Hills, Thomas; Proto, Eugenio; Sgroi, Daniel

Working Paper

Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books

CESifo Working Paper, No. 5906

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

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
MAY 2016

ISSN 2364-1428

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, big data, Google Books, GDP, conflict, Easterlin paradox.

Thomas Hills
Department of Psychology
University of Warwick / UK
t.t.hills@warwick.ac.uk

Eugenio Proto
Department of Economics
University of Warwick / UK
e.proto@warwick.ac.uk

Daniel Sgroi
Department of Economics
University of Warwick / UK
daniel.sgroi@warwick.ac.uk

May 6, 2016

The authors thank several colleagues for discussions on this and related research, especially Sean Allen, Sasha Becker, Thijs Steve Broadberry, Nick Crafts, Andrew Oswald, Luigi Pascali, Giovanni Ricco, Jeremy Smith, Thijs Van Rens. 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.

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*.

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]

With this backdrop our primary objective is to produce a workable proxy for subjective wellbeing going back to 1776, which would enable direct comparisons with GDP over that

[1]For some recent examples see Di Tella, MacCulloch, and Oswald (2001), Deaton (2008), Stevenson and Wolfers (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 multiple 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 wages several hundred years ago there is a deeper issue of what people might buy with that money and hence on how to 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 unrecognisable. This of course poses serious 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 the challenges in more detail and how we can address them.

period (or at least post-1820 when data is readily available for all the countries in our sample) and assess the effect of the improvement in life expectancy, together with the evolution of child mortality, the conflicts and civil wars that have characterised the West in the last two centuries, and the rise of pro-active macroeconomic fiscal and monetary policies.

An initial issue is how we can go about extending existing subjective wellbeing measures when direct survey evidence was only initiated in the 1970s. 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 corpus, 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, and perhaps most importantly, 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, 1776 is consistent with the existing literature on the quantitative use of big data (Michel et al., 2011; Greenfield, 2013).[6]

Our methods rely on making inferences about public mood from large collections ("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 political candidates (Connor, Balasubramanyan, Routledge, & Smith, 2010), predicting stock market trends (Bollen, Mao, & Zeng, 2011), understanding diurnal and seasonal mood variation (Golder & Macy, 2011), detecting the social spread of collective emotions (Chmiel et al., 2011), and understanding the impact of events with the potential for large-scale societal impact such as celebrity deaths, earthquakes, and economic bailouts (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011; Thelwall, Buckley, & Paltoglou, 2011). The approach we take here relies on affective word norms to derive sentiment from text (commonly called "valence" within psychology). In a study of 17 million blog posts, Nguyen, Phung, Adams, Tran, and Venkatesh (2010) found that a simple calculation based on the weighted

[5]For instance consider Kant's *Critique of Pure Reason* which 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.

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 & Danforth, 2010; Dodds et al., 2011).

There is more detail on the data-collection in the next section, but here we give a simple idea of how the method works. First a large 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 countries and we consider the USA, UK, Germany, France, Italy and Spain. This can be straightforward (words like "happy" and "good") but can also be quite abstract (words like "abreast" for instance). In each case the words are measured and rated and a metric that links words, books and entire corpora of language can be constructed. Using this list we can then work through hundreds of thousands of books enumerating the complete published list on Google books by year and by country/language. There will be complications: the UK and USA have distinct publishing traditions and can be separated, but for Spain the use of the Spanish language in Latin America produces some difficulties which leads us to be cautious when examining Spanish data.[7] Added to this is the complication that language is not fixed but is very-much evolving (Hills & Adelman, 2015), just as culture, technology, health and education have changed over time, providing similar difficulties for historical GDP measures.[8]

We check the veracity of our findings in two ways: the first and most straightforward is by analyzing survey-based wellbeing data, 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. First and most simply this is 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 life satisfaction. 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 where they are available, we can proceed to push back further

[7] We argue later that the French language has similar if lesser issues, most likely relating to use in France and northern Africa.

[8] 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 and Gardner (2015). See also footnote 4.

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 the best of our knowledge, available for the longest period of time: life expectancy at birth and per capita GDP. We also add conflicts, democracy, concentration of education and infant mortality (for which long time series data are readily available) to shed some light on the mechanism through which our two explanatory variables, GDP and life expectancy, affect our estimated measure of subjective wellbeing.

This last method provides us with important insights into economic history, but its reliability can be potentially affected by changes in written language and literary styles during the 200 years we are considering. This may generate a bias if the changes have affected our measure of average valence in a systematic way not associated with subjective well-being. We cope with this issue econometrically in two ways. The simplest is to add year fixed-effects. This method is valid under the hypotheses that systematic changes in the valence of languages of written books are common across the countries we are considering. The second method is by detrending 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 trend effects on 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.

