THE UNIVERSITY OF WARWICK

July 2015 No.236

Historical Analysis of National Subjective Wellbeing using millions of Digitized Books

Thomas Hills, Eugenio Proto & Daniel Sgroi

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE Competitive Advantage in the Global Economy

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Historical Analysis of National Subjective Wellbeing using Millions of Digitized Books

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

[a]Department of Psychology, University of Warwick

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

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

July 9, 2015

Abstract: We present the first attempt to construct a long-run historical measure of subjective wellbeing using language corpora derived from millions of digitized books. While existing measures of subjective wellbeing go back to at most the 1970s, we can go back at least 200 years further using our methods. We analyse data for six countries (the USA, UK, Germany, France, Italy and Spain). To highlight some results, we find a positive short-run effect for GDP and life expectancy on subjective wellbeing. An increase of 1% life expectancy is equivalent to more than 5% increase in yearly GDP. One year of internal conflict costs the equivalent of a 50% drop in GDP per year in terms of subjective wellbeing. Public debt, on the other hand, has a short-run positive effect. Our estimated index of subjective wellbeing generally does not feature any positive trend, which is consistent with the Easterlin paradox, although we caution against long term analysis given the historical variation of written texts (which parallel similar issues with historical GDP statistics).

JEL classification: N3, N4, O1, D6

Keywords: Historical Subjective Wellbeing, Big Data, Google Books, GDP, Conflict

Acknowledgements: The authors thank several colleagues for discussions on this and related research, especially Sean Allen, Steve Broadberry, Nick Crafts and Andrew Oswald. We thank CAGE (The Center for Competitive Advantage in the Global Economy) for generous funding and Tomas Engelthaler 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 a *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 period (or at least post-1820 when data is readily available for all the countries in our sample) and assess the effect of the buoyant improvement in life expectancy, together with the conflicts and civil wars that deeply characterised the West in the last two centuries, and the rise of pro-active macroeconomic fiscal and monetary policies.

How we can extend existing subjective wellbeing measures when direct survey evidence

[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.

was only initiated in the 1970s? Our methods rely on 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] Third, 1776 is sensible given the relative dearth of published literature before the end of the eighteenth century, and even where there exists a body of literature we might worry about the intended market or the representativeness of authors (consider for instance, with some famous exceptions, the paucity of female writers before the nineteenth century). A final point to make is that we have access to a wealth of other comparable data going back to 1820 such as GDP itself, which we will use in the analysis below. This does mean that when we wish to directly compare GDP with our measure of wellbeing we cannot go back further than 1820.

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. In a study of 17 million blog posts, Nguyen, Phung, Adams, Tran, and Venkatesh (2010)

[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.

found that a simple calculation based on the weighted affective ratings of words was highly effective (70% accuracy) at predicting the mood of blogs as compared to 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 we can give a simple idea of how the method works here. First a large group of individuals are asked to rate a list of words on how those words make them feel. 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 even more 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]

Crucially we can check the veracity of our findings by analyzing survey-based data, which goes back to the 1970s. As we demonstrate below, we find a remarkable degree of similarity between our valence-based measure and survey measures. This lends some confidence when rolling back our own measure further in time. We also see that major historical negative events (wars, depressions, etc.) are clearly visible as sharp declines in our subjective wellbeing measure across all the countries we examine. Our econometric analysis shows how our measure varies with other important socio-economic variables such as income, life expectancy, political freedom, and macroeconomic policy-related variables such as government debts and inflation rates. We complete our study by considering many uses of the new data that we have gathered, and discussing how else our novel methods might be applied within economics. Finally it is worth noting that there is no reason why this 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.

[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).

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 historic language use. The availability of large copora allowed us to do so for six languages. 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). 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 as a subset of their words a list 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 2, we present a sample of the words covered in all 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 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

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

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). At the same time results after 2000 are not comparable with those before 2000 due to the change of book sampling method (Greenfield, 2013). We will nevertheless show that all our results are robust to the introduction of data after 2000. Table 1 summarizes the valence data from 1800 to 2009 together with the other main variables considered. 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 above 12% at around 1800. Also for German and Italian, the percentage of covered words seems to stabilize after 1800, although this percentage is about 1%, which is consistent with the fact that the number of words covered in Italian and German is 10 times smaller compared to both US and British English. For French and Spanish the percentage of words

[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 as our analysis is scale-free 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.

