “ Very multidisciplinary: Economics, History, Psychology, Computer Science (“Big Data”), Literature and Language

National Income

- “ National Income Accounting and the birth of macroeconometrics
- “ A similar movement to use national wellbeing data (UN, OECD, national governments)
- “ Measuring GDP in the past: The Maddison Project (from 1AD!) but note the limitations of the super-long run

Measuring Happiness

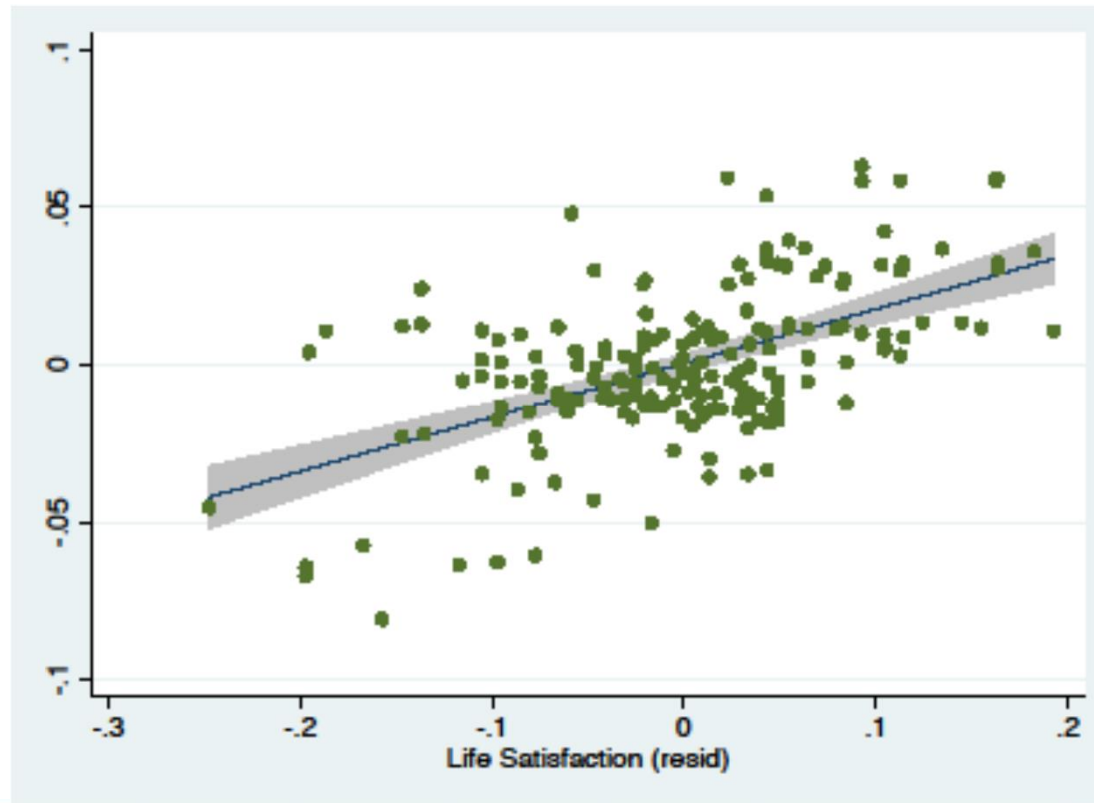
- “ Do we really need historical data? Why?
 - “ Wars, epidemics, depressions, natural disasters..
 - “ Now is fine: we can just ask people!
 - “ But the past, surely impossible? How can we “go back in time” to ask our great grandparents how happy they are?
-

Google Books!