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 analysing 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 source was the Google Books corpus (Lin et al., 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 (Michel et al., 2011). The way they have been selected is not clear, this might be due to the level of availability of the books implying that there may be a bias toward more famous and successful publications. In order to assess the valence of individual words, we used the largest available sets of existing word valence rating norms for each language. We analysed data for 6 languages: English (British), English (American), German, Italian, Spanish, and French.

Valence, is a widely used concept in psychology and can be 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. Of course this is not always the case for any individual user or statement, but any bias should be eliminated since we have little reason to believe such bias will change systematically over time. Hence, any analysis that makes use of the words contained in over 8 million books should easily meet the threshold for the use of the law of large numbers. Word valence rating norms generally 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 & Lang, 1999). This list served as the basis for developing valence ratings for multiple languages through several independent studies.[9]

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, & 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. The 14 thousand words in the database contain a subset of the

[9]The Google Ngrams corpus also includes books in Russian, Chinese and Hebrew but to the best of our knowledge valence has not been calculated for words in these languages.

1034 ANEW words. For German, we used the affective norms for German sentiment terms (Schmidtke, Schröder, Jacobs, & 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, & 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 & Syssau, 2014) and Spanish (Redondo, Fraga, Padrón, & 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 exclusively used the mean valence rating of words.[10]

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. 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 (Greenfield, 2013; Michel et al., 2011). After 2000 there was a change in the book sampling method (Greenfield, 2013) something we note from figure 2, where we observe a drop on the number of words used (especially for Spanish, French and Italian) and in our analysis below we will test the robustness of our results by the exclusion of the years 2000-2009.

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%,

[10] A very reasonable point to make is that our methods do not allow "negation" to be considered. 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 not a problem 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.

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 the word covered is close to around 1% as well, the percentage of words covered does not stabilize, suggesting that the written language in the books went through some evolution with respect to the words considered. This could be a reflection of the fact that literature in Spanish and French encompass a large number of highly heterogeneous countries. For example, it is hard to disentangle Spanish used by Spaniards or by natives of South American countries where Spanish may be the first language, and a similar issue exists for France and certain countries in northern Africa and elsewhere. Given that this can potentially be a source of bias, in all our regressions we will control for words covered, and we will always check the robustness of our results to the exclusion of France and Spain.

2.2 Subjective Wellbeing Indices

We first check the validity of our valence-based measure by comparing it with existing survey-based measures of subjective wellbeing. The life satisfaction we take as the ground truth is the simple 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, to the best of our knowledge, 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). We find that our measure correlates positively with Eurobarometer life satisfaction, as indicated in figure 4 and table 2 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.

2.3 Explanatory variables of the historical analysis

In the next sections, we compare our new measure of estimated wellbeing with the two welfare indicators for which, to the best of our knowledge, the longest series of data are available; 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.[11] The other main explanatory variable is the historical data on life expectancy at birth from the OECD, available from 1820 onwards (van Zanden et al., 2014)

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 an index of freedom, from the OECD data available from 1820 onwards (van Zanden et al., 2014), and infant mortality from the International Historical Statistics 1750-2010 (Ltd, 2013) to better investigate the channel from life expectancy to wellbeing. Table 1 summarises the data.

3 Analysis

We ran three different analyses, 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 analyse how this measure correlates with indicators of wellbeing for which historical times-series data are available.

3.1 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

[11] The results of the next analysis would quantitatively change very little if we used Maddison Dataset instead of Penn.

the country of origin of the language in the corresponding year for the period over which survey data is available on life satisfaction. For all languages we can observe a positive correlation. 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 2 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 2 we note a positive and highly significant relationship between valence and life satisfaction, the coefficient is quite large in magnitude, to an increase of a unit standard deviation of valence correspond a 1.5 standard deviation unit increase in life satisfaction. This relation holds even after excluding the years after 2000 (column 2), the year Google changed the measurement of the word frequency; and after excluding Spain and France (column 3), the two languages that feature clear changes in the 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 given 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). 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 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 2. However, It is highly unlikely that this is the case, since as we saw in column 3 of table 2 the positive and significant correlation between valence and life satisfaction would be even stronger if we exclude Spain from the regression. For Italy, instead 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 $Z(t) = 0.0011$).[12] 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, 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.

3.2 Correlations between Words and Average Life satisfaction

In this section, we conduct a non-parametric analysis that complements the conventional regression analysis described in the previous section. First, we 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 correlation is significant at the usual threshold of the 0.05% level and calculate 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 correlates 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 all indicates that valence is aligned

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

with reported life satisfaction over the period for which both are available.

In figure **??** we have the average valences of positively and negatively correlated words for all languages pooled together and separated. Considering every country separately, these differences are highly significant for the UK and Italy, weakly significant for Germany and non significant for Spain and France. The result for Spain and France confirm that for these languages our measure of wellbeing tends to be noisier. Nonetheless, the results indicate that the correlation between valence and life satisfaction that observe above for data aggregated over multiple words at the same is also visible at the level of individual words, but to a lesser extent.

