The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books

Thomas Hills
Eugenio Proto
Daniel Sgroi

CESIFO WORKING PAPER NO. 5906
CATEGORY 13: BEHAVIOURAL ECONOMICS
ORIGINAL VERSION: MAY 2016
THIS VERSION: DECEMBER 2016

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5906*

Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books

Abstract

We present the first attempt to construct a long-run historical measure of subjective wellbeing using language corpora from millions of digitized books for the USA, UK, Germany, France, Italy and Spain. While existing measures go back at most to the 1970s, our measure goes back at least 200 years further. Our measure correlates positively with existing wellbeing measures where available. The relationship with life expectancy is significant and positive. Infant mortality correlates negatively and independently from the effect of life expectancy. There is no correlation with GDP. Econometric analysis of the data is undertaken to control for potentially confounding factors.

JEL-Codes: N300, N400, O100, D600.

Keywords: historical subjective wellbeing, language, big data, Google Books, GDP, conflict, Easterlin paradox.

Thomas Hills
Department of Psychology & CAGE
University of Warwick
Gibbet Hill Road
UK – Coventry CV4 7AL
t.t.hills@warwick.ac.uk

Eugenio Proto
Department of Economics & CAGE
University of Warwick
Gibbet Hill Road
UK – Coventry CV4 7AL
e.proto@warwick.ac.uk

Daniel Sgroi
Department of Economics & CAGE
University of Warwick
Gibbet Hill Road
UK – Coventry CV4 7AL
daniel.sgroi@warwick.ac.uk

December 18, 2016
The authors thank several colleagues for discussions on this and related research, especially Sean Allen, Sascha Becker, Steve Broadberry, Nick Crafts, Ray Duch, James Fenske, Mark Harrison, Christos Makridis, Michael McMahon, Andrew Oswald, Luigi Pascali, Giovanni Ricco, David Ronayne, Jeremy Smith, Thijs Van Rens as well as the many helpful comments from seminar participants at the European Economic Association annual conference (Geneva 2016), the Royal Economic Society annual conference (Brighton 2016), Society for Institutional and Organizational Economics annual conference(Paris 2016) and at the many universities and government institutions where we have presented early versions of this work. We thank CAGE (The Center for Competitive Advantage in the Global Economy) for generous funding and Tomas Engelthaler and Li Ying for expert research assistance.

HISTORICAL ANALYSIS OF NATIONAL SUBJECTIVE WELLBEING USING MILLIONS OF DIGITIZED BOOKS

Thomas Hills, , Eugenio Proto and and Daniel Sgroi

1. INTRODUCTION

Subjective wellbeing (or "happiness") has played a surprisingly minor role in the development and application of economic policy in the past, despite being central to the United States Constitution and countless tracts by moral philosophers and political scientists from Plato, Aristotle, and Confucius onwards. Within economics and the social sciences more generally there has been a move to rectify this, with a growing literature on international patterns of subjective wellbeing.[1] At the same time governments and international organizations have begun to talk in terms of subjective wellbeing as a sensible objective for maximization.[2] This has been supported internationally most notably in 2011 when the UN released a World Happiness Report and the OECD launched the *Better Life Index*. With this backdrop in mind our primary objective is to produce the first (and only) workable proxy for subjective wellbeing going back to 1776. The unit of assessment would be national subjective wellbeing, which would enable direct comparisons with GDP over that period (or at least post-1820 when data is readily available for all the countries in our sample) and would enable us to assess the effect of the improvement in life expectancy, together with the evolution of child mortality, the conflicts and civil wars that have characterized the West in the last two centuries, and the rise of pro-active macroeconomic fiscal and monetary policies.

In many ways this mirrors the development of national income accounting in the 1930s immediately following the Great Depression.[3] Over the years that followed, GDP became a primary objective for maximization. In line with the rise in the importance of such measures there was an understandable need to roll back figures, which led to the Maddison Historical GDP Project and the consequent development of GDP figures going back to 1820. This process of rolling back GDP measurements shows no sign of slowing down, with recent attempts to go back much further still. For instance, Broadberry, Campbell, Klein, Overton, and Van Leeuwen (2012) reconstruct the national income of Britain and Holland going back to 1270 as part of a broader endeavor to understand the impact of industrialization and urbanization.[4]

[1] For some recent examples see Di Tella, MacCulloch, and Oswald (2001), Deaton (2008), Benjamin, Kimball, Heffetz, and Rees-Jones (2012), and Proto and Rustichini (2013).

[2] Several nations including the UK, Australia, China, France and Canada now collect subjective wellbeing data to use alongside GDP in national measurement exercises.

[3] While the Great Depression and the rise of Keynesian Economics gave National Income accounting its greatest push in the 1930s, there have been attempts as far back as the seventeenth century in England and France to keep some measures, with the work of William Petty (1665) as an early example in the English-speaking world.

[4] An important caveat to make is that any historical analysis going back centuries will always be prone to issues of long-run comparability. Consider GDP comparisons for instance. While it is possible to construct GDP based on output several hundred years ago there is a deeper issue of major change across the centuries in what was produced, how to make cross-time comparisons and deflate the estimated measure of GDP. The bundle of goods used to build price indices change slightly year-on-year to reflect this, but across centuries the bundles would be unrecognizable. This of course poses some limitations to GDP comparability, see Jerven (2012). While these issues are important, they have not prevented the development and use of long-run GDP data nor should similar issues in the evolution of language prevent the development and use of long-run wellbeing data. It does mean that it is important to be cautious both about how far back we are willing to go and our ability to perform very long-run comparisons. We will return to this issue below when we discuss

An initial issue is how we can go about extending existing subjective wellbeing measures when direct survey evidence was only initiated in a systematic way in the 1970s for the countries we consider. To address this we make use of the growth of the internet and in particular the digitization of books, which has made available the Google Books corpora, a mass of data (and part of the growing availability of what is now routinely called "Big Data") on what people thought and wrote going back several centuries. While we can potentially go back as far as any available corpus of words (in printed sources, from c.1500 onwards) we elected to start in 1776 for several reasons. First, 1776 is the date of the American Declaration of Independence, one of the most famous of all historical documents to specifically reference happiness. Moreover, many historians would cite the American Revolutionary War (1775-83) and the French Revolution (1789) as key events denoting the start of the modern era. Both immediately followed on from the "Enlightenment," a period notable for a philosophical and political shift towards more practical studies of human experience.[5] Second, and perhaps most importantly, 1776 is consistent with the existing literature on the quantitative use of big data—reflecting a sufficient volume of digitized texts to begin making accurate comparisons (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. 2011, Greenfield 2013).[6] A key issue for us is that the data we generate (to be of greatest interest and utility) must correlate with national subjective wellbeing and not just reflect the mood of authors. We discuss this issue in depth in section 3, but to summarize: we argue that publishers select the books they publish to match the mood of the nation (potential readers) which fits in well with standard assumptions of profit-maximization. To support this we show that our index captures subjective wellbeing very well over the period where we have both our index and existing survey measures (1972-2009). This gives us the support we need to roll back our own measure to well before survey measures exist.

Our methods rely on making inferences about public mood from large corpora of written text. Inferring public mood (i.e., sentiment) from large collections of written text represents a growing scientific endeavor, with widespread implications for predicting economic, political, and cultural trends. Examples include recovering large-scale opinions about polit-

the challenges in more detail and how they can be addressed.

[5] Kant's *Critique of Pure Reason* saw print in 1781.

[6] While the intersections between linguistics and economics and the scope for greater use of available "Big Data" provides a fertile ground for future research, relatively little has been done so far. Chen (2013) is a rare exception that provides an interesting analysis of the role of the future tense in language in fostering future-oriented behavior.

ical candidates (Connor, Balasubramanyan, Routledge, and Smith 2010), predicting stock market trends (Bollen, Mao, and Zeng 2011), understanding diurnal and seasonal mood variation (Golder and Macy 2011), detecting the social spread of collective emotions (Chmiel, Sienkiewicz, Thelwall, Paltoglou, Buckley, Kappas, and Hołyst 2011), and understanding the impact of events with the potential for large-scale societal effects such as celebrity deaths, earthquakes, and economic bailouts (Dodds, Harris, Kloumann, Bliss, and Danforth 2011, Thelwall, Buckley, and Paltoglou 2011). The approach we take here relies on affective word norms to derive sentiment from text (commonly called "valence" within psychology). The efficacy of this approach is established in a study of 17 million blog posts, where Nguyen, Phung, Adams, Tran, and Venkatesh (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 when compared with direct evidence 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 subjective wellbeing using Tweets (Dodds and Danforth 2010, Dodds, Harris, Kloumann, Bliss, and Danforth 2011).

There is more detail on the data-collection in section 2, but here we give a simple idea of how the method works. First a group of individuals are asked to examine a list of words and rate how those words make them feel. This procedure has been carried out for several languages which enables us to focus on the USA, UK, Germany, France, Italy and Spain. Words can have straightforward valences (like "happy" and "good") but can also be quite abstract (like "abreast"). In each case the words are measured and rated allowing us to produce a metric that evaluates the mood of text based on the valence of the words it contains. Using this method we can then work the through hundreds of thousands of books contained within the Google books corpus by year and by country/language. A nice feature of the Google corpus is the ability to disentangle UK English and American English: the UK and USA have distinct publishing traditions that are separated by Google enabling us to consider nationality rather than just language. This is not the case for the other languages, though for German (which is mainly spoken in Germany or regions just outside Germany) and Italian (which is mainly spoken in Italy) this does not appear to be a significant issue. However, the same cannot be said for Spain where there is possible confounding with the use of the Spanish language in Latin America. To a lesser extent French is widely used in North

Africa. This leads us to be cautious when examining Spanish and French data, and we check robustness to their inclusion or exclusion. Added to this is the complication that language is not fixed but evolves over time (Hills and Adelman 2015): just as culture, technology, health and education have changed over time, providing difficulties for historical GDP measures, we face similar issues with language and the market for books.[7]

We check the veracity of our findings in two ways: the first and most straightforward is by analyzing survey-based wellbeing data from the Eurobarometer, which goes back to the 1970s. As we demonstrate below, we find a remarkable degree of similarity between our language-based measure and survey measures over the period when they are both available. This is most apparent through a strong and significant positive correlation between our measure and existing survey data. This is visible in a simple plot but also remains after including controls in a regression analysis. Second we look directly at words themselves. We find that there is a strong correlation between certain recurring words and survey-based measures of wellbeing. We then find that this set of words is highly correlated with valence thereby justifying our use of the concept and measure.

Given the strength of the evidence suggesting that our measure closely reflects survey-based measures of wellbeing, we can proceed to push back further in time. After rolling back our own measure, we check the correlations of our estimated measure of wellbeing against the two welfare indicators for which data are, to our knowledge, available for the longest period of time: life expectancy at birth and per capita GDP.[8] We then add conflicts, democracy, concentration of education and infant mortality (for which long time series data are readily available) to shed light on the mechanisms through which our two explanatory variables, GDP and life expectancy, affect our estimated measure of subjective wellbeing.

This new measure provides us with important insights into economic history, but its reliability is subject to changes in written language and literary styles during the 200 years we are considering. This will bias our measure to the extent that the changes in written language are not associated with subjective well-being. We resolve this issue econometrically, in two

[7]None of which have of course prevented economic historians from carrying out valuable work both across long periods of time and in drawing comparisons with the current experiences of developing nations. See for instance Broadberry, Campbell, Klein, Overton, and Van Leeuwen (2012) and also footnote 1. Also, language does follow certain predictable patterns of change, allowing us to control for certain likely sources of bias (**?**, see)]hills2015recent.

[8]Life expectancy at birth might be considered a proxy for general physical health especially given the absence of any other good long-run measure of health.

ways. The first method is to add year fixed-effects. This method is valid under the hypotheses that systematic changes in language are common across the countries we are considering. The second method is to detrend the data of the econometric models we estimate, in order to control for the possibility of bias generated by deterministic country-specific medium and long-term cycles. This second procedure has the cost of filtering away valid trends in subjective wellbeing. We will also consider the possibility that the presence of stochastic trends in our estimated wellbeing measures could generate a bias, and test for cointegration between valence and life satisfaction.

