Preliminary and Incomplete

Growth in Income and Subjective Well-Being Over Time*

Daniel W. Sacks
Wharton, University of Pennsylvania

dansacks@wharton.upenn.edu
sites.google.com/site/sacksdaniel

Betsey Stevenson
University of Michigan
CESifo and NBER

betseys@umich.edu
www.nber.org/~bstevens

Justin Wolfers
University of Michigan
Brookings, CEPR, CESifo, IZA and NBER

jwolfers@nber.org
www.nber.org/~jwolfers

Abstract

Recent research has found that richer countries have higher well-being than poorer countries and that the relationship is similar in magnitude to that seen between rich and poor members within countries. However, limited data have constrained previous researchers' ability to detect whether economic growth within countries leads to greater well-being. Thus the question of whether raising the income of all will raise the well-being of all remains open. We combine newer data from many different sources with historical data to study the relationship between well-being and GDP in a panel and time series context. We find strong evidence that well-being and GDP grow together. This finding holds over both the short and long run. Over recent decades the world has gotten happier, and the magnitude of the gains is similar to what would be predicted by the growth in world GDP. Our findings suggest an important role for economic growth in increasing well-being, and cast doubt on the Easterlin paradox and theories of adaptation.

This draft: October 28, 2013

First draft: 10/5/2011

Keywords: Subjective well-being, life satisfaction, quality of life, economic growth, development, Easterlin Paradox, well-being-income gradient, adaptation.

JEL codes: O11, I31, I32

*Sacks gratefully acknowledges financial support from a National Science Foundation Graduate Research fellowship.

I. Introduction

Research on the relationship between subjective well-being and income has established a few clear facts: (a) Within a country¸ richer people are happier than poorer people; (b) Across countries, people in richer countries are, on average, happier than those in poorer countries; (c) In each of these cases, percentage changes in income have roughly similar effects on well-being (that is, the relationship between well-being and income is linear-log); and (d) These comparisons within a country and between countries each yield similar estimates of the well-being–income gradient.

However, the relationship over time between well-being and GDP remains an open question. Indeed, Easterlin et al (2010) say that "the happiness–income paradox is this: at a point in time both among and within nations, happiness varies directly with income, but over time, happiness does not increase when a country's income increases." Clark, Frijters and Shields (2008) refer to "the 'paradox' of substantial real income growth in Western countries over the last fifty years but without any corresponding rise in reported happiness levels." Relatedly, Layard et al (2009) have argued that "the long run effect of higher average income at the country level is quite small," while DiTella and MacCulloch (2009) suggest that "researchers who detect positive trends in some of the happiness time-series have to face up to the fact that these tend to be small." The implication is that raising average incomes does not raise average well-being much, if at all. The juxtaposition of a robust relationship between well-being and income at a point in time, but little evidence of one over time has been emphasized as a key point in favor of theories emphasizing relative income, adaptation, and other rank- and reference-dependent utility functions.[1]

Our goal in this paper is to return to the question originally posed by Easterlin (1974), who asked: does economic growth improve the human lot? The answer speaks to the appropriate goals for government policy. For example, Easterlin et al (2010) argue that a country cannot improve the lives of its citizens through GDP growth since the "escalation of material aspirations

[1] Easterlin (1974, 1995, 2005) emphasizes relative income comparisons, as do Clark, Frijters and Shields (2008), who examine both income relative to others and to oneself in the past; Di Tella, MacCulloch and Oswald (2003) also emphasize relative income, analyzing family income quartile against average national per-capita income; Blanchflower and Oswald (2004) look at income relative to income in an individual's state; Boyce, Brown and Moore (2010) emphasize income ranks; while Caprole et al (2009) use two different operational definitions for reference income: one containing countrymen of similar age and one containing countrymen of similar age and education.

with economic growth, reflecting the impact of social comparison and hedonic adaptation, are of central importance" (p. 5 Easterlin et al PNAS). Other researchers have argued that at higher levels of income there is adaptation to further economic growth. Oswald (2010) says it most starkly: "GDP is a gravely dated pursuit."