3.3 Welfare and Estimated Subjective Wellbeing

Henceforth, we will link the language to its country of origin, hence 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: US, 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.[13] 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 (1915-18) and World War II (1938-45), in the 5 European countries we also added 1848, a year typified by a number of revolutions.[14] For US English we observe a sharp 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

[13]This is the last year available from Google, however after 2000 there have been changes in the way data were collected. In what follows we check the robustness of our results to the exclusion of the latest 8 years.

[14]Moreover, in the US, 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 press which these countries experienced during World War II. We will come back below to the issue of the freedom of press when we present our econometric analysis.

We notice that our estimated subjective wellbeing measure for all countries but Spain and France do not seem to feature a systematic increasing or decreasing trend, this 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, & 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 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.

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

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

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 3; 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. 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 seems to be the one that best fits the data.

In model 2 we estimate the equation:

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

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.8 of the appendix we present a robustness check with different smoothing parameters.[15] As said, this model addresses the possibility of bias generated by systematic trend, 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-terms 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 a more modern vintage

[15] 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.

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 3, 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, its magnitude 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. 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* (Easterlin, 1974; Easterlin et al., 2010). 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, in tables A.2 and A.4 respectively.

Considering now 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 4 we analyse 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 4 (in table 3 the coefficients are similar but not significant). This perhaps suggest 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 effect seems to lose significance 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, in tables A.5 and A.6 respectively.

Stochastic Trends

In the analysis in table 4, 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).[16] 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. Finally, we 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.

4 Concluding Comments

We have produced 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 over 200 years to existing survey-based wellbeing measures which date back to 1973 at best. In order to do this we formed a measure of positive valence from many millions of books digitized in the Google Books corpus. 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 society prefers the mood of books to match the mood of the population.

We caution against thoughtless long-term interpretation of these data since both the market for books, and language itself, have evolved considerably over the period we consider (e.g., Hills & 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 great.[17] A key issue in particular with the limited survey-based data that we

[16] We omit the details of the tests, which are available upon demand.

[17] Consider for instance the arrival of urbanization, huge cultural and political shifts, increased technological

have is our current inability to consider the impact on 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 wellbeing, GDP and life expectancy. We presented two models that seek to make longer-run comparison more feasible and provide an early attempt to use our index to study subjective wellbeing in the USA, UK, Germany, France, Italy and Spain from 1776 to 2009.

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 beyond what we have done in this paper. 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 historic attitudes and (economic) behavior. Certainly the methods employed in our work would be equally applicable to other socio-economic variables aside from subjective wellbeing, such as attitudes towards policy, trust, and interest. 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

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

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

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

Broadberry, S., Campbell, B., Klein, A., Overton, M., & Van Leeuwen, B. (2012). *British economic growth, 1270-1870: an output-based approach* (Tech. Rep.). School of Economics Discussion Papers.

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.

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

Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (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., MacCulloch, R. J., & Oswald, A. J. (2001). Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review*, 335–341.

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

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (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., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences*, *107*(52), 22463–22468.

Golder, S. A., & Macy, M. W. (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., & Adelman, J. S. (2015). Recent evolution of learnability in American English from 1800 to 2000. *Cognition*, *143*, 87–92.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar us business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 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.* Norwood, NJ: Ablex.

Lin, Y., Michel, J.-B., Aiden, E. L., Orwant, J., Brockman, W., & Petrov, S. (2012).

Syntactic annotations for the google books ngram corpus. In *Proceedings of the acl 2012 system demonstrations* (pp. 169–174).

Ltd, P. M. (Ed.). (2013). International historical statistics. , 0. Retrieved from `http://dx.doi.org/10.1057/9781137305688` doi: 10.1057/9781137305688

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... others (2011). Quantitative analysis of culture using millions of digitized books. *science*, *331*(6014), 176–182.

Monnier, C., & Syssau, A. (2014). Affective norms for French words (fan). *Behavior research methods*, *46*(4), 1128–1137.

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior research methods*, *46*(3), 887–903.

Nguyen, T., Phung, D., Adams, B., Tran, T., & Venkatesh, S. (2010). Classification and pattern discovery of mood in weblogs. In *Advances in knowledge discovery and data mining* (pp. 283–290). Springer.

Proto, E., & Rustichini, A. (2013). A reassessment of the relationship between gdp and life satisfaction. *PloS one*, *8*(11), e79358.

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (affective norms for English words). *Behavior research methods*, *39*(3), 600–605.

Schmidtke, D. S., Schröder, T., Jacobs, A. M., & Conrad, M. (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., & Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the easterlin paradox. *Brookings Papers on Economic Activity*, 1–87.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, *62*(2), 406–418.

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

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

Figures and Tables

Table 1: Main Variables

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

Figure 1: **A Sample of Word Valence in Different Languages.**

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 2: **The Number of Words and Share of Words Covered.** The red line represents the share of words covered over the total, the blue line represents the total number of words, for all countries considered in the analysis.