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 the literature in Spanish and French reflects 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.

In figure 3 we show the valences of the 6 languages. 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. We started from 1776, the year of independence and we considered the UK, US, Spain and France until year 2000, 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 the 1848, the year of the revolutions.[11] 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. 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 press when we present our econometric analysis. Differently from the other languages, Spanish does not feature a drop in World War I or 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.

We notice that our valence 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

[11] 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)

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).

However, longterm 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 within literature. The fact that literature portrayed 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. Fortunately, the complexity of the market for books can be controlled econometrically to some extent (though this means at least for long-run temporal comparisons we lose the ability to rely solely on simple plots).

2.2 Subjective Wellbeing Indices

The level of life satisfaction is constructed using Eurobarometer data as the simple average per year and country of all individuals surveyed. The question is "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 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).[12]

2.3 Explanatory variables

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

[12]Data from the US 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 run our analysis by including this measure from the GSS as well, obtaining very similar results which are available upon request.

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 explanatory variables are historical data from the OECD, available from 1820 onwards (Zanden et al., 2014). As a general rule, we selected all the existing variables with yearly coverage that are traditionally used in subjective wellbeing literature; they are life expectancy at birth (as an index of health), internal conflict, external conflict, education inequality (measured as a GINI index), the index of democracy (originally, from the Polity IV project; as an index of freedom), total gross central government debt as a percentage of GDP (as an index of public expenditures), and inflation rates.[14]

3 Econometric Analysis

We ran two 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 we analyse the historical determinants of valence.

3.1 Valence and Aggregate Life Satisfaction

In figure 4 we present the relationship between the valence of each language and year and the aggregate life satisfaction of the country of origin of the language in the corresponding year for the period over which survey data is available on life satisfaction. 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.

In table 2 we show that this positive relationship holds even after controlling for year fixed effects and words covered (column 1), per capita GDP (column 2), after restricting the analysis to data before 2000 (columns 1,2) or considering the entire set of data available (column 3). We also present results excluding Spanish and French data (column 4). In this last case, the relationship appears even stronger as one would expect given our previous point that literature in French and Spanish derive from a heterogenous set of countries, so it would be reasonable to argue that the link with France and Spain is weaker and noisier.

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

[14]Ideally, it would have been worthwhile including measures of average education and wealth concentration as well, but they are not available on a yearly basis in the OECD dataset.

3.2 Historical Determinants of Subjective Wellbeing

As discussed above, 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. We deal with this problem in two 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) 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 the year fixed effect should correct any source of bias; ii) we relax the assumption that each country faced the same evolution in markets for books and language more generally, but instead assume that all countries are characterised by a different evolution in the written texts which yields linear trends in the valence of the words from the different languages.

We aim to provide an estimation of the historical evolution of the life satisfaction, accordingly we use predicted life satisfaction as a dependent variable. This is obtained as the estimated linear prediction, $\hat{Sat}_{i,t}$, from the simple model $Sat_{i,t} = a_i + bVal_{i,t}$, where a_i is the country-specific coefficient. $Sat_{i,t}$ is the average life satisfaction in country i in the year t, and $Val_{i,t}$ is the valence at year t, of the language corresponding to country i as defined in 1. Henceforth we will refer to our estimated index $\hat{Sat}_{i,t}$ as "estimated life satisfaction".

Accordingly, 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 countries considered (an assumption we relax for the second model that follows):

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

where i and t denote respectively the country and the year, α_i and η_t denote the country and year-specific effects, respectively. $wc_{i,t}$ measures the word coverage at time t in the language corresponding to country i. The variables x_z are the Z determinants of wellbeing discussed earlier, and listed in the different models of table 3.

For our second model we make use of a country-specific trend, which (as indicated at the end of section 2) relaxes the assumption of the same evolution of language across countries, but instead imposes a linear trend. Hence:

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

so that the term $\delta_i t$ is the country specific trend.