- “ Big Data: Ngrams! ([Viewer](#))
- “ Sentiment (“valence”) and psychology
- “ Building the HPS index

ENGLISH	VALENCE	GERMAN	VALENCE	FRENCH	VALENCE	ITALIAN	VALENCE	SPAIN	VALENCE
aardvark		6.26 Misshandlung	-2.6	abeille	4.22	abbaglio	3.94	abandonado	1.68
abalone		5.3 Unfall	-2.35	abonné	4.53	abbandonato	2	abejas	3.18
abandon		2.84 Erfolg	2.1	abricot	6.55	abbondanza	6.82	aborto	2.8
abandonment		2.63 aktivieren	0.81	absent	3.42	abbraccio	7.7	abrasador	2.46
abbey		5.85 bewundert	1.44	abstrait	4.72	abete	6.17	abrazo	8.13
abdomen		5.43 anbetungswi	0.4	accordéon	5.7	abitante	5.67	abrumado	2.9
abdominal		4.48 Zuneigung	1.9	acide	3.47	abitazione	6.46	absurdo	3.8
abduct		2.42 ängstlich	-1.38	agneau	6.35	abito	7.27	abundancia	6.8
abduction		2.05 aggressiv	-1.6	agréable	8.29	abitudini	4.91	aburrido	2.33
abide		5.52 Qual	-2.7	aide	7.08	aborto	2.06	accidente	1.32
abiding		5.57 Alarm	-2.55	aigle	6.53	abuso	1.74	ácido	3.41
ability		7 allein	-1.8	aiguille	3.9	accettazione	5.79	acogedor	7.64
abject		4 entfernt	-0.6	ail	4.22	accogliente	8.03	acontecimier	5.99
ablaze		5.15 Ehrgeiz	1.5	aile	6.05	accomodanti	6.4	acre	4.23
able		6.64 Krankenwage	-0.79	aisance	7.26	accordo	6.71	activar	6
abnormal		3.53 Engel	2.15	album	6.34	acqua	7.78	acuerdo	7.24
abnormality		3.05 Ärger	-1.9	alcool	5.64	adorabile	7.33	acurrucarse	6.98
abode		5.28 verärgert	-1.27	algèbre	3.87	adulto	5.78	adicto	2.41
abolish		3.84 verängstigt	-1.56	allégorie	5.42	aereo	6.56	adinerado	6.21
abominable		4.05 verängstigt	-1.56	alligator	4.05	affamato	4.74	admirado	7.33
abomination		2.5 ärgern	-1.22	allumette	5.32	affascinare	7.97	adorable	7.48
abort		3.1 gespannt	0.73	ambition	7.6	affaticato	3.73	adulto	5.68
abortion		2.58 Beweglichkeit	1.53	ambulance	3.22	affetto	7.48	afectar	3.48
abracadabra		5.11 Armee	-2.1	âme	7.12	affizione	1.94	afecto	8.1
abrasive		4.26 erregt	1	amer	2.8	affogare	1.79	afianzar	5.93
abreast		4.62 arrogant	-1.93	ami	7.94	aggressione	2.53	afigido	1.96
abrupt		3.28 ermorden	-2.21	amitié	8.38	aggressivo	3.48	afortunado	7.71

Valence correlates heavily with life satisfaction 1972- 2009:



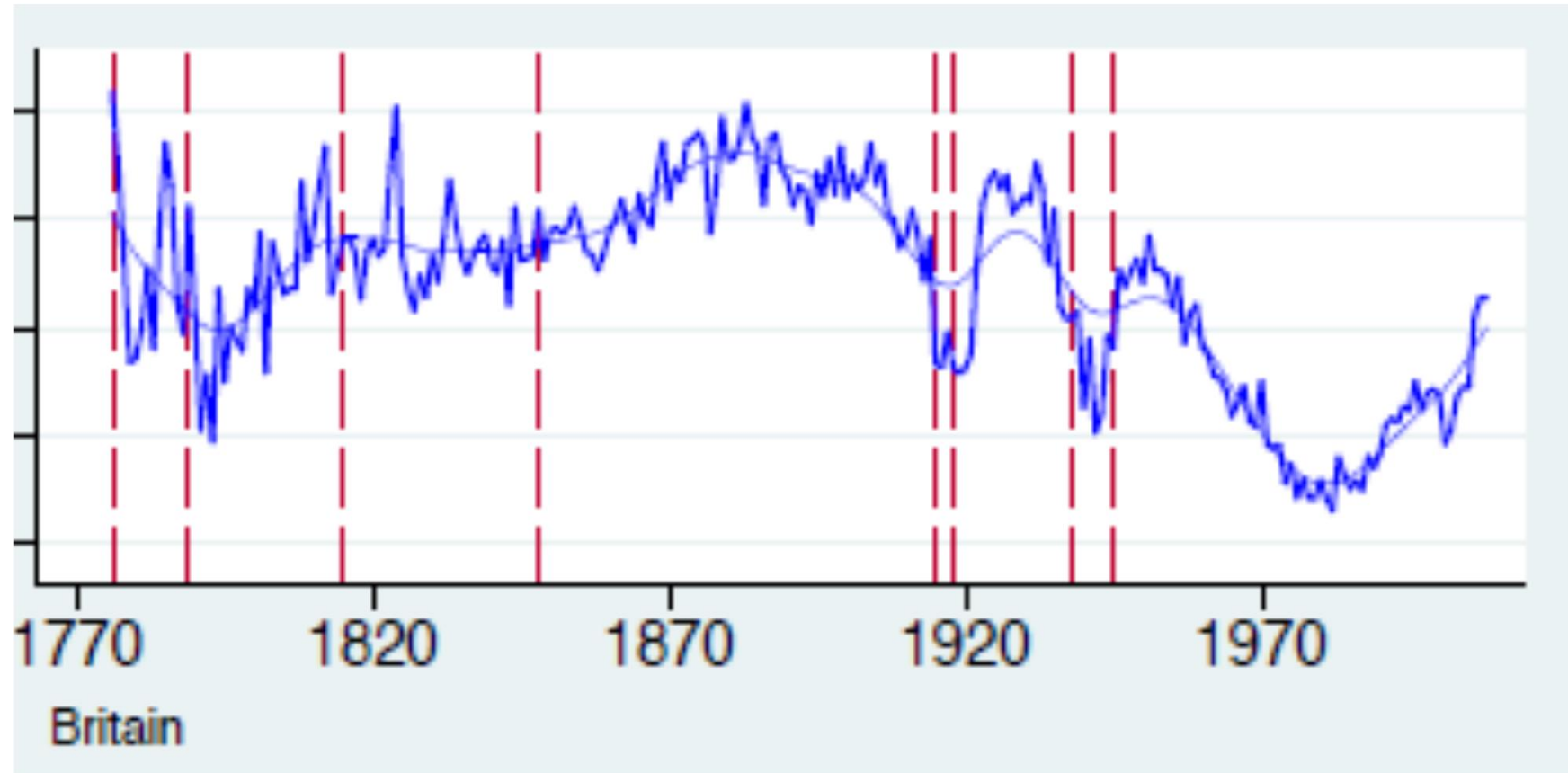
Other checks

- “ Words with high valence occur significantly more often at times of higher life satisfaction (1972-2009, Eurobarometer)
- “ Happy people read happy books (not happy-sad) – competitive publishing industry perhaps?