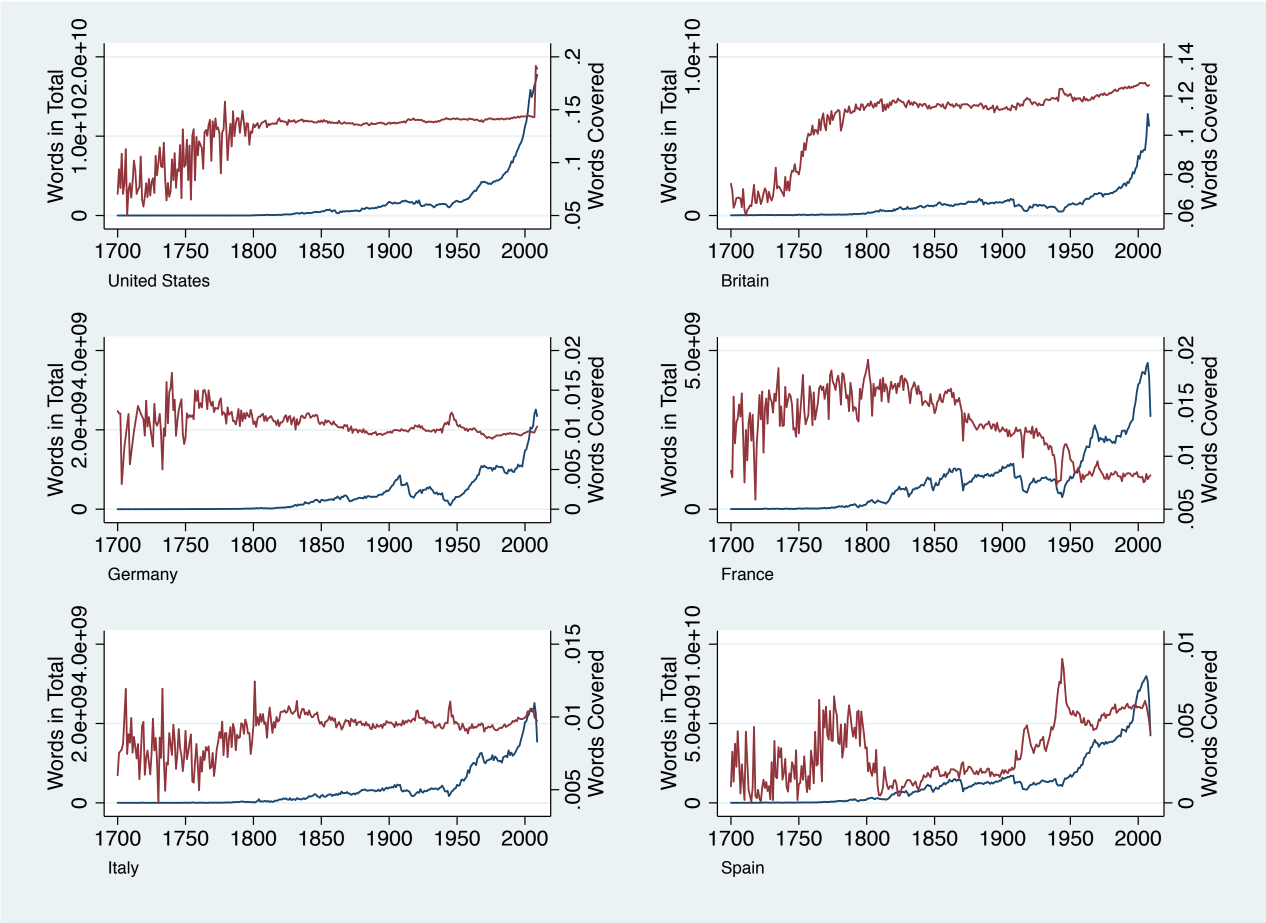

Figure 3: **Average valence and correlations with life satisfaction** The bars in the panel represent the average valence of the words featuring levels of correlations with the corresponding levels of average life satisfaction significantly different from 0 (at the 5% level). The bars represent the 95 % confidence interval.