3.2.1 Results of Historical Determinants Analysis

In table 3, we present the results of the estimations of the different specifications of model 2 and 3. Non-dichotomic variables have been transformed into logarithms to simplify the interpretations of the results. From Column 1 and 2, we note that both life expectancy and GDP have positive effect on estimated life satisfaction. This last result is consistent with the literature generally showing a positive short-term relationship between GDP and subjective wellbeing (e.g., Stevenson & Wolfers, 2008).) Comparing the coefficients (and recalling that the variables are in logarithmic forms) one can argue that an increase of 1% life expectancy is equivalent to more that 5% increase in yearly GDP. We also note that democracy is negative and significant, supporting the notion that a freer press decreases our valence measure because the press is less constrained to support the current regime (consider for instance the press in Germany or Italy in the period just before and during World War II). In columns 2 and 3 we added a measure of internal and external conflicts. While both are important, an internal conflict costs about 3 times more than an external conflict in terms of estimated life satisfaction. Furthermore comparing the conflict measure to the coefficients on GDP in column 3, one can argue that one year of internal conflict costs (in terms of estimated life satisfaction) the equivalent of a 50% drop in GDP in that year. Finally in columns 5 and 6, we note that coefficients generally become bigger and more significant when we exclude Spain and France. As previously argued, we believe this is a facet of the extra noise implicit in the Spanish and French language data.

In table 4, we can see that national debt has positive effects on estimated life satisfaction, at least in the short-run. Taken literally this indicates that expansionary fiscal policies (for example a reduction in tax rates or increase in public spending) are likely to boost estimated life satisfaction. Notice also that debt seems to mediate the positive effect of GDP, since its inclusion eliminates the significance of GDP-induced changes to subjective wellbeing. It is important to note however that this is a short-term analysis, hence this cannot say anything on the longer-run ramifications of expansionary government policies for the economy. Finally from columns 3 and 4, we note that the inflation rate does not seem to have a significant effect, although the coefficient is negative.

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 the wellbeing measures we had previously which date back to 1973. In order

to do this we formed a measure of positive valence from many millions of books digitized in the Google Books corpus. Our new measure is highly consistent with existing wellbeing measures going back to 1973.

We caution against 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 emphasised 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 with 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 intertemporal comparisons difficult. For the same reason that we would herald longer-run GDP data as important and useful despite these issues, we would similarly point to the many uses of a longer-run measure of wellbeing.

For a first use of this new data we present an analysis of the determinants of wellbeing going back to 1820 (when GDP figures for all the countries we consider are available) finding that GDP and life expectancy have a strong positive effect on our wellbeing measure, while conflict (especially internal) has a negative effect. We also find a positive effect of public government debt, suggesting an increase in welfare following a fiscal expansion. Nonetheless, we would not draw any conclusions concerning the impact on long-run economic development or indirectly through to long-run wellbeing, which is presently beyond the scope of our analyses. We also find that democracy seems to have a negative effect, which we would interpret through control of the press and published literature under dictatorial regimes.

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. 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.

5 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	2.984	0.175	2.52	3.248	190
per capita GDP (Maddison)	6771.196	6362.951	1007.867	31357	984
per capita GDP (Penn)	25064.164	6553.946	13069.197	43511.594	232
Life Expectancy	59.771	14.774	25.81	82.400	798
External Conflict	0.427	0.495	0	1	1206
Internal Conflict	0.111	0.314	0	1	1206
Democracy	3.983	6.548	-10	10	1079
Govern. Debt (in % of GDP)	68.793	52.352	0.003	261.759	885
Inflation	4.212	18.43	-67.605	344.569	1202
Education Inequality	26.964	19.559	6.111	98.935	784
Words Covered	0.049	0.057	0	0.191	1259

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	Abenddämm	-2.35	abonné	4.53	abbandonatc	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	anbetungswi	-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