Historical analysis of our measure shows that life expectancy has a robust and significant negative impact on our measure of subjective wellbeing across all specifications and models. This is robust to the introduction of conflicts, world wars and infant mortality; this last variable correlates negatively with our estimated wellbeing measures and separately from the effect of GDP and life expectancy. This provides some evidence for the existence of a strong relationship between physical health (for which life expectancy at birth is the best long-run proxy available) and happiness. Our index can also be used to consider the "Easterlin paradox" which calls into question any relationship between income and happiness (Easterlin 1974, Easterlin, McVey, Switek, Sawangfa, and Zweig 2010), as opposed to the considerable controversy and work that questions this finding (Stevenson and Wolfers 2008). We provide easily the longest time-series dataset that enables an evaluation of the Easterlin paradox: all previous attempts to evaluate the claim draw on either cross-sectional or much shorter-run time-series. For the period from 1776 to 2009 we find no clear relationship between GDP and our index.

We complete our study by considering uses of the new data that we have gathered, and discussing how our novel methods might be applied elsewhere within economics. There is no reason why our approach should be restricted to developing proxies for wellbeing; data series can similarly be provided for a number of other socio-economic variables of interest which we also discuss in the conclusion.

2. DATA

2.1. *Word Valence*

By analyzing language corpora we aim to gain retrospective insight into the rise and fall of subjective wellbeing as derived from the use of written language in the past. Our key historical data source was the Google Books corpus (Lin, Michel, Aiden, Orwant, Brockman, and Petrov 2012), which is a collection of word frequency data for over 8 million books. Overall, this data represents about 6% of all books ever published. The corpus is based on a database of digitized versions of physically published books from "over 40 university libraries around the world" (?, p.176,)]michel2011quantitative. Further metadata on the exact selection of books is not yet available. However, for the UK and USA, Google has separated the books published in each country, allowing us to focus on the nations rather than the language. For the non-English languages, books are pooled together without country identifiers. As we argued in the introduction this can be problematic especially for Spanish (given the widespread use of the Spanish language in South America) and French (the use of French in North Africa) and for this reason we check the robustness of our results by excluding these two countries.

To assess the valence of individual words we used the largest available sets of word valence norms for each of 6 languages: English (British), English (American), German, Italian, Spanish, and French. In psychology, valence is defined as the intrinsic attractiveness (positive valence) or aversiveness (negative valence) of an event, object, or situation. It can also be used to capture the hedonic tone of feelings and affect, and it is in this spirit that we use the term here. Words with positive valence are taken to have positive connotations for the subjective wellbeing of the user, and those with negative valence are taken to have an equivalent negative connotation.[9] Word valence norms ask participants to rate each word from a list on how positive or negative they perceive the word to be. To allow for comparison across languages, all of our valence norms contain a subset of approximately 1000 words adapted from ANEW, the "Affective Norms for English Words" (Bradley and Lang 1999).[10]

In figure 1 we present a sample of the words covered in all the languages we are considering. For English we used the affective rating norms (Warriner, Kuperman, and Brysbaert 2013). These norms are a database of nearly 14 thousand English words, all rated on a 1 to 9 valence scale. Each word was rated by 20 participants and the mean valence rating was used for the purpose of our study.[11] The 14 thousand words in the database contain a subset of the

[9]The strength of a large dataset is that, even though valence usage may vary over individuals and texts, by averaging over them idiosyncratic noise will be averaged out.

[10]The Google Books corpus also includes books in Russian, Chinese and Hebrew but to the best of our knowledge valence norms have not yet been collected for words in these languages.

[11]20 participants may induce concerns about measurement error, but since we are considering a thousand words per participant with unbiased measurement error, this should cancel out in aggregate. Moreover, the

1034 ANEW words. For German we used the affective norms for German sentiment terms (Schmidtke, Schröder, Jacobs, and Conrad 2014). This is a list of 1003 words, and 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 in our analysis. For Italian we used an adaptation of the ANEW norms (Montefinese, Ambrosini, Fairfield, and Mammarella 2014), which contains 1121 Italian words. As with the English words, the ratings were collected on a 1 to 9 scale. Similarly, the French (Monnier and Syssau 2014) and Spanish (Redondo, Fraga, Padrón, and Comesaña 2007) norms were also adaptations of the ANEW. These contained 1031 and 1034 words respectively. Both used a 1 to 9 points Self Assessment Manikin scale (Lang 1980). All of these norm databases measure multiple psychometric attributes. For the purpose of our study we used only the mean valence rating of words.[12]

For each language i we compute the weighted valence score, $Val_{i,t}$, for each year, t, using the valence, $v_{j,i}$ for each word, j, as follows,

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

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. The proportion is computed over all words in the corpus for that year for which valence norms are available. The Google Book database includes books from 1500 to 2009, but the number of books included for the first three centuries is very small so we would caution against their use at this time (Greenfield 2013, Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. 2011). After 2000 there was a change in the book sampling method (Greenfield 2013). This can be observed in figure 2, where we observe a drop in the number of words used (especially for Spanish, French and Italian). In our analysis below we will test the robustness of our results by the exclusion of the years 2000-2009.

ratings are highly reliable across participants. This explains why this particular dataset has become an important resource within psychology.

[12] Our methods do not take into account negation. For instance if a statement reads "...this gives me no happiness" our methods would pick up the word "happiness" but not realize it has been negated by the word "no". This is a common problem in sentiment analysis. However, this is less of a problem for our analysis so long as negation does not dramatically and suddenly change in a short period of time. If it is used consistently then it will not have any qualitative effect (for instance if some fixed percentage of all words are proceeded by "no", "not" or similar). Even if the use of negation does change, so long as it does so slowly or changes across more than one country it will not significantly change our analysis. This is part of a more general point about how we handle the evolution of language, which we discuss further below.

In figure 2 we plot, for each of the 6 languages we consider, the number of words in total and the percentage of words covered. For US English and British English we can observe that this percentage stabilizes between 10-12% at around 1800. Also for German and Italian, the percentage of covered words stabilizes after 1800, although this percentage is about 1%, which is consistent with the number of words covered in Italian and German being 10 times smaller compared to both US and British English. For French and Spanish, although word coverage is close to around 1% as well, the percentage of words covered does not stabilize, suggesting that word usage distributions may still be evolving in these languages.

2.2. *Explanatory variables used in the historical analysis*

In section 4 we compare our measure of estimated wellbeing with the two welfare indicators for which we have the longest series available data; namely GDP and life expectancy at birth. Following convention, we use per capita GDP for the first analysis containing only observations after 1972 from the Penn dataset (version PWT 8.0) where data are in 2005 international dollars and are adjusted for purchasing power parity. For the historical analysis we use data from the Maddison Project (`http://www.ggdc.net/maddison/maddison-project/home.htm`, 2013 version.) where data are in 1990 international dollars.[13] The other main explanatory variable is the historical data on life expectancy at birth from the OECD, available from 1820 onwards (van Zanden, Baten, Mira dErcole, Rijpma, Smith, and Timmer 2014). Life expectancy from birth is computed by adding up all the current probabilities of death at various ages. It is based on the total of deaths in society matched against the cohorts in which the deaths occur. We can take it literally as a measure of life expectancy, or perhaps more reasonably, we can consider it a proxy of general health which incorporates features such as the state of medical care, diet, hygiene and any major contributor to longevity.[14]

Other variables we will use as controls are internal conflict, external conflict, education inequality (measured as a GINI index, which we use a proxy for the inclusivity of the demand for books within society); the index of democracy (originally, from the Polity IV project) as

[13]The results of the analysis in section 4 changes very little if we use the Maddison dataset instead of the Penn dataset.

[14]It is hard to believe that concern about life expectancy at birth directly drives happiness, but as a general measure of health it is easy to see how it might be important. Another possible mechanism might be as an indicator of demographic change since higher life expectancy at birth is indicative of an ageing population.

an index of freedom, from the OECD data available from 1820 onwards (van Zanden, Baten, Mira dErcole, Rijpma, Smith, and Timmer 2014), and infant mortality from International Historical Statistics 1750-2010 (Mitchell 2013) to better investigate the channel from life expectancy to wellbeing. Table I summarizes the data.

3. UNDERSTANDING THE INDEX: IMPORTANT CONSIDERATIONS

Before proceeding to examine the index derived from our valence-based measure there are a number of important considerations that we need to address that relate to the market for books, and forming a better understanding of what our measure is actually measuring. We will tackle these in turn in this section.

There are other important issues relating to the evolution of language and literature, potential cointegration and detrending but we will cover these in section 4.3 when we show how we can deal with many of these issues econometrically.

3.1. *Comparison with Subjective Wellbeing Indices: Are We Really Measuring National Mood?*

Is the valence-based index a measure of national mood or just author sentiment? At first glance it might look as though by performing sentiment analysis on published books we are capturing the mood of the authors of those books, in which case why should this have anything to do with the mood of an entire population? Are we perhaps arguing that authors are somehow typical? It might well be that the authors of the books in question are typical but that is not the argument we are making. Rather we argue that the index we have produced is a good match for the mood of a nation because books reflect, both in production and consumption, the kinds of things people are thinking about. We can demonstrate this both theoretically and empirically. Theoretical considerations focus on the publishing industry, how it works and in what sense published books might be linked to national sentiment. We develop this argument further in the next section.

However, before considering the book industry, we provide two empirical observations. First, we perform a direct empirical check on the validity of our valence-based measure by comparing it with existing survey-based measures of subjective wellbeing. The measure of life satisfaction we take as the ground truth is the average per year and per country data taken from the Eurobarometer. The question surveyed individuals answered was “On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?", coded in a 4 point scale from "Very satisfied" to "Not at all satisfied". This is the oldest survey available containing most of the countries of origin for the languages in the digitized Google Books we used. The first wave dates back to 1973 and it covers every year. In particular, it contains data from the UK (104,068 interviews), Germany (102,795 interviews, and we consider only West Germany before 1990), Italy (103,789 interviews), France (102,692 interviews) and Spain (75,259 interviews, with data available only from 1985). Our measure correlates positively with the Eurobarometer measure of life satisfaction, as indicated in figure 4 and table II below but we will describe our findings in detail in the Analysis section below. Finally we should note that data from the USA is not included in the Eurobarometer, but it is available from the early 1970s in the General Social Survey. However, the measure of subjective wellbeing is "Happiness" rather than "Life Satisfaction" (used in the Eurobarometer), hence it is not directly comparable. We nevertheless ran a similar analysis by including this measure from the GSS as well, obtaining very similar (positively correlated) results which are available upon request. We explore these results in much greater detail in section 4 below, indicating that it remains robust to the inclusion of other controls.

Second, one can investigate the historical trends of our measure against known peaks and valleys of public mood. As will become clear in our subsequent analysis, world wars, the roaring twenties, the great depression, and the winter of our discontent, along with other ups and downs are all clearly reflected in our measure.

This empirical arguments stand irrespective of the underlying reason why the index is a good match for national sentiment, and applies whether or not the index also happens to capture the mood of authors or not. Nevertheless, it would be good to have a clearer idea why this is the case and so let us start with a description of the publishing industry which offers a plausible rationale.