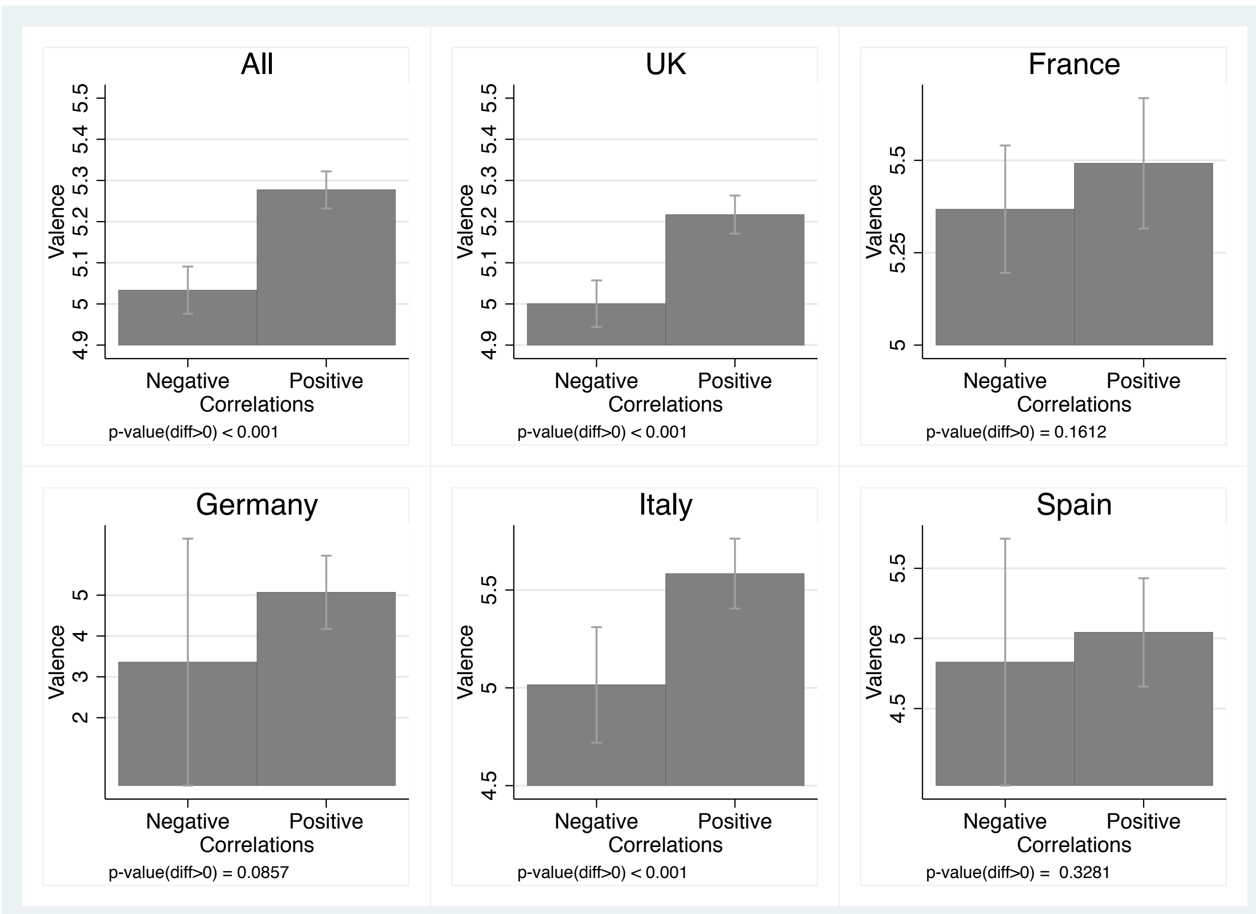

Figure 4: **Valence and Aggregate Life Satisfaction.** In the first 5 panels presenting the data in time-series, Valence is represented in red (values in the left axis), life satisfaction is represented in blue (values in the right axis). In the last panel, we plotted valence against life satisfaction 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.