Policymakers influenced by these findings argue for a shift in economic discourse from traditional measures of well-being. For example, French President Sarkozy put together a commission that recommended shifting emphasis from measuring economic production towards more multi-dimensional measures of objective and subjective well-being (Sarkozy, 2010). While many of these recommendations may be relevant even if raising GDP does indeed raise well-being, it is clear that answering this question is of utmost importance to directing policy. Even in the US, where arguably subjective well-being research has had a smaller impact on policy-makers, the current Chairman of the US Federal Reserve appealed to the Easterlin paradox and argued that "economic policymakers should pay attention to family and community cohesion" as well as traditional objectives such as GDP (Bernanke, 2010).

In this paper we test Easterlin's hypothesis by examining the relationship between growth in subjective well-being and income over time. Our previous research has focused particularly on comparing the income–well-being relationship in the cross-section, showing that the relationship observed when comparing rich and poor people is similar to that observed when comparing rich and poor countries (Stevenson and Wolfers 2008, Sacks et al, 2010). In this paper, we focus on the time series dimension. Our contribution is to compile the most thorough analysis of national time series possible. We incorporate data from all of the major cross-national survey efforts, and add to this previously unexploited archival data. Thus, our aim is to compile the most comprehensive such study that is possible.

To preview, our findings show a robust positive relationship between well-being and GDP over time. Our point estimates of the relationship between well-being and GDP over time are similar to that when comparing countries at a point in time. We can reject Easterlin's hypothesis that economic growth is not associated with greater subjective well-being. Countries which have experienced greater economic growth have experienced a larger rise in well-being. Moreover, we find that global well-being has risen since the 1960s, and the magnitude of this rise is consistent with the gain in global output over this period. As average incomes have risen around the world, the subjective well-being of its people has risen.

These findings overturn the status quo in the literature. As such, it is worth sketching out a brief reconciliation of our results. One consistent difference between our work and those of other scholars is that we test the null of the Easterlin paradox directly, comparing the magnitude of the relationship between well-being and GDP over time with that estimated at a point in time within or between countries. With limited and noisy data, previous scholars have often failed to reject the null hypothesis that there is no time series relationship, although this largely reflects the imprecision of estimates based on sparse datasets with limited variation. We show that the primary challenge in analyzing the relationship between growth in subjective well-being and income over time is the very limited variation in economic growth that has occurred within existing datasets. The variation in income both within countries and across countries at a point in time is simply much greater than that observed in our time series. We show that this imprecision means that in most of the cases in which past analyses have failed to reject a null of a zero relationship, they also fail to reject the polar opposite null that there is no Easterlin paradox. With more data and careful attention to the details surrounding sample, questions and question order effects, we are better able to be more precise about which hypothesis the data can reject.

The rest of this paper is organized as follows. In the next section we give some of the background and stylized facts in the literature. We then lay out our conceptual framework. We show how various comparisons—between rich and poor people, rich and poor countries, and a given country over time—identify different aspects of the relationship between well-being and income. We next turn to the data, compiling data from all six of the major repeated cross-national survey efforts, paying careful attention to how unrepresentative samples, changes in questions, and changes in question ordering might lead one astray. Using consistently-coded data we show that simple panel regressions reveal a clear and quantitatively important association between growth in subjective well-being and growth in income. Moreover, this relationship is robust to controlling for cyclical movements in GDP, unemployment, and inflation. We then estimate the relationship between economic growth and growth in well-being using only very long-run changes in income, finding similar results. In order to use all of the data available, and increase the precision of our results, we then combine all our data into a single panel of panels, which yields robust evidence of a relationship between GDP growth and rising well-being. We then briefly consider individual country time series estimates, as many of

them have been considered in other research. Finally, we estimate average well-being in the world in each year, net of country and dataset fixed effects, yielding an estimate of "global well-being," which has grown over recent decades.

II. Background and Stylized Facts

The literature documenting the link between income and subjective well-being is, by now, rather voluminous. So we begin by simply reviewing the key stylized facts that have become clear over the past decade as summarized by the Sarkozy Commission (2010, p. 149):

i) countries with a higher level of GDP per capita do report higher life evaluations;

ii) the relationship between life-evaluations and the logarithm of GDP is broadly linear (i.e. it does not flatten out at higher income levels beyond the flattening that is implicit in the log-linear relation between the two variables); and

iii) the relation between country level GDP and average life-evaluations is similar to the one that applies to individuals