3.2. *The Publishing Industry*

What do publishers actually do? We can think of publishers as fulfilling two roles. First and foremost, they publish books: attending to the physical production of books, which for the period in question almost entirely concerns the manufacture and distribution of print texts: this alone means that they cannot afford to publish every book they receive. This leads into their other role historically which is to filter from the mass of written books those they feel happy to publish. In this way they fulfil a role as a middleman, similar to a bank that brings lenders and borrowers together, or a stock market bringing investors and firms together. Publishers take the huge supply of (largely) unsolicited written books and select from them books they feel will match the demand of the reading public.[15] The end result is that only a very tiny minority of authors end up with a publishing contract: some estimates suggest that publishers (and more recently, agents) can receive hundreds or thousands of unsolicited manuscripts a year and might select only a handful,[16] On that basis the text of published books represents a tiny proportion of the words written by all (published and unpublished) authors. Similarly they also decide whether a book published in a foreign country is a good match for a different country, for instance when publishing an American book in the UK or translating a book from Italian to another language. The insight here is that publishers (and agents) filter in a non-random way: they filter in an attempt to fulfil their role in matching the books that are being published to the demand from potential readers. The size of what is often called "the slush pile" of unsolicited manuscripts compared with the tiny number that are selected means that publishers can in effect choose any book they wish, or from our perspective, any set of words which reflect a particular mood. To be more precise, we are arguing that publishers profit-maximize in a completely standard way by selecting a book (a set of words) with an implied mood that they feel are demanded by the population in general in order to maximize sales.

How do publishers (agents) select books? Unless we have reason to suspect some behavioral forces or some market failure we would normally assume that firms aim to profit maximize. There is no special reason to think otherwise in the case of publishing firms: we would therefore argue that they select from the mass of over-supply, books that they feel with earn the most in sales. That means matching supply with demand and carefully ensuring that the books they select are going to be popular with the paying public. This introduces a complication: will publishers sell more if they match books that are characterised by positive mood to a population that also is characterized by a positive mood, or rather should they

[15] Recently this role has been partly spun out of publishing and is now carried out at least in part by "agents" who receive unsolicited manuscripts and select from those they wish to bring to the attention of publishers. After a process of editing and rewriting, these selected books will be seen by publishers who in turn might perform another layer of filtering before deciding to print and distribute the books in question.

[16] Alberge (2015) gives two specific examples of publishers acceptances from unsolicited fiction submissions: 3/5,000 at Jonathan Cape, and 1/400 at Harper Collins.

match happier more upbeat books to sadder more downbeat readers? Do they even care about the sentiment in books? These questions do not have obvious answers: while we might think that happy people want to read happy books, might sad people not also want to be"cheered up" and also read happy books, or do readers even care about the implicit mood of the books they read? Indeed, is there ever a demand for books characterised by negative word use? As this is an open question theoretically, it becomes an empirical matter which we address by once again looking at the period from 1972-2009 in which we have both our text-based index and the Eurobarometer survey measure of happiness. From the discussion above we know that there is a positive correlation between our index and life satisfaction so we know the answer to both questions: the mood implicit within published texts is a good proxy for life satisfaction, and since the correlation is positive, publishers match books typified by predominantly high valence words ("happy books") to "happy people" and books typified by predominantly low valence words ("sad books") to "sad people".

Similar to banks and stock markets, publishers and agents provide a match-making service. Their role incidentally ensures that the mood characterised by published books matches the mood of the population in general.[17]

3.3. *Finding the Right Lag*

A further concern is the year of publication, in particular we need to consider what matters: when a book is published or when it was written? The above arguments suggest that if a publisher is concerned with sales maximisation then the key issue is to match the words in a book to the demand right now (or as soon as is practically possible). It would be possible for a book that is considered a poor match for demand in one year to be a good match a few years later. That would at least partly explain how books might be rejected multiple times before being selected and then becoming great successes. What matters then is the year of publication. In fact, since it can take a year (or longer) for a book that is accepted for publication to see print it might be that some short time lag is also important. A related concern is the fact that a book can be considered a durable good, since it can be read

[17] An alternative explanation which would also explain the correlation between our index and survey measures would be that authors happen to match the population average. This seems unlikely if authors themselves are not typical members of the population (they might be more educated on average, wealthier, or quite the opposite) but it could be true. However, if this is true this would be every bit as supportive of our use of the index as a measure of national wellbeing and so this is not an argument against the use of our index as a measure of national-level subjective wellbeing.

more than once and potentially read years after purchase (or second-hand many years after publication) which will add further noise to any measure. This all reinforces the need to check empirically which lag makes best sense. The analysis in the appendix indicates that a one-year lag from publication provides the very best correlation between life satisfaction taken from the Eurobarometer (from the 1970s onwards) and our index.

3.4. *How Far Back Should We Go?*

A more general issue is whether 200 years is too far back to go. We know from literary history that the early nineteenth century represents an important period in the history of books: the arrival of social commentary, the rise of the novel as a form of entertainment (which existed of course in earlier ages, but became accessible to the masses), and increasing standardisation of language. Despite this, language has changed in the past 200 years. This means that comparisons over very long periods will always suffer from difficulties. For this reason, for instance comparing the flat level of our measure in 1800 to 2000 seems difficult. This is of course also true of other major time-series (comparisons of GDP for instance) and so we would recommend prioritising local comparisons. So, while language has changed between 1800 and 2000, between 1800 and 1850 the changes will have been less pronounced, similarly between 1950 and 2000. Second, by providing a measure from 1776 onwards we are of course allowing its use from any date after 1776: so those who have reason to doubt the voracity of the index before any given date for a particular country can elect to use it only for the more recent past. Starting the index in 1800 or 1850 or even 1900 would still result in many decades more subjective wellbeing data than is otherwise available.

4. ANALYSIS

We next move on to our econometric analyses. We followed three different approaches. The first aims to show that the valence for each language and year is a significant predictor of average life satisfaction for the country of origin for the language. In the second, which can be considered a robustness check for the first, we ran an analysis based on correlations between the frequency of each word for which we know the valence, and the average life satisfaction in the corresponding language and year. In the third we calculate the estimated subjective wellbeing using the predictions from the first analysis and analyze how this measure correlates with indicators of wellbeing for which historical times-series data are available.

4.1. *Valence and Aggregate Life Satisfaction*

In figure 4 we present the relationship between the time-series data of the valence of each language and year, and the aggregate life satisfaction (derived from Eurobarometer data) of the country of origin of the language in the corresponding year for the period over which survey data is available on life satisfaction. As discussed in section 3 we observe a positive correlation for all languages. In the bottom-right panel both series of data are presented in the form of residuals after controlling for country fixed-effects. The relationship is clearly positive and highly significant.

However, the correlations presented in figure 4 could be due to omitted variables like GDP, deterministic or stochastic trends, or other external influences. The aim of the analysis we present in table II is to show that the positive relationship in the bottom right panel is robust to the introduction of the most plausible omitted variables.

In column 1 of table II we note a positive and highly significant relationship between valence and life satisfaction. The coefficient suggests that an increase of one standard deviation of valence corresponds to an increase of 1.5 standard deviations in life satisfaction. This relation holds even after excluding Spain and France (column 3), the two languages that feature clear changes in word coverage. In this case the relationship becomes stronger as one would expect given our previous point that the link between published language and national subjective wellbeing may be weaker for French and Spanish due to the heterogeneity of countries publishing in these languages. We also introduce controls for country-specific trends (column 4), GDP (column 5), and year fixed-effects (column 6). Furthermore, we note that in 2000 Google changed the way they measure word frequency and so we also checked our results for robustness to this change by excluding the data from 2000-2009 (column 2).[18] In all these specifications the coefficient of valence is always positive and highly significant.

In columns 4 and 5 we introduced a control for deterministic trends. However, stochastic trends may also bias our results. To address this issue we used the Augmented Dickey-Fuller unit-root test for stationarity of valence from 1970 onwards for all countries separately. The test for a unit root can be rejected in all but Italy (MacKinnon approximate $p-value$ for $Z(t) = 0.6898$) and Spain (MacKinnon approximate $p-value$ for $Z(t) = 0.7101$), whose series

[18]The robustness of our results to exclusion or inclusion of the data from 2000-2009 leads us to include those extra years whenever possible in the analysis to follow. We perform further robustness checks in section 4.3.

resulted integrated of order 1. For the UK, the unit root can be rejected at 10% confidence levels (MacKinnon approximate $p-value$ for $Z(t) = 0.0696$). For these 3 countries we performed the same test on the life satisfaction variable. For life satisfaction in the UK the test for a unit root can be strongly rejected (MacKinnon approximate $p-value$ for $Z(t) = 0.0000$). This implies that for the UK a stochastic trend cannot be a source of confounding in the relationship between valence and life satisfaction.

For life satisfaction in Italy and Spain the unit root test cannot be rejected (Italy: MacKinnon approximate $p-value$ for $Z(t) = 0.2743$; Spain: MacKinnon approximate $p-value$ for $Z(t) = 0.2564$), but can be rejected on the first differences; the two series are then integrated of degree 1. Accordingly, there are stochastic trends in both life satisfaction and valence for Spain and Italy.

We therefore tested cointegration between valence and life satisfaction in both countries. In Spain the test for cointegration can be rejected: we can reject the unit root test on the residuals of the regression of valence on life satisfaction for Spanish data (MacKinnon approximate $p-value$ for $Z(t) = 0.2564$). Hence for Spain a stochastic trend can potentially be an omitted variable biasing some results in II. However, it is highly unlikely that this is the case, since as we saw in column 3 of table II the positive and significant correlation between valence and life satisfaction would be even stronger if we exclude Spain from the regression. For Italy the test for cointegration between valence and life satisfaction cannot be rejected: in the residuals of the regression of valence on life satisfaction in Italy the test allows us to reject the existence of a unit root (MacKinnon approximate $p-value$ for $Z(t) = 0.0011$).[19] The existence of cointegration between two variables provides a further test of the existence of a link between these variables, establishing a correlation between long-term shocks in both variables. Hence a permanent shock in life satisfaction is featured in the valence as well.

Finally, we note that the positive correlation between life satisfaction and valence is non-trivial. It might have been the case that mood derived from published books negatively reflected popular mood, for instance if “sad” people desire “happy” books. However, our results indicate that the mood in books reflects the mood of the population at large even including a wealth of relevant controls.

[19]The details of all tests can be provided upon request.

4.2. *Correlations between Words and Average Life satisfaction*

In this section, we conduct a non-parametric analysis that complements the conventional regression analysis described above. Specifically, this analysis asks which words best correlate with changes in survey-based measures of wellbing. To do this, we first calculated the relative frequency of all words for which there is a valence measure for every year. The relative frequency is simply the number of times the word appears in each year t and country j in the Google book corpus data, divided by the average frequency of every word in the same language j and year t; then we select the words for which the level of correlation with the survey-based measures is significant at the usual threshold of the 0.05% level. From this set of words, we compute the averages of the valence across the words correlating positively and negatively.

If the valences of the words carry information about life satisfaction then the average valence of all words that correlate positively with life satisfaction should be significantly higher than the average valence of the words that correlate negatively. This is exactly what the bars of figure 3 suggest. Words that correlate positively (negatively) with life satisfaction also correlate positively (negatively) with valence. This indicates that valence is aligned with reported life satisfaction over the period for which both are available.

4.3. *Welfare and Estimated Subjective Wellbeing*

Henceforth we will link the language to its country of origin, and following a mild abuse of notation, the i index used earlier to indicate a language will be used hereafter to also denote the corresponding countries: the USA, Britain, Germany, Italy, France and Spain. In figure 5 we show the estimated subjective wellbeing of the 6 countries, defined as $\hat{Sat}_{i,t}$. This is calculated as the prediction deriving from a simple OLS regression of average life satisfaction for the country and years these measures are available on corresponding valence and country dummy variables.

Starting from 1776 we considered the UK, US, Germany, Italy, Spain and France until the year 2009. While the red vertical lines represents key political events in the country of origin of each language, for all countries we draw lines for 1789, the year of the French Revolution, World War I (1914-18) and World War II (1939-45), in the 5 European countries we also added 1848, a year typified by a number of revolutions.[20] For US English we observe a sharp

[20]Moreover, in the US, the vertical lines represent: the Civil War (1861-65), the Wall Street Crash (1929),

drop during the civil war and other drops corresponding to World War I and World War II. Interestingly the data show a peak in 1929, the year of the Wall Street crash, supporting the view that the crash followed a period of over-optimism. In all countries apart from Spain, we also observe a drop during World War I and World War II. Spain was not directly involved in these wars (of course, the Spanish data may also be influenced by the buoyant development of South American literature). In Italy, France and Germany the effect of World War I seems stronger than for World War II, reflecting perhaps the strong control of the press which these countries experienced during World War II. We will come back below to the issue of the freedom of the press when we present our econometric analysis.