Table 2: **Valence Predicts Aggregate Life Satisfaction** Average life satisfaction per country and year from the Eurobarometer dataset is the dependent variable. In columns 1,2 and 4, the years are 1973-2000. The Countries are: France, Germany, Italy, Spain, UK. Per Capita GDP (expressed in terms of purchasing power parity) is from the PWT 8.0 dataset. Coefficients are standardized. Robust standard errors are given in brackets. * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	with GDP	until 2009	W/O Spain and France
	b/se	b/se	b/se	b/se
Valence	1.4646***	1.3795***	1.3892***	2.1837***
	(0.3535)	(0.3847)	(0.2483)	(0.3453)
Log GDP		0.1747	0.2186	0.5076
		(0.3102)	(0.2327)	(0.3624)
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Words Covered	Yes	Yes	Yes	Yes
r2	0.903	0.903	0.904	0.953
N	119	119	163	78

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: **The Average Valences Over the Period 1776-2000** 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 draw in 1848, the year of the revolutions. 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)

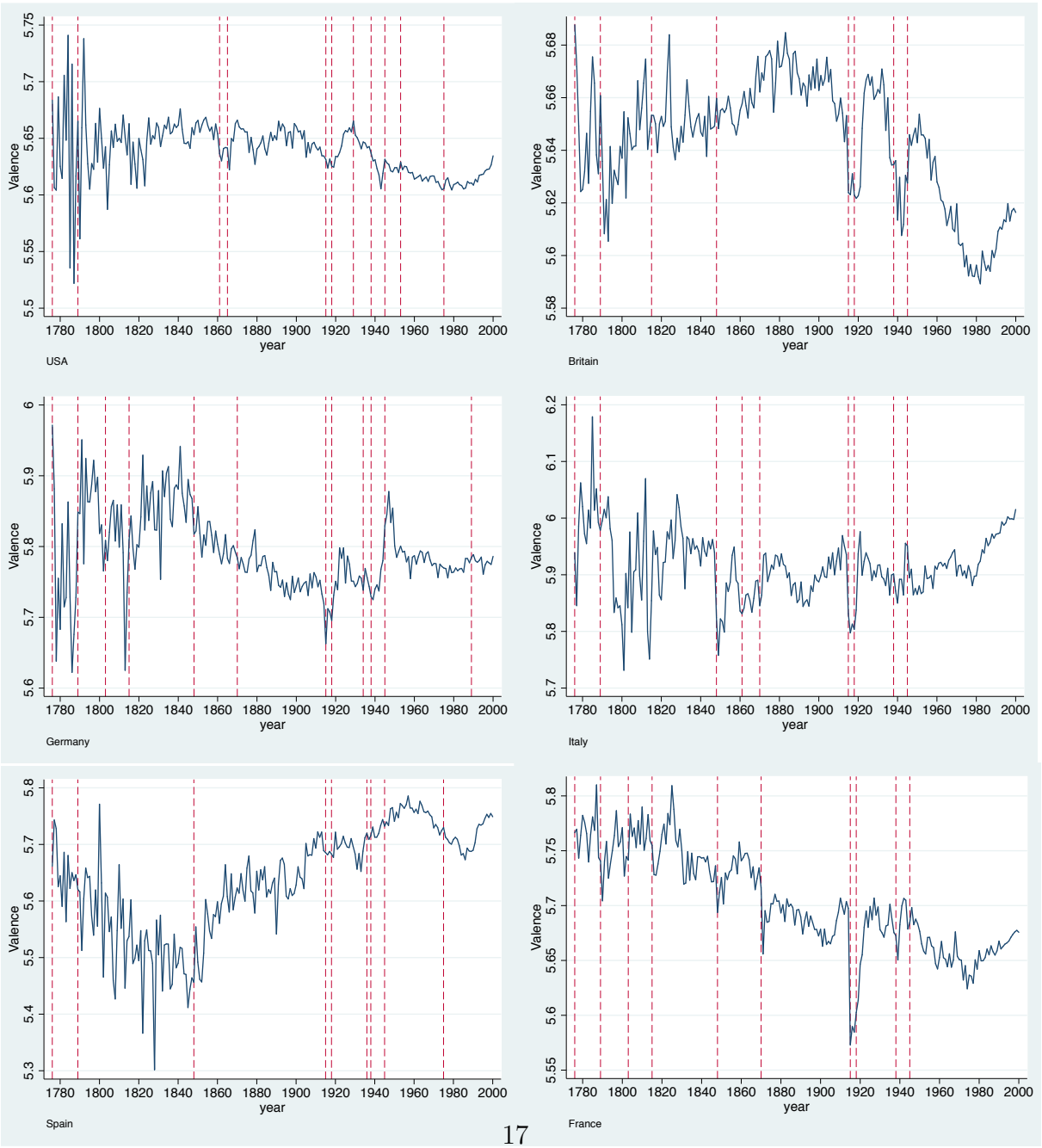

Figure 4: **Valence and Aggregate Life Satisfaction.** The countries are Germany, Italy, Spain, UK in the period 1972-2009. Both variables are expressed in the form of residuals after controlling for country fixed-effects. The grey area represents the 95 % confidence interval of the regression line.