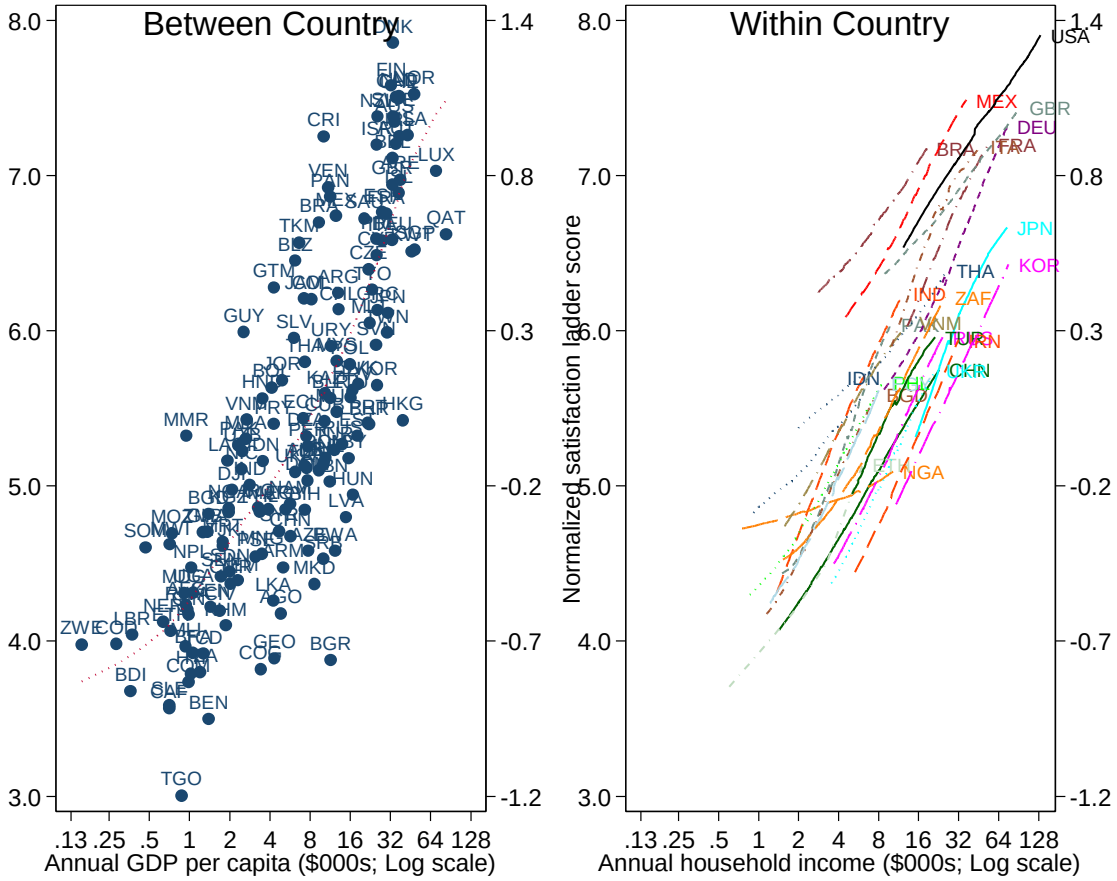
The first fact became clear with the release of the Gallup World Poll and subsequent waves of other large cross-national surveys (Deaton, 2008, Stevenson and Wolfers, 2008, Sacks et al, 2010, Deaton, 2008, Stevenson and Wolfers, 2008, Sacks et al, 2010, Inglehart, 2008, Hagerty and Veenhoven, 2003). Many of these papers have also shown that the relationship between subjective well-being and the log of GDP per capita is linear, with no evidence of satiation. Finally, Stevenson and Wolfers (2008) and Sacks et al (2010) show that the gradient is similar when estimated between countries and within countries and in most cases a gradient in the range of 0.3-0.4 cannot be rejected.[2]

Each of these findings can be illustrated in Figure 1, which draws on data from the Gallup World Poll, since it is the most comprehensive cross-national study of well-being. The left panel shows the between-country cross-section, analyzing national averages of well-being and GDP per capita from 2005 to mid-2011. On average, countries with a higher level of per capita GDP over this period report higher life evaluations, and these data hew tightly to the linear