We notice that our estimated subjective wellbeing for all countries except Spain and France does not seem to feature a systematic increasing or decreasing trend, despite these countries going through very high economic growth rates from 1800 onwards. Spain and France show increasing and decreasing trends, respectively, but as we argued above, data in these countries may be subject to additional influences not associated with the countries themselves. The absence of a systematic rising trend in most countries is, however, consistent with the Easterlin paradox (Easterlin 1974, Easterlin, McVey, Switek, Sawangfa, and Zweig 2010).

Long-term changes in valence may arise from changes in the market for literature, albeit the direction of the bias is not clear. Firstly we might expect that, over the long run, as the target for a typical published book moved from the wealthy elite to the mass public, the content of these books would change. Moreover patterns in literary style changed considerably in the early part of the nineteenth century with the advent of greater literary realism (and social commentary) within literature. Literature portraying reality may have boosted the usage of words with lower levels of valence. On the other hand, books became more widely available and used for entertainment in addition to academic purposes, which might have biased the words towards higher levels of valence. We will address some of the issues deriving from the complexity of the market for books below in our econometric analysis and should also refer back to the discussion in section 3.

The valence measure we use is likely to be affected by the market for literature and,

---

the end of Korean War (1953) and the fall of Saigon (1975). In the UK, the Napoleonic Wars (1803-15). In Spain, the start of the Civil War (1936). In France, the Napoleonic Wars (1803-15), the end of the Franco-Prussian War (1870). For Germany, the vertical lines represent the Napoleonic Wars (1803-15), the Franco-Prussian War and reunification (1870), Hitler's ascendency to power (1934), the reunification (1990). In Italy, the unification (1861-70).

more generally, by the evolution of literature and language. As well as adding some control variable which can correct part of this omitted variables bias, we deal with this problem in two alternative ways corresponding to two different hypotheses on the evolution of literature and language, and we will show that the resultant models generate similar findings: i) in model 1 we assume that the market for books and language itself evolved in a similar way across the different countries we are considering, hence the introduction of year fixed-effects should correct any source of bias; ii) in model 2 we assume that the evolution of the market for books and of language itself affects written texts of different languages only in the medium and long term, hence by filtering away the data from medium and long term cycles we should be able to correct any source of bias to the extend that this generates deterministic cycles or trends. At the end of the section, we will also test for the presence the stochastic trends and cointegration in the main variables.

Econometric Models

Starting with model 1, we will then estimate our first model with time-specific fixed-effects, under the assumption that literature and language evolve in the same way in the different countries considered (an assumption we relax in model 2):

$$(2) \qquad \hat{Sat}_{i,t} = \beta_1 LifeExp_{i,t-1} + \beta_2 GDP_{i,t-\tau} + \sum_{z=1}^{Z} \delta_z x'_{z,i,t-1} + \gamma wc_{i,t} + \alpha_i + \eta_t + u_{i,t};$$

where i denotes the country and t the year; $\hat{Sat}_{i,t}$ is the estimated life satisfaction as previously defined; $LifeExp_{i,t-1}$ and $GDP_{i,t-\tau}$ are respectively per capita GDP (in logarithm) and life expectancy, our main explanatory variables; $x_{z,t-1}$ is the vector of control variables discussed earlier and listed in the different models of table III; α_i and η_t denote the country and year-specific effects, respectively. In the main specification of the model, the expression 2, we lag the regressors by one period as discussed in section 3. This corresponds to the hypothesis that the publication lag (the time passing between acceptance and actual publication) is one year and, the more general hypothesis discussed above, that books accepted for publication are the ones which better reflect a country's aggregate subjective wellbeing. In the appendix we consider different possible lags to show that the specification in model 2 is the one that best fits the data.

In model 2 we estimate the equation:

$$(3) \qquad S\tilde{\tilde{\hat{a}}}t_{i,t} = \beta_1 Lif\tilde{e}Exp_{i,t-1} + \beta_2 G\tilde{D}P_{i,t-1} + \sum_{z=1}^{Z} \delta_z \tilde{x}'_{z,i,t-1} + \gamma \tilde{wc}_{i,t} + \alpha_i + u_{i,t};$$

where the notation ˜ indicates that the original variable has been filtered using the HP filter ((Hodrick and Prescott 1997)). Initially we use a smoothing parameter $\lambda = 523.53$ corresponding to a *cutoff* point of 30 years. In table A.6 of the appendix we present a robustness check with different smoothing parameters.[21] As said, this model addresses the possibility of bias generated by systematic trends and in the last part of this section we will consider the possibility of stochastic trends. As it is well known, the HP filter takes away low frequency data, hence it filters both stochastic and deterministic long and medium-term cycles. Nevertherless, as a robustness check, we also test the order of integration of the main variables and if they are cointegrated. We then discuss the extent to which stochastic trends may bias our results.

Finally, note that we cluster errors at the language level to calculate standard errors. However, the disturbances in our model are likely to be changing over years for several reasons as already mentioned: the number of books available is lower in the earlier years, and while languages evolve our measure of valence is calibrated to more modern usage of language. For this reason in the appendix we will present the same regressions we present in the main text by clustering the standard errors at the year level.

Econometric Analysis

In table III we present the estimation of different specifications of model 1. From column 1 and 2 we note that the effect of life expectancy is positive and significant. As seems reasonable, the magnitude of the coefficient on life expectancy declines once we add a control for infant mortality (column 3), but its coefficient is still highly significant in column 4, where we added the full set of controls including infant mortality. The effect of GDP is not significant. This is consistent with the literature on happiness generally showing no relationship in time-series analysis between GDP and subjective wellbeing, that is *the Easterlin paradox*

[21] We will consider $\lambda = 1649.33$ (40 years) $\lambda = 253.288$ (25 years) , and $\lambda = 100$ the value that the European Central Bank uses for economic business cycles, see StataCorp (2015), page 612, for technical details on the HP filter.

(Easterlin 1974, Easterlin, McVey, Switek, Sawangfa, and Zweig 2010).As a robustness check we ran the same set of regressions without Spain and France and omitting the last 9 years of data.[22] The results are presented in the appendix in tables A.2 and A.3 respectively.

Considering the second model, where we applied a HP filter, we start by comparing our valence-based measure of expected subjective wellbeing with life expectancy, presented in figure 6. From the figure we note that there is a positive correlation between the two filtered variables in every country. In Spain this is weak. As we argued before Spanish data on valence can be particularly noisy due to the confounding effect of Latin American literature. In table IV we analyze the relationship between these two variables more systematically by estimating different specifications of model 2.

In this estimation we filtered out long-term cyclical components from all regressors, hence the coefficients are a measure of the short and medium-term effects. The coefficients of life expectancy are even stronger than before and are robust to the introduction of infant mortality with an almost unchanged coefficient, suggesting that the two effects on our valence-based measure of wellbeing are almost independent. The coefficient on GDP is again insignificant in this model.

Finally, it is interesting to note the negative and significant sign of democracy in the model presented in table IV (in table III the coefficients are similar but not significant). This perhaps suggests that democracy has a negative effect on the valence of words in publication in the short and medium-term. Hence an increase in democracy results in an increased freedom of the press that pushes authors to write more realistic (and potentially critical) books in the medium and in the short-term. This might also relate to the small literature on the relationship between subjective wellbeing and voting behavior which indicates that the mood of the nation has an important impact on electoral support for the incumbent administration (even if this changes in mood are not a result of any action taken by government) which lends support to the idea of a relationship between mood and democracy (Bagues and Esteve-Volart 2016, Liberini, Proto, and Redoano 2016). For whatever reason, the relationship between democracy and our valence measure does seem to weaken once long-term effects are included. As a robustness check we ran the same set of regressions without Spain and France and omitting the last 9 years of data. The results are presented in the appendix,

[22]As noted earlier, Google changed the way they measure word frequency in 2000, though omitting the last 9 years of data does not qualitatively change any of our results. See also table II, column 2 and the discussion surrounding this table in section 4.1.

in tables A.4 and A.5 respectively.

Stochastic Trends

In the analysis in table IV, we addressed the possibility that trends and long/medium-term cycles generated by languages, culture or other omitted factors might have biased our initial results. Here we explicitly address the possibility that omitted variables might have generated stochastic trends and biased the correlations presented above. If our estimated life satisfaction and the other regressors are integrated of order bigger than 0, this could potentially be a source of spurious correlation.

We tested the order of integration of our estimated life satisfaction for all languages and years we are considering in this section with the Augmented Dickey-Fuller unit-root test, and we find that for all, but Spanish, the presence of a unit root hypothesis can largely be rejected (while, as it is expected, for both GDP and life expectancy the same hypothesis cannot be rejected).[23] The presence of Spanish data can then be a source of bias due to the presence of stochastic trends in the dependent and in the independent variables. We can however rule out this possibility, as we checked the robustness of our results to the exclusion of Spain above.

Note that the fact that GDP and estimated life satisfaction have a different order of integration can explain the lack of correlation that we observed in some of the above regressions.

5. CONCLUDING COMMENTS

We have produced the first long-run time-series data for a number of countries that allows us to assess subjective wellbeing going back to 1776 (the American Declaration of Independence), adding 200 years to existing survey-based wellbeing measures which typically date back to the 1970s. In order to do this we formed a measure of positive valence from many millions of books digitized in the Google Books corpus.

Being the very first such index it comes with caveats: section 3 discusses many of the key issues, such as how far back we might reasonably go, what the index actually measures and the right uses (and wrong uses) of the data. Using conventional regression analysis and non-parametric methods we show that our new measure is highly consistent with existing wellbeing measures going back to 1973 and incidentally indicates that on average the mood

[23] We omit the details of the tests, which are available upon request.

of books matches the mood of the population (which is a result rather than an assumption). This is further supported by another results that corresponds to general intuition: historical events that are popularly regarded as positive and negative (i.e., wars, depressions, and exhuberance) have corresponding rises and falls in our sentiment measure.

We reiterate that having provided the first long-run data of this sort there are no easy ways to compare our results with other results which by their nature cannot go back as far in time as ours. For example, while we offer some support for the Easterlin paradox with long-run time-series, other discussions of the paradox are necessarily based on cross-sectional data or much shorter-run analyses. Our measure gives some insight into why the relationship with GDP is difficult to establish and perhaps reconcile those who do see a relationship with those who do not. Similarly, our findings on the relationship between happiness over the long-run and life expectancy and child mortality are completely new findings and indicate that there is a significant long-run relationship between our measure of subjective wellbeing and life expectancy at birth. Given the absence of any other good long-run proxies for physical health, this is indicative of a long-run relationship between health and happiness.

We caution against long-term interpretations of simple trends in our measure since both the market for books, and language itself, have evolved considerably over the period we consider e.g. Hills and Adelman (2015). We nevertheless argue that this is a similar issue in spirit to the problem of comparing economic growth and income levels across many centuries when lifestyles have changed beyond recognition: essentially we would argue that caution is needed when considering any very long-run socio-economic data, but the utility of having long-run data remains non-trivial.[24] We also partly correct for these problems by following different econometric approaches (and robustness checks) and also by controlling for stochastic trends and potential cointegration in the data.

A key issue in particular with the limited survey-based data that is traditionally used to measure subjective wellbeing is our current inability to consider the impact on national subjective wellbeing of major nation-level shocks such as wars, epidemics, natural disasters and our only very limited ability to consider the long-run relationship between subjective

[24]Consider for instance the arrival of urbanization, huge cultural and political shifts, increased technological advances (mechanization, computerization, mobile telephony, the internet and so on) and countless other important changes that make inter-temporal comparisons of national income difficult. However, for the same reason that we would support the work done by economic historians and macroeconomists in pushing back GDP data as far as possible despite these issues, we would similarly point to the many uses of a longer-run measure of wellbeing.

wellbeing, GDP and life expectancy. We presented two models that seek to make longer-run comparison more feasible and provide a first attempt to use our measure to study subjective wellbeing in the USA, UK, Germany, France, Italy and Spain for more than two centuries.