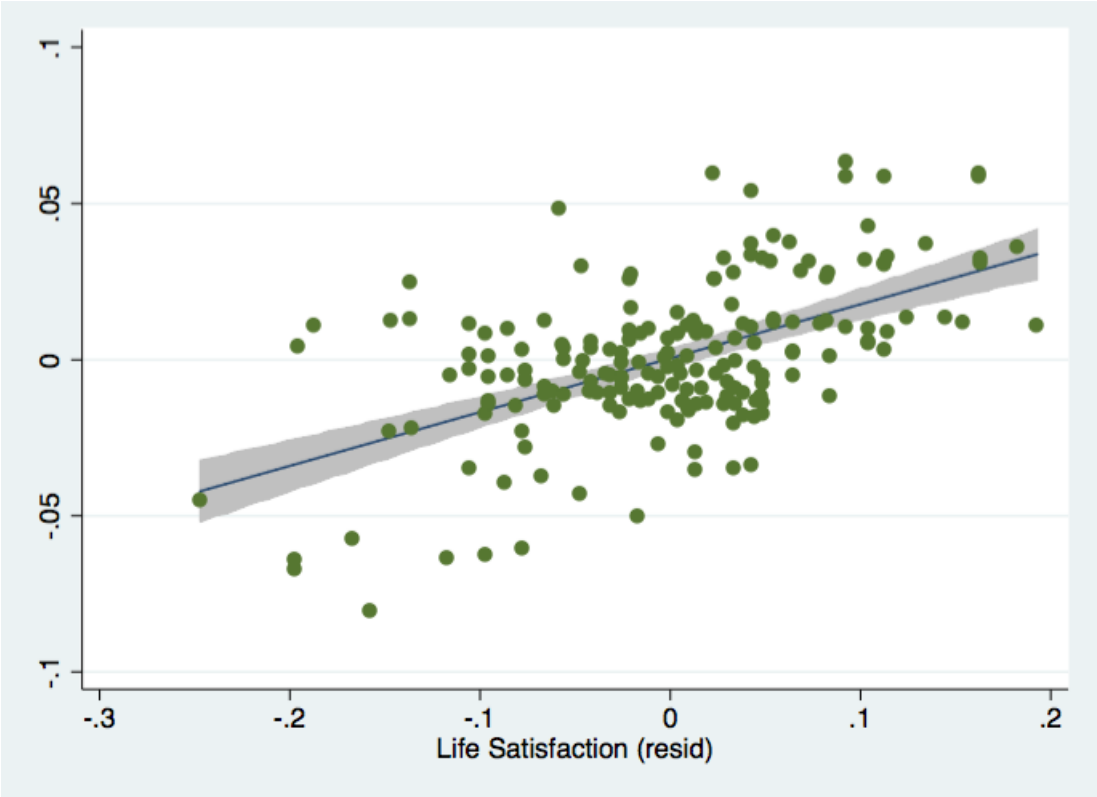

Table 3: **The Historical Determinant of Estimated Subjective Wellbeing** The Countries are: France, Germany, Italy, Spain, UK, United States. Per Capita GDP is from Maddison's dataset. All non-dichotomic variables are in logarithmic form. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

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

Table 4: **The Effect of Public Debt and Inflation on Estimated Subjective Wellbeing** The Countries are: France, Germany, Italy, Spain, UK, United States. Per Capita GDP is from Maddison's dataset. All non-dichotomic variables are in logarithmic form. Robust standard errors are given in brackets. * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

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

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.

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., et al. (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.

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).

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., et al. (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.

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.

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.

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