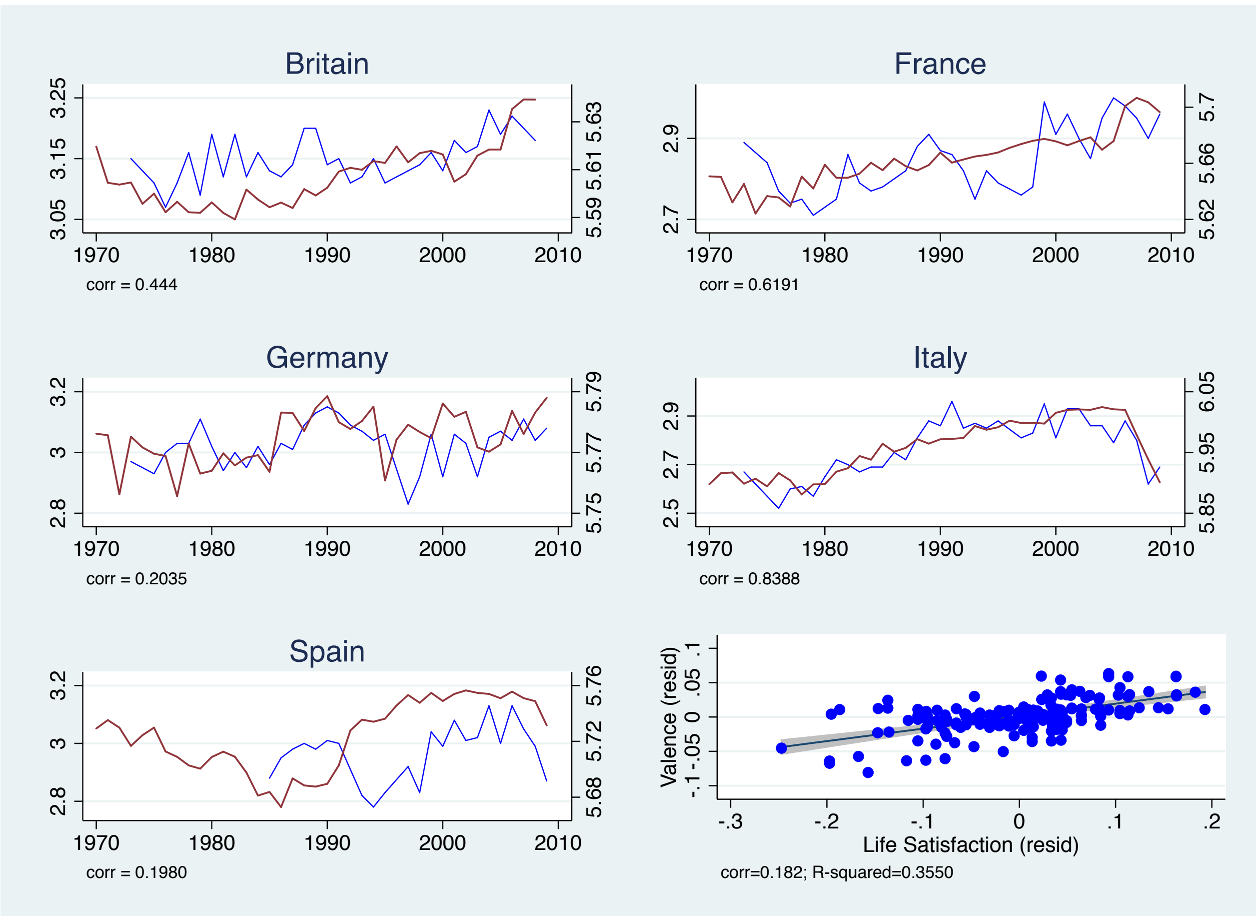

Figure 5: **Estimated Subjective Wellbeing Over the Period 1776-2009** For all countries the vertical red lines correspond to 1789, the year of the French Revolution, to World War I (1915-18) and to World War II (1938-45). In the 5 European countries a line is drawn in 1848, the "Year of Revolution". 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.