We finish with two points. First we would like to emphasize that the dataset we have developed can and should be used to help further our understanding of subjective wellbeing and we will make our methods and data available to other researchers for further analysis. Second, we would also like to draw attention to a broader message. The availability of "big data" opens up many new doors to a better understanding of historical attitudes and (economic) behavior. Certainly the methods employed in our work would be equally applicable to other socio-economic variables aside from subjective wellbeing. As we note in the introduction work already exists (and is ongoing) which uses similar methods to examine opinions about political candidates, stock-market trends, mood variations and the impact of specific events, but there is still much potential for future research in this area.

REFERENCES

Alberge, J. (2015): "Publishers bypass literary agents to discover bestseller talent," Discussion paper, The Observer, February 22.

Bagues, M., and B. Esteve-Volart (2016): "Politicians luck of the draw: Evidence from the Spanish Christmas lottery.," *Journal of Political Economy*, 124(5), 1269–1294.

Benjamin, D. J., M. S. Kimball, O. Heffetz, and A. Rees-Jones (2012): "What do you think would make you happier? What do you think you would choose?," *The American economic review*, 102(5), 2083.

Bernard, A. (1990): *Rotten Rejections*. Robson Books.

Bollen, J., H. Mao, and X. Zeng (2011): "Twitter mood predicts the stock market," *Journal of Computational Science*, 2(1), 1–8.

Bradley, M. M., and P. J. Lang (1999): "Affective norms for English words (ANEW): Instruction manual and affective ratings," Discussion paper, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Broadberry, S., B. Campbell, A. Klein, M. Overton, and B. Van Leeuwen (2012): "British Economic Growth, 1270-1870: an output-based approach," Discussion paper, School of Economics Discussion Papers.

Chen, M. K. (2013): "The Effect of Language on Economic Behavior: Evidence from Savings

Rates, Health Behaviors, and Retirement Assets," *American Economic Review*, 103(2), 690–731.

Chmiel, A., J. Sienkiewicz, M. Thelwall, G. Paltoglou, K. Buckley, A. Kappas, and J. A. Hołyst (2011): "Collective emotions online and their influence on community life," *PloS one*, 6(7), e22207.

Connor, B., R. Balasubramanyan, B. R. Routledge, and N. A. Smith (2010): "From tweets to polls: Linking text sentiment to public opinion time series," *ICWSM*, 11, 122–129.

Deaton, A. (2008): "Income, Health, and Well-being Around the World: Evidence from the Gallup World Poll," *Journal of Economic perspectives*, 22(2), 53–72.

Di Tella, R., R. J. MacCulloch, and A. J. Oswald (2001): "Preferences over inflation and unemployment: Evidence from surveys of happiness," *American Economic Review*, pp. 335–341.

Dodds, P. S., and C. M. Danforth (2010): "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents," *Journal of Happiness Studies*, 11(4), 441–456.

Dodds, P. S., K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth (2011): "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter," *PloS one*, 6(12), e26752.

Easterlin, R. A. (1974): "Does economic growth improve the human lot? Some empirical evidence," *Nations and households in economic growth*, 89, 89–125.

Easterlin, R. A., L. A. McVey, M. Switek, O. Sawangfa, and J. S. Zweig (2010): "The happiness–income paradox revisited," *Proceedings of the National Academy of Sciences*, 107(52), 22463–22468.

Golder, S. A., and M. W. Macy (2011): "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, 333(6051), 1878–1881.

Greenfield, P. M. (2013): "The changing psychology of culture from 1800 through 2000," *Psychological science*, 24(9), 1722–1731.

Hills, T. T., and J. S. Adelman (2015): "Recent evolution of learnability in American English from 1800 to 2000," *Cognition*, 143, 87–92.

Hodrick, R. J., and E. C. Prescott (1997): "Postwar US business cycles: an empirical investigation," *Journal of Money, credit, and Banking*, pp. 1–16.

JERVEN, M. (2012): "An Unlevel Playing Field: National Income Estimates and Reciprocal Comparison in Global Economic History," *Journal of Global History*, 7, 107–128.

LANG, P. J. (1980): *Behavioral treatment and bio-behavioral assessment: Computer applications*. Ablex, Norwood, NJ.

LIBERINI, F., E. PROTO, AND M. REDOANO (2016): "Happy Voters," Discussion paper, Forthcoming in the Journal of Public Economics.

LIN, Y., J.-B. MICHEL, E. L. AIDEN, J. ORWANT, W. BROCKMAN, AND S. PETROV (2012): "Syntactic annotations for the google books ngram corpus," in *Proceedings of the ACL 2012 system demonstrations*, pp. 169–174. Association for Computational Linguistics.

MICHEL, J.-B., Y. K. SHEN, A. P. AIDEN, A. VERES, M. K. GRAY, J. P. PICKETT, D. HOIBERG, D. CLANCY, P. NORVIG, J. ORWANT, ET AL. (2011): "Quantitative analysis of culture using millions of digitized books," *Science*, 331(6014), 176–182.

MITCHELL, B. (2013): *International Historical Statistics*. Palgrave Macmillan.

MONNIER, C., AND A. SYSSAU (2014): "Affective norms for French words (FAN)," *Behavior research methods*, 46(4), 1128–1137.

MONTEFINESE, M., E. AMBROSINI, B. FAIRFIELD, AND N. MAMMARELLA (2014): "The adaptation of the Affective Norms for English Words (ANEW) for Italian," *Behavior research methods*, 46(3), 887–903.

NGUYEN, T., D. PHUNG, B. ADAMS, T. TRAN, AND S. VENKATESH (2010): "Classification and pattern discovery of mood in weblogs," in *Advances in Knowledge Discovery and Data Mining*, pp. 283–290. Springer.

PROTO, E., AND A. RUSTICHINI (2013): "A reassessment of the relationship between GDP and life satisfaction," *PloS one*, 8(11), e79358.

REDONDO, J., I. FRAGA, I. PADRÓN, AND M. COMESAÑA (2007): "The Spanish adaptation of ANEW (affective norms for English words)," *Behavior research methods*, 39(3), 600–605.

SCHMIDTKE, D. S., T. SCHRÖDER, A. M. JACOBS, AND M. CONRAD (2014): "ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words," *Behavior research methods*, 46(4), 1108–1118.

STATACORP, L. (2015): *Stata Time-series: Reference Manual release 14*. StataCorp LP.

STEVENSON, B., AND J. WOLFERS (2008): "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*, pp. 1–87.

THELWALL, M., K. BUCKLEY, AND G. PALTOGLOU (2011): "Sentiment in Twitter events," *Journal of the American Society for Information Science and Technology*, 62(2), 406–418.

van Zanden, J. L., J. Baten, M. Mira dErcole, A. Rijpma, C. Smith, and M. Timmer (2014): "How was life?: Global well-being since 1820," .

Warriner, A. B., V. Kuperman, and M. Brysbaert (2013): "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior research methods*, 45(4), 1191–1207.

FIGURES AND TABLES

Variable	Mean	Std. Dev.	Min.	Max.	N
Valence	5.72	0.114	5.302	6.07	1259
Life Satisfaction	2.949	0.163	2.52	3.23	163
Est. Life Satisf.	2.9	0.216	2.125	3.352	1259
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
Education Inequality	26.964	19.559	6.111	98.935	784
Infant Mortality	96.657	81.822	3	332	892
Words Covered	0.049	0.057	0	0.191	1259
Life Satisfaction	2.949	0.163	2.52	3.23	163

TABLE I

Main Variables.

ENGLISH	VALENCE	GERMAN	VALENCE	FRENCH	VALENCE	ITALIAN	VALENCE	SPAIN	VALENCE
aardvark	6.26	Aas	-2.6	abeille	4.22	abbaglio	3 94	abandonado	1.68
abalone	5.3	Abenddämmerung	-2.35	abonné	4.53	abbandonato	2	abejas	3.18
abandon	2.84	Abendessen	2.1	abricot	6.55	abbondanza	6 82	aborto	2 8
abandonment	2.63	Abenteuer	0.81	absent	3.42	abbraccio	7.7	abrasador	2.46
abbey	5.85	Abfall	1.44	abstrait	4.72	abete	6.17	abrazo	8.13
abdomen	5.43	abkochen	0.4	accordéon	5.7	abitante	5.67	abrumado	2 9
abdominal	4.48	Abschaum	1.9	acide	3.47	abitazione	6.46	absurdo	3 8
abduct	2.42	Abscheu	-1.38	agneau	6.35	abito	7 27	abundancia	6 8
abduction	2.05	Absturz	-1.6	agréable	8.29	abitudini	4 91	aburrido	2 33
abide	5.52	absurd	-2.7	aide	7.08	aborto	2 06	accidente	1 32
abiding	5.57	Abtreibung	-2.55	aigle	6.53	abuso	1.74	ácido	3.41
ability	7	aggressiv	-1.8	aiguille	3.9	accettazione	5.79	acogedor	7.64
abject	4	aktivieren	-0.6	ail	4.22	accogliente	8 03	acontecimier	5 99
ablaze	5.15	Alarm	1.5	aile	6.05	accomodante	6.4	acre	4 23
able	6.64	Alimente	-0.79	aisance	7.26	accordo	6.71	activar	6
abnormal	3.53	Alkoholiker	2.15	album	6.34	acqua	7.78	acuerdo	7 24
abnormality	3.05	Allee	-1.9	alcool	5.64	adorabile	7 33	acurrucarse	6 98
abode	5.28	allein	-1.27	algèbre	3.87	adulto	5.78	adicto	2.41
abolish	3.84	Allergie	-1.56	allégorie	5.42	aereo	6 56	adinerado	6 21
abominable	4.05	Alptraum	-1.56	alligator	4.05	affamato	4.74	admirado	7 33
abomination	2.5	anbetungswürdig	-1.22	allumette	5.32	affascinare	7 97	adorable	7.48
abort	3.1	angeekelt	0.73	ambition	7.6	affaticato	3.73	adulto	5.68
abortion	2.58	angespannt	1.53	ambulance	3.22	affetto	7.48	afectar	3.48
abracadabra	5.11	Angriff	-2.1	âme	7.12	afflizione	1 94	afecto	8.1
abrasive	4.26	ängstlich	1	amer	2.8	affogare	1.79	afianzar	5 93
abreast	4.62	Anreiz	-1.93	ami	7.94	aggressione	2 53	afligido	1 96
abrupt	3.28	Anstellung	-2.21	amitié	8.38	aggressivo	3.48	afortunado	7.71

Figure 1.— A Sample of Word Valence in Different Languages.

The first words (alphabetically) and their respective valences in the languages considered. The words are coded from 1 to 9 in all languages, apart from German where words have been coded from -3 to 3.