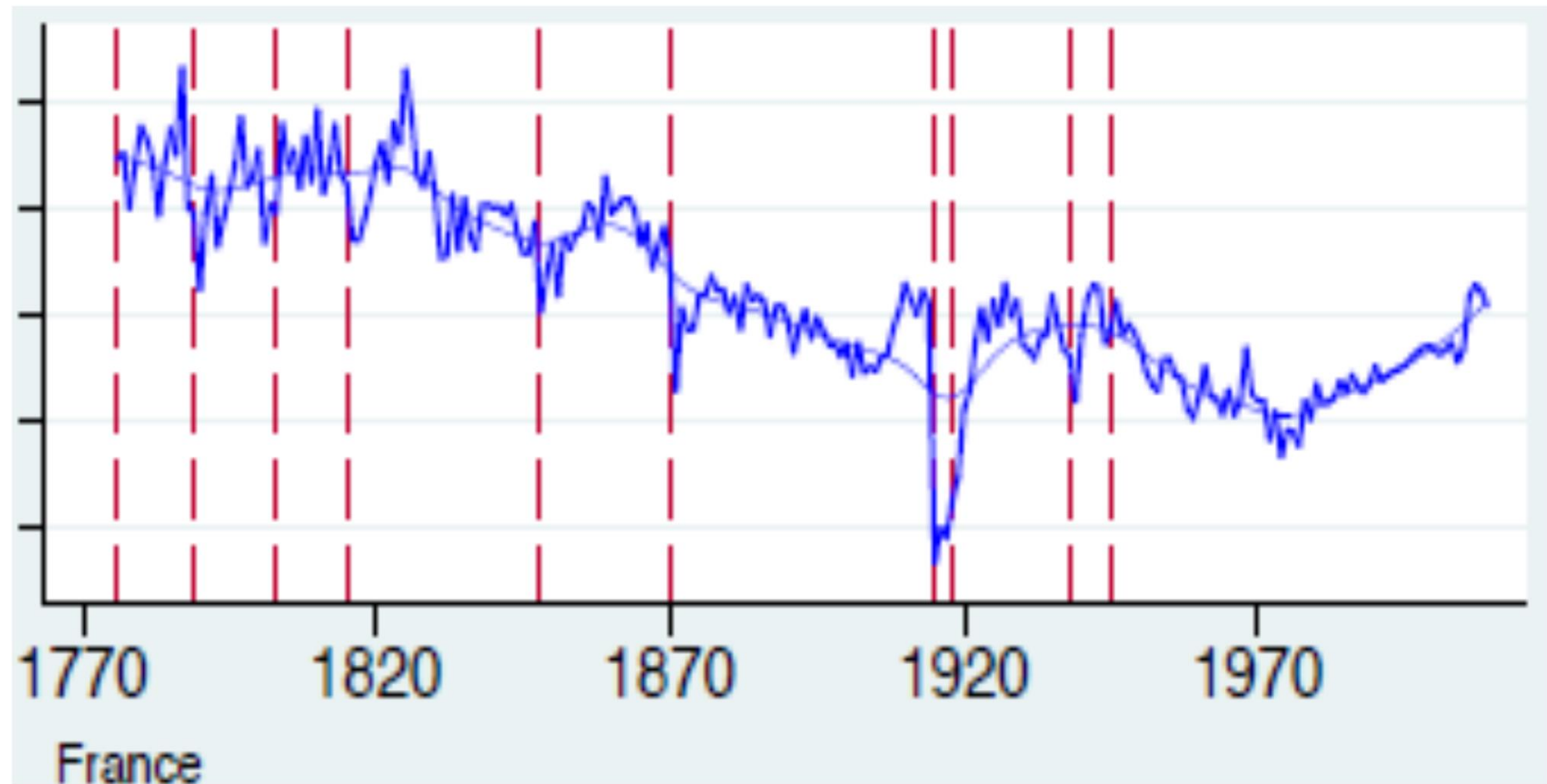
Rolling back through time...

- “ For the UK, USA, Germany, France, Italy, Spain
 - “ From 1776 to 2009
 - “ Lets look at the UK and France for now
 - “ Note censorship/control of publishing in France under occupation
-

The HPS Index: UK



The HPS Index: France



So now what?

- “ Scale, direction, comparisons...
- “ At the margin what matters more for happiness (health vs money, wars vs natural disasters, etc.)
- “ But avoid super-long run comparisons (for now...)

Next steps

- “ Text-analysis on other forms of sentiment, looking at specific words and themes
- “ Refining the method (how language evolves through time)

Thank You!