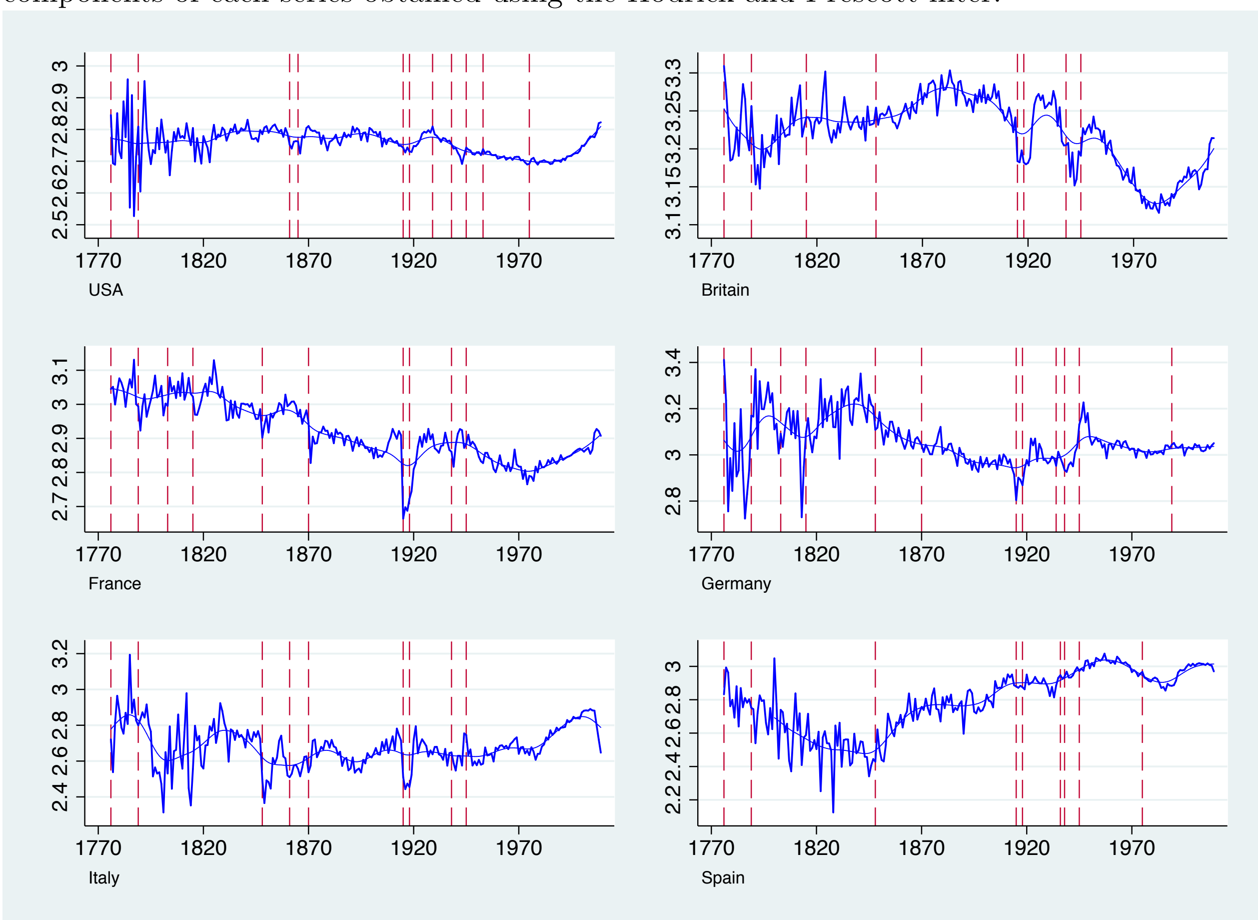

Figure 6: **Detrended Estimated Subjective Wellbeing and Life Expectancy**. The red line represents life expectancy at time $t-1$, the blue line represents estimated subjective wellbeing at time t. Both series have been detrended by using the Hodrick and Prescot filter