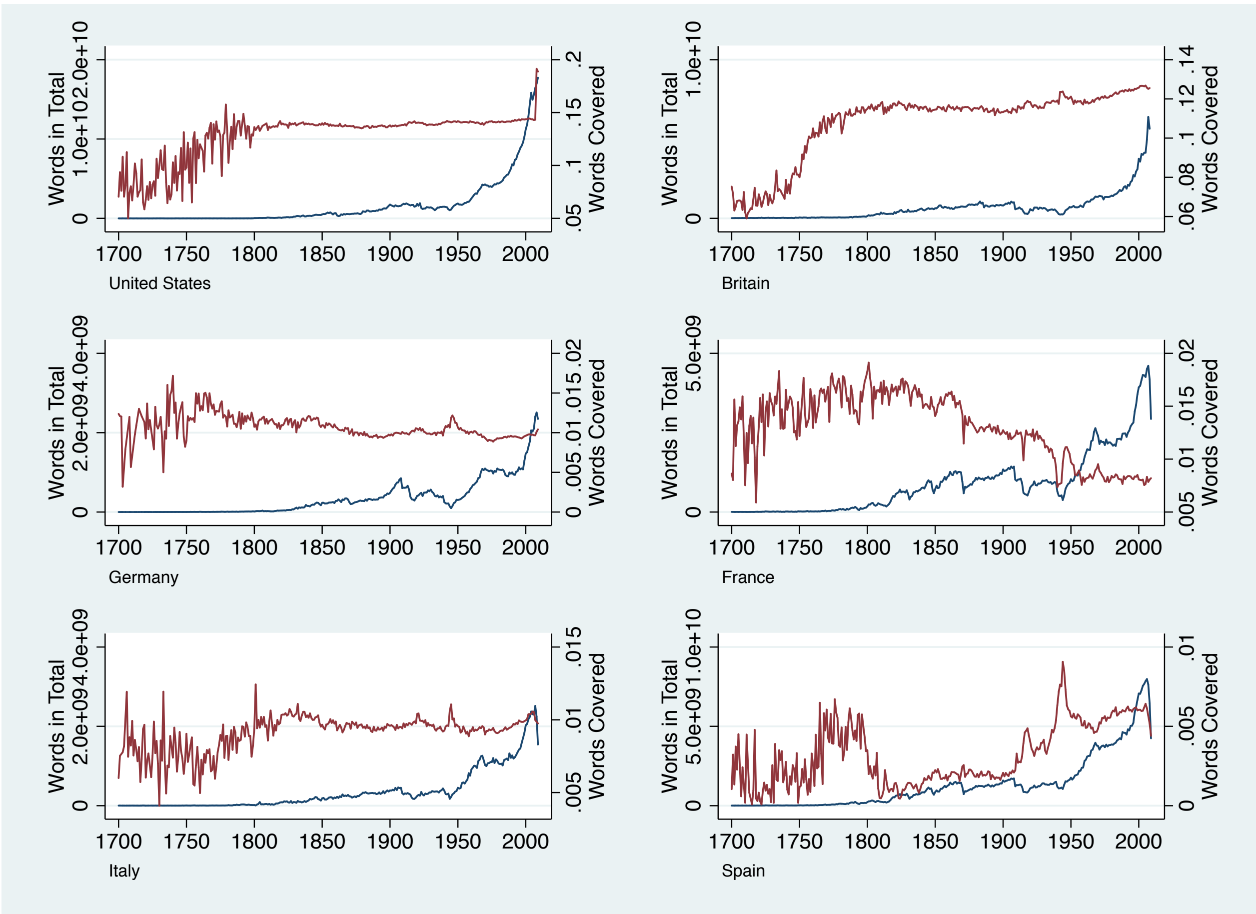

FIGURE 2.— The Number of Words and Share of Words Covered.

The red line represents the share of words covered over the total measured in the right axis, the blue line represents the total number of words measured in the left axis, for all countries considered in the analysis.

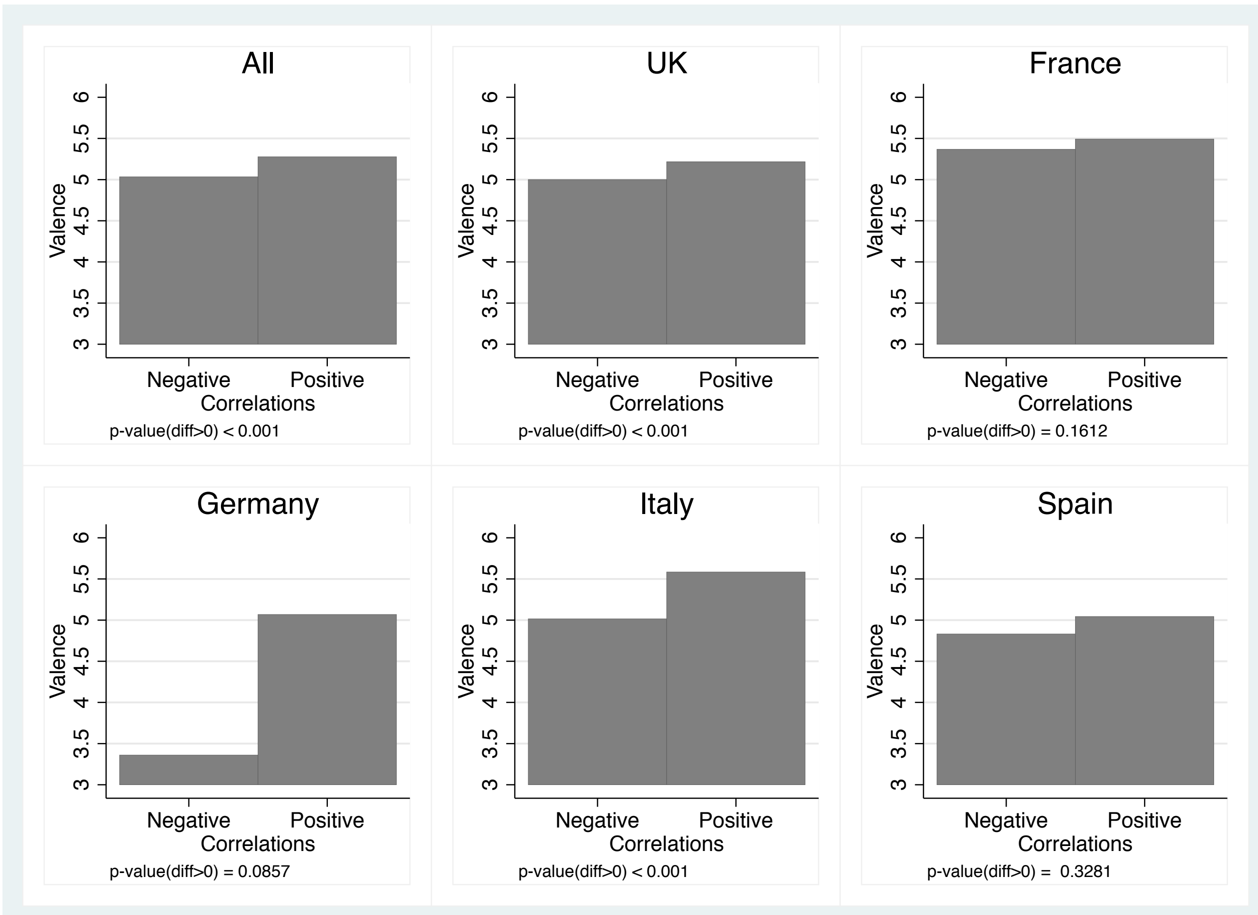

FIGURE 3.— Average valence and correlations with life satisfaction.

The bars in the panel represent the average valence of words featuring significantly (at the 5% level) positive and negative levels of correlations with the corresponding levels of average life satisfaction significantly different from 0.

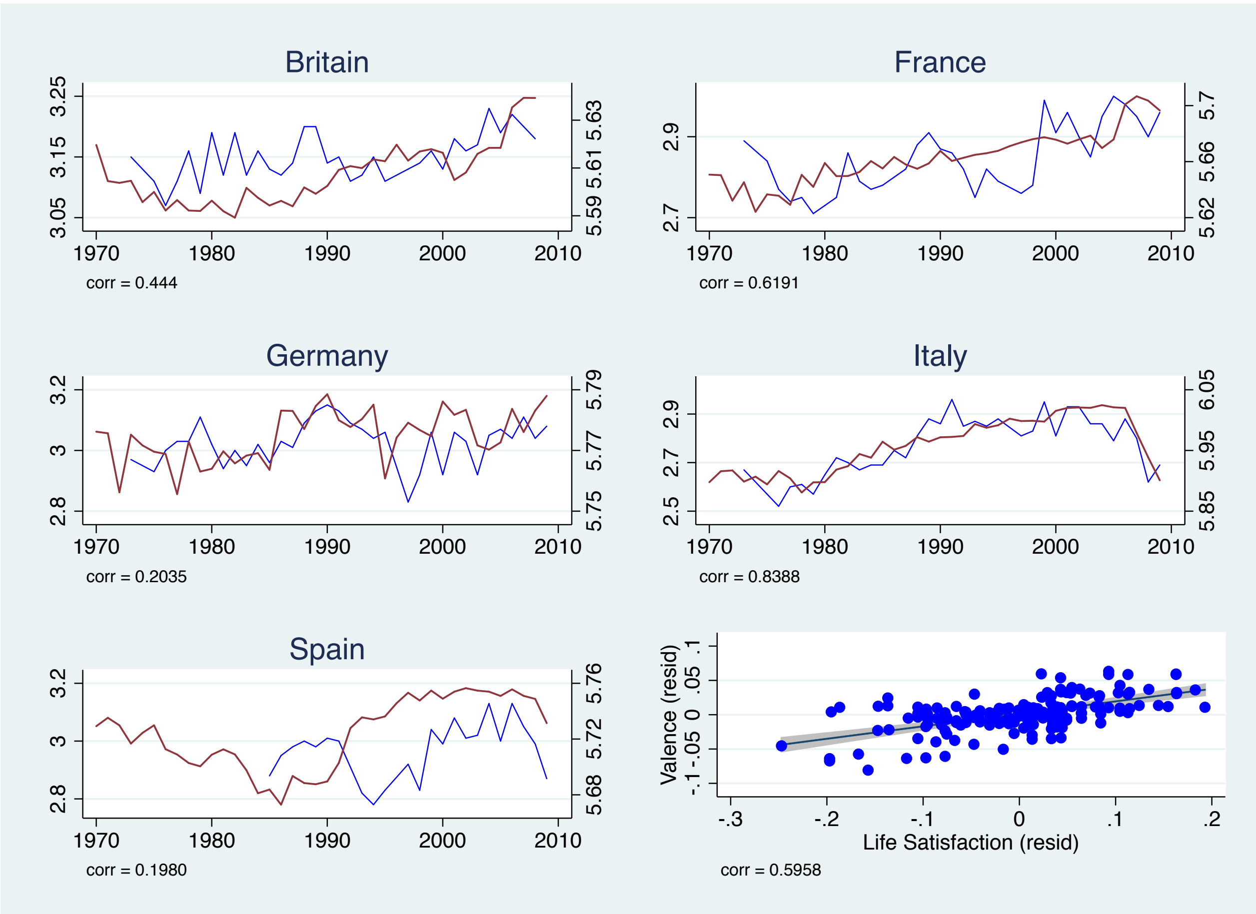

FIGURE 4.— Valence and Aggregate Life Satisfaction.

In the first 5 panels presenting the data in time-series. Valence is represented in red (values in the left axis), life satisfaction is represented in blue (values in the right axis). In the last panel, we plotted valence against life satisfaction for the same countries and periods; both variables are expressed in the form of residuals after controlling for country fixed-effects. The grey area represents the 95% confidence interval of the regression line.