[2] There is a long literature beginning with Easterlin's pioneering work in the 1970s, establishing that richer people are happier than poorer people within a country (Blanchflower and Oswald, 2004, Clark and Oswald, 1986, Diener, 1984, Diener et al., 1993, Diener and Biswas-Diener, 2002, Diener and Oishi, 2000, Duncan, 1975, Easterlin, 1973, Easterlin 1974, Frank, 1985, Frey and Stutzer 2002, Gardner and Oswald 2001, Hagarty, 2000, Schyns, 2001).

regression line. We add a non-parametric lowess fit to allow the data to dictate the functional form of this relationship, and with GDP plotted on a logarithmic scale, this dotted line is approximately linear. This confirms the second observation, which is that the well-being–income relationship is roughly linear in the log of GDP, which means that similar percentage increases in GDP yield similar measured changes in well-being. Also, notice that this lowess fit does not flatten out at higher levels of GDP. That is, there is little evidence of satiation at higher income levels.[3]

Figure 1



[3] Layard et al (2009) point to the fact that Deaton (2008) shows estimates that are imprecisely estimated at incomes per head over $20,000 as leaving open the possibility that the relationship between well-being and income does flatten out at high incomes. Using the 2005-mid-2011 data we find a statistically significant slope for countries with incomes per head over $20,000, with a point estimate of .58 that exceeds that estimated for poorer countries. Examining incomes per head over $25,000 the estimated gradient is .49, but with a limited number of countries it is imprecisely estimated. Stevenson and Wolfers (2011) examine satiation in greater detail.

The right panel analyzes individual responses, highlighting the within-country cross-section for each of the world's most populous 25 countries. For each country, we show a non-parametric lowess fit summarizing the relationship between responses to a question asking about overall life evaluations on a 0-10 scale, and household income, measured at purchasing power parity. Again, the horizontal axis shows household income on a logarithmic scale. The common feature in each of these countries is not just that rich people within a country are happier than poor people (that is, the lines slope up), but also, the line rises roughly linearly, and so these data also suggest a linear-log relationship. Note that both panels are plotted against comparable axes to highlight the fact that the well-being–income relationship observed within-countries is both similar across countries, and similar to that observed in the between-country cross-section.

This final observation points to an important development in this literature, which is the focus on the relative magnitude of the estimated well-being–income gradient, rather than its statistical significance. This is a point that has caused substantial confusion in the literature. Studies of the within-country cross-section relationship typically involve comparisons of the responses of thousands of individuals, and so the bivariate well-being–income relationship is nearly always statistically significantly different from zero. By contrast, early comparisons of the between-country cross-section involved few countries, and so the bivariate income–well-being relationship was often statistically insignificant with large standard errors. It was the juxtaposition of a statistically significant finding with a statistically insignificant finding that many labeled paradoxical. Stevenson and Wolfers (2008) show that in most of the early cross-country studies the size of the estimated gradient was not statistically significantly different from that estimated using the within-country cross-section, even though it was also not statistically significantly different from zero.

The same concerns are even more important as we now turn our attention to national time series data. In particular, there are few comparable observations of populations within the same country asked the same well-being question over time. Moreover, the range of variation of GDP when comparing the same country over time is substantially more limited than when comparing rich and poor countries. And so with relatively few observations of changes in well-being over time, and limited variation in GDP over time, it should be no surprise that many studies have found a statistically insignificant result when analyzing the time series. This is compounded by

the fact that time series data are often impacted by changes in survey design—including changes in the question, the ordering of questions, and even the population surveyed.

The problem of insufficient statistical power was ultimately solved in the between-country cross-section by collecting data from a larger number of countries. Accumulating more data in the time series requires time and it will take decades for more data to accumulate. Instead, our research strategy is to both analyze all available datasets as systematically as possible, and to mine the historical archive to allow longer-run comparisons to be made. In so doing, we will analyze data from a variety of surveys asking different questions. We focus primarily on evaluative questions—life evaluations, life satisfaction, and happiness overall. By contrast, recent work by Deaton and Kahneman (2010) suggest that affective measures (such as "did you smile or laugh a lot yesterday?") may have a very different relationship with income. Thus, we should be clear that our main focus will be in assessing time series movements in evaluative measures of well-being. This is partly a question of necessity rather than a choice, as consistent cross-national data probing affect simply have not been collected for a sufficiently long period to allow a sustained study of their time series.