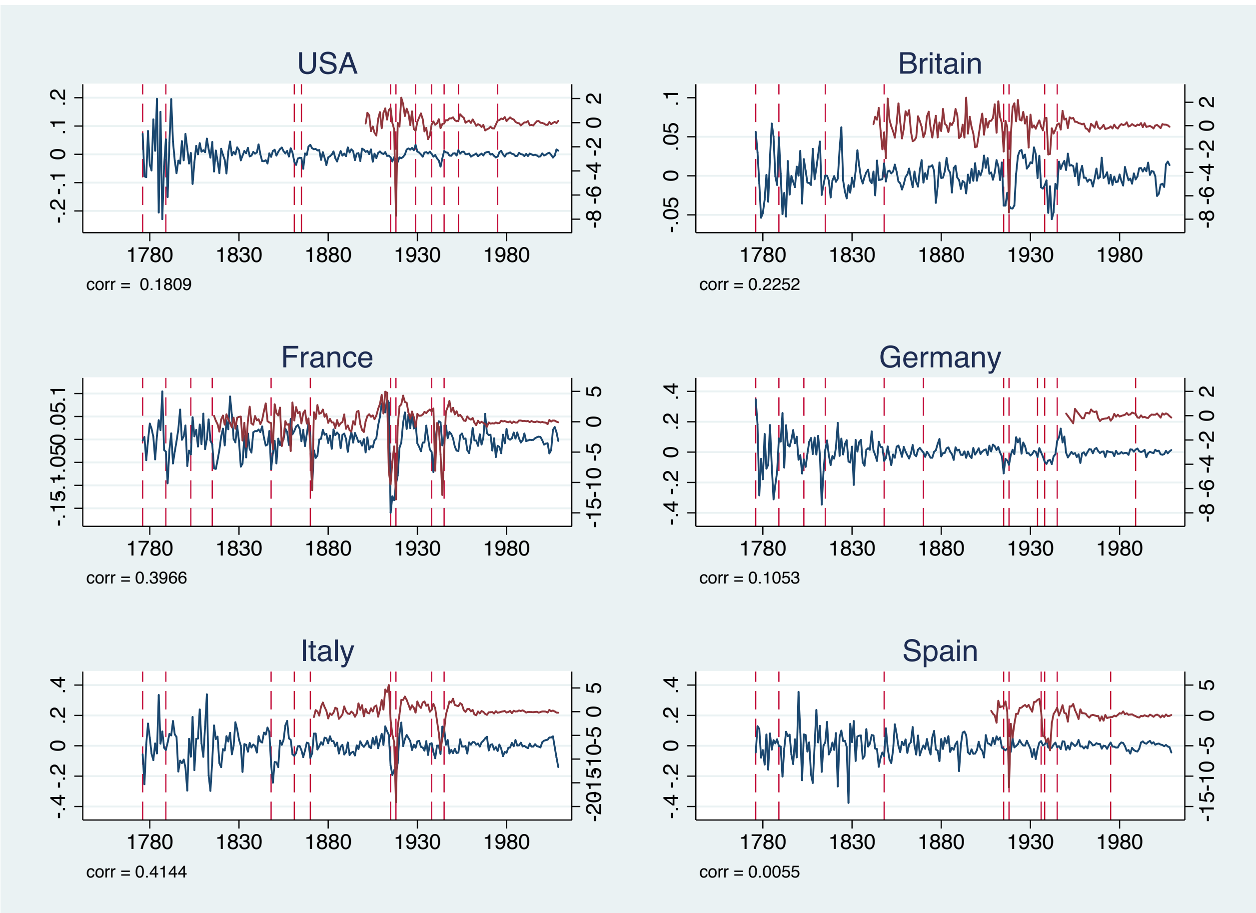

Table 2: **Valence Predicts Aggregate Life Satisfaction** Average life satisfaction per country and year from the Eurobarometer dataset is the dependent variable. In columns 1, 4, 5 and 6 the years are 1983-2009, in column 3, 1972-2009, and in column 2, 1983-2000. The countries in columns 1,2,4,5 and 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 Until 2000 b/se	3 w/o Sp.and Fr. b/se	4 Trends b/se	5 +GDP b/se	6 Year FE 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 3: **The Historical Determinant of Estimated Subjective Wellbeing** The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust Standard Errors clustered at country levels 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	0.0089***	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 4: **The Historical Determinant of Estimated Subjective Wellbeing, with Detrended Data** Data have been detrended using the Hodrick and Prescott filter. The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Standard Errors clustered at country levels 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

Appendix: For Online Publication

Different lags of the regressors

Lets more generally consider the expression

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

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 in 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 the estimations corresponding to the above models, we 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.

Table A.1: **The Historical Determinants of the Estimated Subjective Wellbeing, using Different Time Lags in the Regressors** The countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors clustered at country levels are given in brackets. * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	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.2: **The Historical Determinant of Estimated Subjective Wellbeing, excluding Spain and France** The Countries are: Germany, Italy, the UK, and the United States. Robust standard errors clustered at country levels 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.3: **The Historical Determinant of Estimated Subjective Wellbeing, with Errors Clustered at Year Levels** The Countries are: Germany, Italy, the UK, and the United States. Robust standard errors clustered at year levels 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	0.0089***	0.0084***	0.0045***	0.0045***
	(0.0010)	(0.0010)	(0.0015)	(0.0014)
GDP (log) t-1	0.0121	0.0396	0.0252	–0.0047
	(0.0177)	(0.0250)	(0.0259)	(0.0214)
Infant Mortality t-1			–0.0012***	–0.0011***
			(0.0003)	(0.0003)
Internal Conflict$^-$ t-1				–0.0061
				(0.0074)
External Conflict$^-$ t-1				–0.0070
				(0.0066)
WW1 t-1				–0.0512***
				(0.0126)
WW2 t-1				–0.0295**
				(0.0126)
Democracy		–0.0020***	–0.0023***	–0.0029***
		(0.0008)	(0.0008)	(0.0007)
Education Inequality		–0.0006**	–0.0007**	–0.0014***
		(0.0003)	(0.0003)	(0.0003)
Words Covered		–1.1861	–1.0702	–8.8518***
		(1.6367)	(1.6659)	(1.6856)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
N	781	657	639	592

Table A.4: **The Historical Determinant of Estimated Subjective Wellbeing, 1800-2000** The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors clustered at country levels 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.5: **The Historical Determinant of Estimated Subjective Wellbeing, w/o Spain and France and Data Detrended** Data detrended using the Hodrick and Prescott Filter. The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors clustered at country levels 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	0.0071**	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.6: **The Historical Determinant of Estimated Subjective Wellbeing, 1800-2000 and Data Detrended** Data detrended Using the Hodrick and Prescott filter. The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors clustered at country levels 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.7: **The Historical Determinant of Estimated Subjective Wellbeing, with Errors Clustered at Year Levels and Data Detrended** Data detrended using the Hodrick and Prescott filter. The Countries are: Germany, Italy, the UK, and the United States. Robust standard errors clustered at year levels are given in brackets. * $p\text{−}value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	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***
	(0.0013)	(0.0013)	(0.0014)	(0.0011)
GDP (log) t-1	–0.0586***	–0.0657***	–0.0724***	–0.0525***
	(0.0204)	(0.0195)	(0.0188)	(0.0175)
Infant Mortality t-1			–0.0004	–0.0002
			(0.0003)	(0.0003)
Internal Conflict^{-} t-1				–0.0012
				(0.0033)
External Conflict$^{-}t - 1$				0.0018
				(0.0028)
WW1 t-1				–0.0746***
				(0.0115)
WW2 t-1				0.0077
				(0.0086)
Democracy		–0.0026***	–0.0027***	–0.0021***
		(0.0007)	(0.0007)	(0.0007)
Education Inequality		0.0005	–0.0003	–0.0000
		(0.0013)	(0.0022)	(0.0020)
Words Covered		0.6402***	0.7196***	3.1986**
		(0.2234)	(0.2172)	(1.5640)
Country FE	Yes	Yes	Yes	Yes
N	765	648	633	586

Table A.8: **The Historical Determinant of Estimated Subjective Wellbeing with Different Smoothing Parameters** Data have been detrended using the Hodrick and Prescott filter. The Countries are: France, Germany, Italy, Spain, the UK, and the United States. Robust standard errors clustered at country levels 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