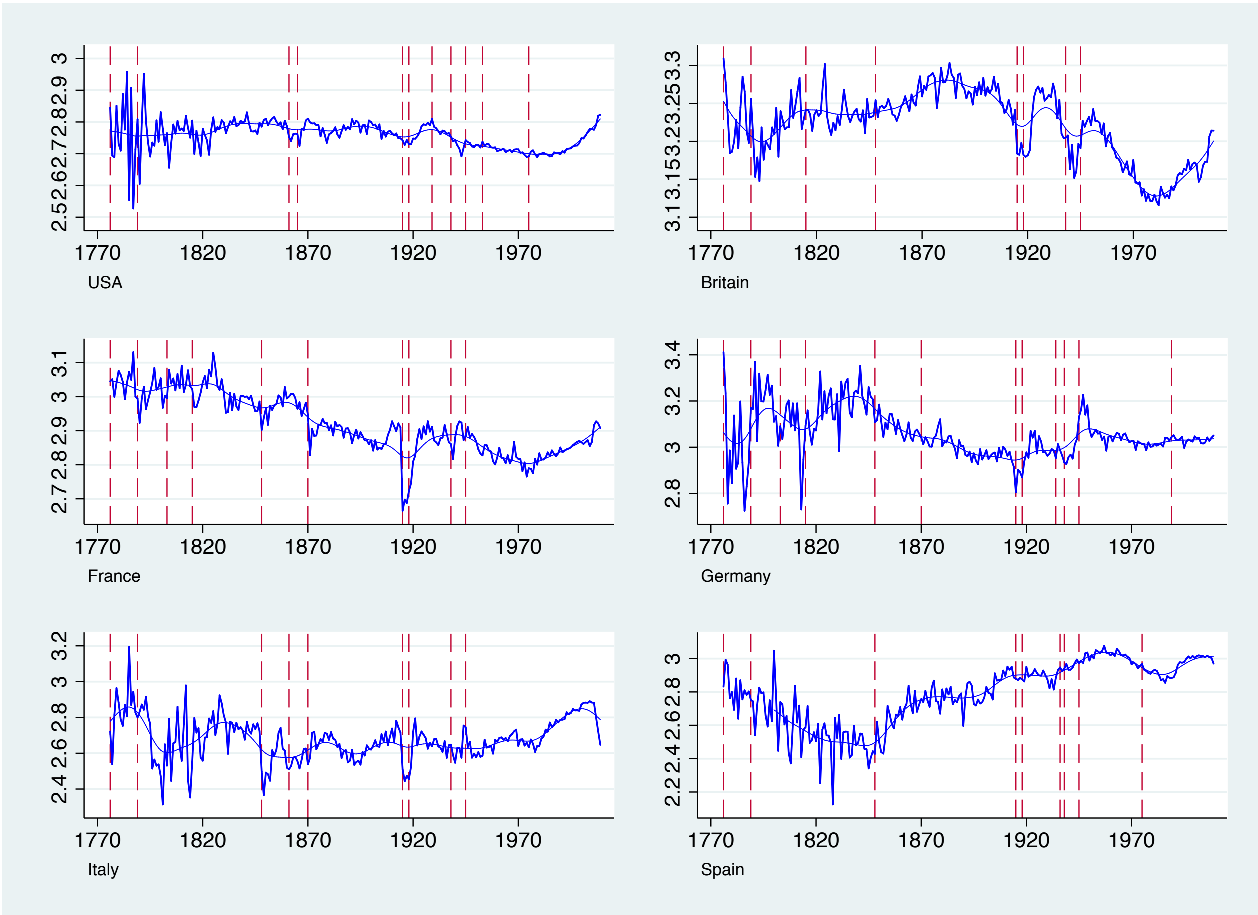

FIGURE 5.— Estimated Subjective Wellbeing Over the Period 1776-2009.

The blue lines represent the estimated life satisfaction derived from the valence measures for each country. The vertical red lines correspond to 1789, the year of the French Revolution, to World War I (1914-18) and to World War II (1939-45). In the 5 European countries a line is drawn in 1848, the "Year of Revolution". In the USA, the vertical lines represent: the Civil War (1861-65), the Wall Street Crash (1929), the end of Korean War (1953) and the fall of Saigon (1975). In the UK, the Napoleonic Wars (1803-15). In Spain, the starting of Civil War (1936). In France, the Napoleonic Wars (1803-15), the end of the Franco-Prussian War (1870). For Germany, the vertical lines represent the Napoleonic Wars (1803-15), the Franco-Prussian War and reunification (1870), Hitler's ascendency to power (1934), the reunification (1990). In Italy, the unification (1861-70). The thin blue lines represent the trend components of each series obtained using the Hodrick and Prescott filter.

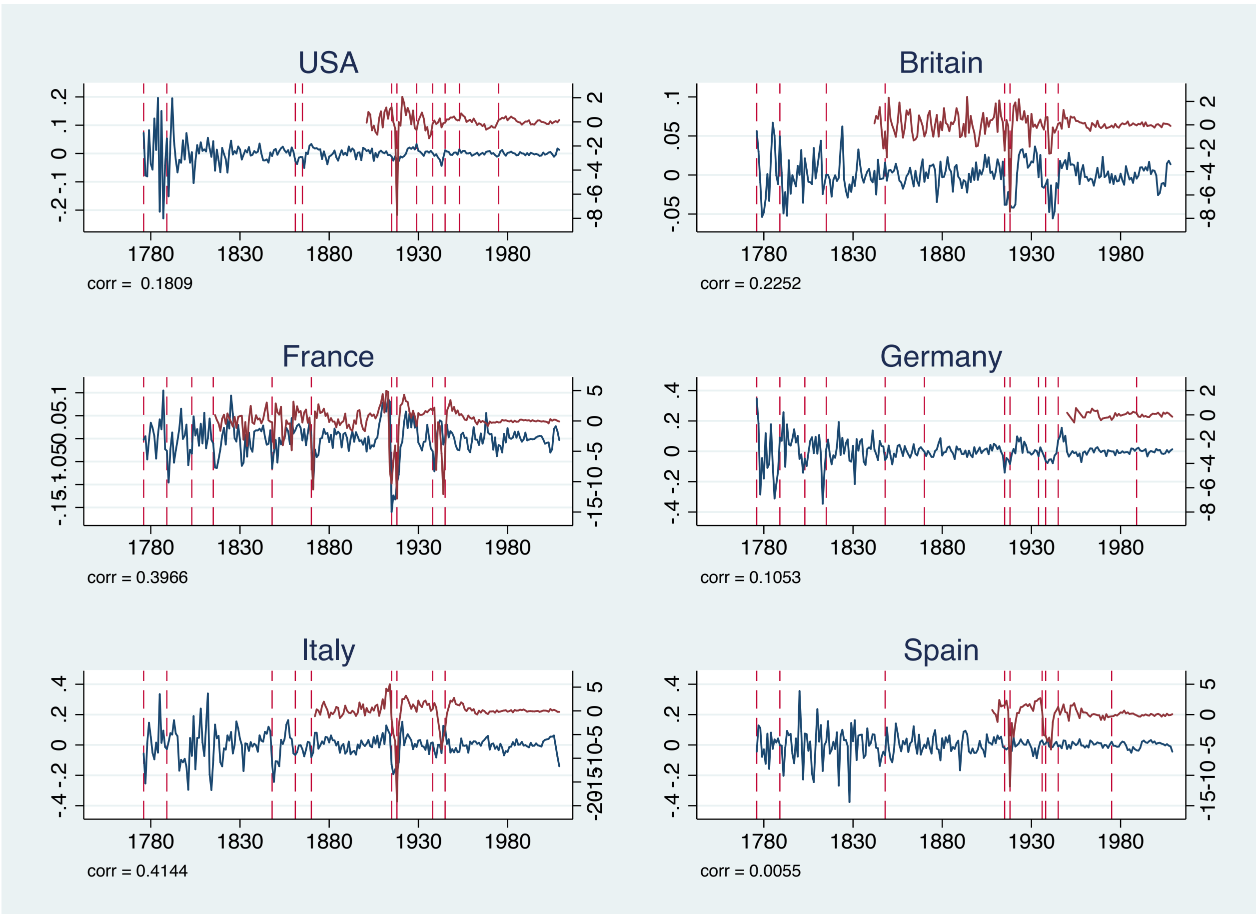

FIGURE 6.— Detrended Estimated Subjective Wellbeing and Life Expectancy.

The red line represents life expectancy at time $t-1$, the blue line represents estimated subjective wellbeing at time t derived from the valence measures for each country. Both series are detrended by using the Hodrick and Prescot filter.

	1	2	3	4	5	6
	Baseline	Until 2000	w/o Sp.and Fr.	Trends	+GDP	Year FE
	b/se	b/se	b/se	b/se	b/se	b/se
Valence	1.9554***	1.6941***	2.1696***	1.5549***	0.7180**	1.6107***
	(0.2221)	(0.3093)	(0.2339)	(0.3408)	(0.3499)	(0.2784)
Log GDP					0.8243***	0.1452
					(0.1537)	(0.1300)
Words Covered	0.9816	–0.0037	–0.4491	8.7147	–0.1693	–15.6331*
	(6.2645)	(9.1248)	(5.7111)	(15.0425)	(13.9245)	(8.5557)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Country Specific Trend	No	No	No	Yes	Yes	No
Year FE	No	No	No	No	No	Yes
r2	0.358	0.227	0.501	0.387	0.485	0.645
N	163	119	104	163	163	163

TABLE II

VALENCE PREDICTS AGGREGATE LIFE SATISFACTION.

Average life satisfaction per country and year from the Eurobarometer dataset is the dependent variable. In columns 1, 4, 5 and 6 the years are 1983-2009, in column 3, 1972-2009, and in column 2, 1983-2000. The countries in columns 1,2,4,5 and 6 are: France, Germany, Italy, Spain, and the UK. Per Capita GDP (expressed in terms of purchasing power parity) is from the PWT 8.0 dataset. Standard errors are given in brackets. * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1 Baseline b/se	2 Controls b/se	3 Infant Mort. b/se	4 Conflicts b/se
Life Expectancy t-1	0.0089***	0.0084***	0.0045***	0.0045**
	(0.0010)	(0.0013)	(0.0011)	(0.0013)
GDP (log) t-1	0.0121	0.0396	0.0252	–0.0047
	(0.0847)	(0.0707)	(0.0711)	(0.0620)
Infant Mortality t-1			–0.0012***	–0.0011**
			(0.0003)	(0.0003)
Internal Conflict^{-} t-1				–0.0061
				(0.0090)
External Conflict^{-} t-1				–0.0070
				(0.0080)
WW1 t-1				–0.0512
				(0.0264)
WW2 t-1				–0.0295
				(0.0321)
Democracy		–0.0020	–0.0023	–0.0029
		(0.0016)	(0.0013)	(0.0015)
Education Inequality		–0.0006	–0.0007	–0.0014**
		(0.0007)	(0.0007)	(0.0005)
Words Covered		–1.1861	–1.0702	–8.8518*
		(2.2457)	(2.3227)	(4.1258)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
r2	0.548	0.482	0.506	0.549
N	781	657	639	592

TABLE III

Historical Determinants of Estimated Subjective Wellbeing.

Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust Standard Errors are given in brackets. * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	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

TABLE IV

Historical Determinants of Estimated Subjective Wellbeing, with Detrended Data.

Estimated Subjective Wellbeing per country and year is the dependent variable. Data are detrended using the Hodrick and Prescott filter. Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust Standard Errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

APPENDIX A: FOR ONLINE PUBLICATION

Below we discuss different possible lags of the regressors and also provide more detail on decision-making by publishers as referenced in the main text.

A.1. *Different Lags of the Regressors*

Consider the expression

$$\text{(A-1)} \quad \tilde{\hat{S}}at_{i,t} = \beta_1 Lif\tilde{e}Exp_{i,t-\tau} + \beta_2 G\tilde{D}P_{i,t-\tau} + \sum_{z=1}^{Z} \delta_z \tilde{x}_{z,i,t-\tau} + \gamma \tilde{wc}_{i,t} + \alpha_i + u_{i,t};$$

where $\tau \geq 1$ represents a generic temporal lag.

We compare three different models determining the channels through which a country's subjective wellbeing is factored into the different written languages:

1. the books published in the market reflect current subjective wellbeing (publishers selecting books that match the current mood of the population). In this case $\tau = 0$

2. as before, but the publisher's decision to publish is taken on the basis of the subjective wellbeing one year before, i.e there exists a publishing lag of one year. In this case $\tau = 1$

3. a book published at time t reflects the subjective wellbeing of the population three years prior to publication, i.e. there exists a publishing lag of three years. In this case we assume $\tau = 3$.

In the first 3 columns of table A.1, we present estimation corresponding to the above models. Note that GDP is never significant, while life expectancy is always significant with a coefficient which seems to be decreasing from lag 1 to lag 3. In column 3 we consider all models altogether. From this specification we note that the one-year lag *wins the horserace*, being the only one where life expectancy remains significant.