III. Conceptual Framework

The obvious omission from the set of stylized facts discussed above is that none speak to the time series relationship between GDP and national well-being aggregates. Our goal is to fill this void. While we shall have little to say about a causal interpretation, these stylized facts are often used to speak to questions about the role of income and economic factors in determining well-being.

For instance, Easterlin's Paradox—the claim that raising the incomes of all did not raise the well-being of all—has been used to argue that well-being is determined by relative income concerns. Thus, in the within-country cross-section, higher income is associated with greater well-being, only because it makes you richer than the Joneses', where "Joneses" denotes people in your country, community or other national or sub-national comparison group to whom you compare yourself. Related models emphasize instead the role of one's position or rank in the income distribution, rather than your position relative to a local average. In all of these cases, broadly-based economic growth raises both your income, and the Joneses', and hence makes

neither of you better off. This perspective, if true, has first-order policy implications. For instance, Layard (2003) argued that the Joneses' higher income has a negative externality, because his success lowers your relative income, making you feel less happy.[4] As such, it suggests a new rationale for taxing income (or consumption). Likewise, Easterlin (2003) and Oswald (2010) have argued that these findings suggest de-emphasizing economic growth as a target of policy.

There is substantial debate within this literature as to what the reference group is—whether it is national, local, within your social circle or social class, or demographic group (Clark et al. 2008, Clark et al., 2009, Clark, 2010, Luttmer, 2005) or even a comparison with yourself at some other point in time Clark (2008). Even so, to the extent that economic growth makes any and all of these groups better off at the same rate that it makes you better off, it will not raise your well-being. Thus, theories in which relative income, relative consumption, or rank determine well-being all predict that higher average economic growth will not yield higher average well-being.

Likewise, these theories predict that people born into richer countries—who will also be born into richer reference groups—will not be happier than those born into poorer countries with poorer reference groups. If relative income is all that matters, then the blessings of greater individual riches are exactly offset by the curse of richer (within-country) reference groups. Consequently these theories predict that those born into rich societies should enjoy no greater well-being than those born into poorer societies. The data in Figure 1 clearly falsify this prediction. We have been puzzled why this observation has not called theories based on relative comparison into greater question, although our reading is simply that many researchers are uncomfortable making inferences from purely cross-sectional evidence, preferring instead to emphasize national time series.

One possibility is that the relevant reference group is not a subset of fellow citizens, but rather a global reference group in which we are all citizens of one world and assess our lives relative to the lives of those in the world. The implication is that people born into richer

[4] A narrow interpretation is that it is the consumption of certain goods that make your neighbors less happy. For example, Frank (2005) also points to the negative spill-over effects of "positional" consumption goods, or goods such as housing, for which relative position appears to matter most. Frank argues that increased spending by top earners on positional goods indirectly exerts upward pressure on median earner spending, resulting in an equilibrium in which society as a whole spends too much on positional goods and too little on non-positional goods. Thus, tax cuts for the wealthy, which are spent mostly on positional goods, increase the size of consumption externalities.

countries are happier than those born into poorer countries, but if income rises around the world it will not raise well-being. Thus theories emphasizing a global reference group—which could be the global average, the richest country, the richest citizens in a particular country, or even just archetypes seen in movies—suggest that broad-based economic growth which raises global income without changing the distribution of income will not raise well-being.

A related set of theories emphasize adaptation, in which the relevant reference point is not the economic success of others, but rather your own past economic successes. By this view, people get used to higher incomes, and eventually, their well-being returns to a pre-determined set point. Kahneman and Thaler (1991) describe adaptation as dooming people "to march forever on a hedonic treadmill" (p. 342). Thus, increased income will yield higher well-being for a period, until people get used to their greater riches, and well-being returns to its baseline level.