	1	2	3	4
	no lag	1 year lag	3 years lag	All
	b/se	b/se	b/se	b/se
Life Expectancy	0.0086***			0.0023
	(0.0021)			(0.0024)
Life Expectancy t- 1		0.0084***		0.0064**
		(0.0013)		(0.0016)
Life Expectancy t- 3			0.0065**	0.0001
			(0.0019)	(0.0027)
GDP (log)	0.0292			–0.0154
	(0.0668)			(0.0350)
GDP (log) t-1		0.0349		–0.0680***
		(0.0745)		(0.0159)
GDP (log) t-3			0.0657	0.1329
			(0.0995)	(0.0939)
Democracy	–0.0019			–0.0048
	(0.0014)			(0.0024)
Democracy t- 1		–0.0015		0.0028
		(0.0014)		(0.0022)
Democracy t- 3			–0.0015	0.0000
			(0.0012)	(0.0014)
Education Inequality	–0.0006			–0.0040*
	(0.0007)			(0.0018)
Education Inequality t-1		–0.0005		0.0035
		(0.0007)		(0.0026)
Education Inequality t-3			–0.0005	–0.0003
			(0.0008)	(0.0044)
Words Covered	–2.4926	–1.0898	–0.7023	–2.3755
	(2.7456)	(2.1823)	(1.7609)	(2.5093)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
r2	0.468	0.477	0.445	0.521
N	653	652	640	628

TABLE A.1

THE HISTORICAL DETERMINANTS OF ESTIMATED SUBJECTIVE WELLBEING, USING DIFFERENT TIME LAGS IN THE REGRESSORS.

Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors are given in brackets. * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1 Baseline b/se	2 Controls b/se	3 Infant Mort. b/se	4 Conflicts b/se
Life Expectancy t-1	0.0096**	0.0075**	0.0047	0.0031***
	(0.0023)	(0.0023)	(0.0024)	(0.0004)
GDP (log) t-1	0.1289	0.1234*	0.1268*	0.0688
	(0.0691)	(0.0491)	(0.0457)	(0.0331)
Infant Mortality t-1			–0.0006*	–0.0008**
			(0.0002)	(0.0002)
Internal Conflict^{-} t-1				0.0030
				(0.0099)
External Conflict^{-} t-1				–0.0006
				(0.0117)
WW1 t-1				–0.0565
				(0.0481)
WW2 t-1				0.0057
				(0.0307)
Democracy		0.0037**	0.0030*	0.0009
		(0.0011)	(0.0012)	(0.0004)
Education Inequality		0.0007	0.0002	–0.0003
		(0.0003)	(0.0002)	(0.0002)
Words Covered		0.1986	0.1944	–11.0697**
		(2.0306)	(1.9726)	(2.0251)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
r2	0.621	0.692	0.698	0.744
N	487	412	396	365

TABLE A.2

The Historical Determinant of Estimated Subjective Wellbeing, excluding Spain and France.

Estimated Subjective Wellbeing per country and year is the dependent variable. The Countries are: Germany, Italy, the UK, and the United States. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	1	2	3	4
	Baseline	Controls	Infant Mort.	Conflicts
	b/se	b/se	b/se	b/se
Life Expectancy t-1	0.0088***	0.0085***	0.0052**	0.0044**
	(0.0010)	(0.0017)	(0.0014)	(0.0013)
GDP (log) t-1	0.0047	–0.0007	–0.0147	–0.0060
	(0.0839)	(0.0635)	(0.0633)	(0.0621)
Infant Mortality t-1			–0.0011**	–0.0011**
			(0.0003)	(0.0003)
Internal Conflict^{-} t-1				–0.0076
				(0.0081)
External Conflict^{-} t-1				–0.0075
				(0.0079)
WW1 t-1				–0.0521
				(0.0266)
WW2 t-1				–0.0288
				(0.0315)
Democracy		–0.0023	–0.0025*	–0.0028
		(0.0014)	(0.0013)	(0.0014)
Education Inequality		–0.0011**	–0.0013**	–0.0014**
		(0.0004)	(0.0004)	(0.0004)
Words Covered		–8.5931*	–9.0217*	–8.5797*
		(3.6496)	(3.6485)	(4.0337)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
r2	0.556	0.510	0.537	0.546
N	728	604	586	586

TABLE A.3

Historical Determinants of Estimated Subjective Wellbeing, 1800-2000.

Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	1 Baseline b/se	2 Controls b/se	3 Infant Mort. b/se	4 Conflicts b/se
Life Expectancy t-1	0.0071**	0.0073***	0.0073**	0.0053**
	(0.0014)	(0.0011)	(0.0014)	(0.0012)
GDP (log) t-1	–0.0698	–0.0815	–0.0843	–0.0666
	(0.0519)	(0.0724)	(0.0747)	(0.0554)
Infant Mortality t-1			–0.0001	0.0002
			(0.0003)	(0.0004)
Internal Conflict$^-$ t-1				–0.0006
				(0.0022)
External Conflict$^- t-1$				–0.0015
				(0.0020)
WW1 t-1				–0.0610
				(0.0336)
WW2 t-1				0.0095
				(0.0180)
Democracy		–0.0014	–0.0015	–0.0003
		(0.0030)	(0.0030)	(0.0016)
Education Inequality		0.0013	0.0008	0.0009
		(0.0011)	(0.0021)	(0.0023)
Words Covered		0.4244	0.4451	1.1691
		(0.5865)	(0.5960)	(4.5234)
r2	0.151	0.164	0.166	0.256
N	474	406	391	360

TABLE A.4

Historical Determinants of Estimated Subjective Wellbeing, excluding Spain and France and with Data Detrended.

Data detrended using the Hodrick and Prescott Filter. Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	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)	(0.0013)	(0.0009)
GDP (log) t-1	–0.0598	–0.0669	–0.0743	–0.0526
	(0.0390)	(0.0451)	(0.0401)	(0.0265)
Infant Mortality t-1			–0.0004**	–0.0002
			(0.0002)	(0.0002)
Internal Conflict^{-} t-1				–0.0013
				(0.0018)
External Conflict$^{-}t-1$				0.0017
				(0.0025)
WW1 t-1				–0.0748**
				(0.0224)
WW2 t-1				0.0075
				(0.0138)
Democracy		–0.0027*	–0.0028*	–0.0021**
		(0.0011)	(0.0011)	(0.0007)
Education Inequality		0.0001	–0.0009	–0.0001
		(0.0012)	(0.0016)	(0.0012)
Words Covered		1.7633	2.5019	3.4562
		(4.3166)	(4.7718)	(4.2242)
Country FE	Yes	Yes	Yes	Yes
r2	0.134	0.178	0.189	0.297
N	712	595	580	580

TABLE A.5

Historical Determinants of Estimated Subjective Wellbeing, 1800-2000 and Data Detrended.

Data detrended using the Hodrick and Prescott filter. Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	1	2	3	4	5	6
	$\lambda = 100$	$\lambda = 253.288$	$\lambda = 523.53$	$\lambda = 1649.33$	No Filter	No Filter
	b/se	b/se	b/se	b/se	b/se	b/se
Life Expectancy t-1	0.0058***	0.0061***	0.0062***	0.0063***	0.0053***	0.0045***
	(0.0014)	(0.0014)	(0.0014)	(0.0013)	(0.0012)	(0.0011)
GDP (log) t-1	–0.0608	–0.0681	–0.0724	–0.0757*	–0.0132	0.0252
	(0.0384)	(0.0404)	(0.0387)	(0.0311)	(0.0356)	(0.0711)
Infant Mortality t-1	–0.0005***	–0.0004**	–0.0004**	–0.0004*	0.0009**	–0.0012***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)	(0.0003)
Democracy	–0.0027**	–0.0027**	–0.0027**	–0.0027**	0.0005	–0.0023
	(0.0009)	(0.0009)	(0.0009)	(0.0007)	(0.0023)	(0.0013)
Education Inequality	–0.0023	–0.0017	–0.0003	0.0015	0.0003	–0.0007
	(0.0013)	(0.0014)	(0.0015)	(0.0016)	(0.0010)	(0.0007)
Words Covered	0.6722	0.7057	0.7196	0.7479	–0.9102	–1.0702
	(0.5976)	(0.6654)	(0.7141)	(0.7703)	(2.7268)	(2.3227)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	No	No	No	Yes
r2	0.160	0.172	0.175	0.174	0.066	0.506
N	633	633	633	633	639	639

TABLE A.6

Historical Determinants of Estimated Subjective Wellbeing with Different Smoothing Parameters.

Data detrended using the Hodrick and Prescott filter. Estimated Subjective Wellbeing per country and year is the dependent variable. The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

A.2. *Rejection letters*

In this appendix we note some quotes from publishers and authors concerning their rationale for rejecting books submitted for publication. The aim is to provide some supporting evidence for the importance of the potential demand-side to publishers which is sufficient for publishers to routinely reject books that they do not feel are a good match for the market. While this is a natural assumption for economists in any case as part of a broader assumption of profit-maximizing behavior, it might still be worth providing some evidence that publishers are not an exception to normal behavior.

We first need to note that there is a strong "survivor bias when looking at rejection letters/decisions that are easily observable: the vast majority of books that are rejected by publishers will not see print and it is highly unlikely that rejection letters will then ever come to light. In fact, the rejection letters that survive tend to be for books which become successful. What is helpful for us is that the bias works in favour of our hypothesis: if publishers are rejecting books that later do become a success on market-based grounds, how much more do they reject those that never come to print on those same grounds? In essence what we are saying is that given what we can find it is likely that there are many more such accounts of rejections on a market basis that will never be found as the books in question never see print and the authors drop out of the supply-side of the market. What follows then are a few notable examples for quite famous books which hint at the importance that publishers place on the marketable nature of books and whether books are a good match for readers: note that these authors and books were eventually printed at some later date which might mean that a book was not a good match at one point but later became a better match for the market, or of course that different publishers had different ideas about what might be a good match.

The examples included here are derived from a very much longer list that can be found in (Bernard 1990) and directly relate the decision to reject to demand from readers:

- John Gallsworthy's book "A Man of Property" from "The Forsyte Saga" was rejected on the grounds that "The author writes to please himself rather than to please the novel reading public and accordingly his novel lacks popular qualities" and that the book "would have no real sale in this country".

- Simon Brett recalled the following rejection: “I’m afraid the current state of the fiction market is too depressing for me to offer you any hope for this”: this could mean that literally the market demanded depressing books but more likely it is a statement that the publisher felt that demand in the market offered no hope to Brett whose work was not a good match. Either way it supports the argument in section 3.

- Harlan Ellison recalls having a piece rejected by Playboy magazine because, while the story was “a knockout piece of writing” it did not match the philosophy of action of the “young urban male readership”.

- Laurence J. Peter’s book “The Peter Principle: Why Things Always Go Wrong” was rejected by McGraw-Hill in 1964 with the following words: “I can foresee no commercial possibilities for such a book and consequently can offer no encouragement”.

- Stephen King remarks that he sent three chapters of a book to a publisher before he had published anything else and the rejection informed him that “We are not interested in science fiction which deals with negative utopias, they do not sell”.

- Thomas Hardy’s book “Tess of the D’Urbervilles” was rejected on the grounds that the readership might be concerned by “improper explicitness”.

- Sherwood Anderson’s book “Winesburg, Ohio” was rejected on directly on the grounds that readers might find it “far too gloomy”.

- George Moore was told about his book “Esther Waters” that it would “hardly go down here” because of certain scenes (such as childbirth) that might upset the potential readers.

- Herman Melville was told that “Moby Dick” would be “unsuitable for the Juvenille Market in [England]”.

- Laurence Wylie's chronicle of French country life "A Village in the Vaucluse" was rejected on the grounds that "It is so far from being a book for the general reader that nothing can be done about it".

- Barbara Pym was told after submitting her novel "An Unsuitable Attachment": "Novels like (this), despite their qualities, are getting increasingly difficult to sell." Barbara Pym was also told of her novel "The Sweet Dove Died" that is was "Not the kind of thing to which people are turning."