The empirical implications of adaptation theories depend on income dynamics. If income levels have been stable for long enough for adaptation to be complete, there will be no relationship between levels of well-being and levels of income. In reality though, incomes are rising for some, and falling for others. Thus, in the within-country cross-section, the rich may be happier than the poor if—as seems likely—people who have recently experienced positive income shocks are over-represented among the rich, and people who have suffered negative shocks are over-represented among the poor. The implications for the between-country cross-section are different, because the income dynamics are different. In particular, differences between countries in levels of GDP are extremely persistent—the correlation between log(GDP per capita) in 1960 and 2010 is 0.83—and so presumably the populations of both rich and poor nations have largely adapted to these differences. Thus, adaptation-based theories predict a much weaker (and possibly nil) relationship between well-being and income in the between-country cross-section. The more direct implications of these theories is in the time series, where adaptation predicts that rising income will raise well-being for a short period, with the effect decaying over time. Thus, comparing changes in well-being with changes in GDP over short periods may yield large effects, while comparisons over periods long enough for adaptation to have occurred will yield smaller effects. And extremely long differences—over the period long enough for complete adaptation to have occurred—will yield nil effects.

Finally, the very simplest view of the relationship between income and well-being is that well-being rises with one's level of income. The implications of this are clear: in within-country

cross-sections, richer people will be happier than poorer people, and in the between-country cross-section, people in richer countries will on average be happier than people in low-income countries. And likewise, in the time series, rising GDP will be associated with rising well-being. Thus, this theory predicts that greater income—whether accruing to an individual, a country, or over time—is associated with greater well-being in roughly similar magnitudes. The quantitative prediction is that the well-being–income gradient measured in the within-country cross-section is similar to that measured in the cross-country cross-section, which is similar to that measured in national time series. That is, the "no paradox" null hypothesis is not about whether the income–well-being gradient measured in various different ways is statistically significantly different from zero. Instead, it is a hypothesis about the similarity of the well-being–income gradient in the within-country cross-section, the between-country cross-section, and national time series.

We should be clear: the claim that income exerts a strong force on well-being is a claim about the size of β, which we refer to as the well-being–income gradient. It is possible for this gradient to be large, even if the correlation between well-being and income is small. That is, if factors other than income account for much of the variation in well-being, then the correlation may be low even though income has a quantitatively important effect on well-being. None of the theories that we have discussed rule out other factors influencing well-being.

We should also be clear about the precision necessary to distinguish among these different theories. For instance, failing to reject the null that well-being rises with increases in income at the same magnitude as seen in the cross-section is evidence for a role for absolute income, but it does not eliminate a role for relative income or adaptation. That requires a stricter finding of precisely estimated coefficients that are identical to each other. Moreover, there are conceptual differences in income measured across people within a country—at a point in time these measurements include transitory and permanent income—between countries, and over time. Beyond conceptual differences there are measurement issues with income as well as subjective well-being that makes a stricter finding of "no adaptation" or "no relative income effects" more difficult to establish.

We now turn to assessing the evidence from various cross-country panel datasets.

IV. Analyzing Cross-Country Panel Data

We have compiled data from every large-scale cross-national well-being research effort of which we are aware. In this section, we analyze data from each of these international panels separately. Throughout, we will run regressions of the form:

$$Well-being_{c,t} = \beta \log(GDP_{c,t}) + \sum_t Survey\ wave_t + \sum_c Country_c + \epsilon_{c,t} \quad [1]$$

Our goal is to focus on the time series dimension of the data, and so we include country and survey-wave fixed effects. The country fixed effects allow us to partial out permanent differences in well-being across countries, due to, for example, cultural or climatic differences.[5] The survey-wave fixed effects allow us to partial out common time patterns, which ensures that our findings reflect the different time paths of GDP growth across countries. (The common global pattern of well-being is also of interest, and so we return this in Section VIII.) Another reason we control for survey-wave fixed effects is that typically survey designs are constant within a survey wave, and so this allows us to hold constant design effects such as question order. Answers to well-being questions are sensitive to nearby questions and the magnitude of question order effects on measured well-being can be quite large (Stracker et al, 1988; Schwartz et al, 1991; Deaton, 2011). Including survey-wave fixed effects allows us to partial out these effects.

The independent variable of interest in these regressions is the log of GDP per capita, measured at purchasing power parity. We draw these data from several sources. Our main source is the World Bank's World Development Indicators (WDI). These data provide annual, PPP-adjusted per capita GDP figures for most countries. The PPP adjustments are based on the 2005 round of the International Comparisons Project, and all of our estimates are in 2005 international dollars. When the WDI data are missing, we supplement them with data from the Penn World Tables (mark 6.3); failing that, we use data from the IMF's World Economic Outlook. When the IMF's data are unavailable, we use data from Angus Maddison and, in a few cases, the CIA Factbook.

One concern is that variation in GDP reflects both the long-run economic growth that we are interested in analyzing, and also short-run business-cycle variation. Moreover, we know

[5] Some researchers have argued that the positive cross-country gradient reflects the correlation of GDP with these country fixed effects (Easterlin et al, 2010).

from Di Tella, MacCulloch and Oswald (2001, 2003) and Wolfers (2003) that well-being is sensitive to the state of the business cycle. In order to focus only on low frequency variation we consider an alternative independent variable—a measure of trend GDP that has been purged of business-cycle frequency variation. We construct this variable by applying a Hodrick-Prescott filter to the annual time series of log GDP per capita for each country.[6]

Turning to our measure of well-being, we have assembled six datasets which ask general subjective well-being questions multiple times across multiple countries. Each of our datasets asks slightly different questions, or allows responses on different scales. While some of these scales have a natural quantitative interpretation (such as a 0-10 or 1-10 scale), others are qualitative in nature ("not at all satisfied"; "not very satisfied"; "fairly satisfied" and "very satisfied"). In previous work we have experimented with different ways of scaling these qualitative data (see Appendix A in Stevenson and Wolfers 2008). Here, we follow the simplest transformation, coding the least satisfied category as one, the next as two, and so on. We then transform answers from different scales into a comparable metric so that we can compare results across datasets. Thus for each of our datasets, we begin with the individual well-being responses, and standardize well-being so that it has a standard deviation of one, net of country and year fixed effects. This rescaling yields a naturally interpretable metric, as differences are all measured relative to the cross-sectional distribution of well-being within a typical country. Importantly, it also allows us to compare the estimated well-being–GDP gradient across datasets.

For each of our datasets we assembled repeated cross-sections of countries in which a single survey effort (e.g. the World Values Survey) asked an identical well-being question across repeated cross-sections. We exclude observations of countries that lack a nationally representative sampling frame, since it is not possible to estimate average well-being. We provide a basic description of each dataset as we go through the results and our appendix describes these datasets in much more detail.

World Values Survey

The World Values Survey asks: "All things considered, how satisfied are you with your life as a whole these days?" and "Taking all things together, would you say you are: very happy; quite happy; not very happy; not at all happy?" Through time, this survey has expanded its

[6] Since our data are annual, we use a smoothing parameter 6.25. To minimize end-point problems, we use the IMF's GDP projections for 2010-2016 in constructing trend.

scope enormously, from 21 countries (mostly middle- and upper-income) in the 1981-84 wave, to a more representative 56 countries in the most recent 2005-08 wave. As such, it comprises a heavily unbalanced panel.

There are two challenges, beyond the unbalanced panel, in doing time series analysis using the World Values Survey. First, many national samples—particularly in the early years—are not nationally representative. In particular, these samples often over-represent urban areas, more educated and English-speaking populations; these are all groups which tend to be both richer, and more satisfied with their lives.[7] Because whole segments of the population are entirely absent from these non-representative surveys, there is no way to way devise sampling weights to make them comparable with later representative surveys. In short, it is not possible to estimate average well-being in these cases and since we are studying the change in average well-being over time, it is essential to have measures of average well-being.[8] As such, we drop all non-representative surveys from our samples.

Second, there are important changes in question ordering in successive waves. These question order issues effect both the life satisfaction and happiness questions. In the 1994–99 and 1999–2004 waves, the life satisfaction question was preceded by a question asking about one's financial satisfaction. Stevenson and Wolfers (2008) show that life satisfaction is more correlated with the responses to the financial satisfaction question when they are proximate. In those same waves, the happiness question was part of a battery of questions probing the importance of friends, family, leisure, politics, and religion, and a similar analysis reveals that the correlation of measured happiness with these variables rose.[9] Additionally, Easterlin et al (2010) point to a change in the ordering of the response options—whether one is offered options that range from happiest to unhappiest or vice versa—as biasing estimates since those that move from most happy to least happy tend to generate higher average measured happiness (p.22464).

[7] Stevenson and Wolfers (2008) contains a detailed appendix that discusses the sampling frame in each country-wave that is non-representative. In some cases, such as when the language that the survey was conducted in changes, it would be impossible to find a consistent sample across the waves as one could not know which people who later take the survey in a different language would have been able to participate when the survey was not offered in that language.

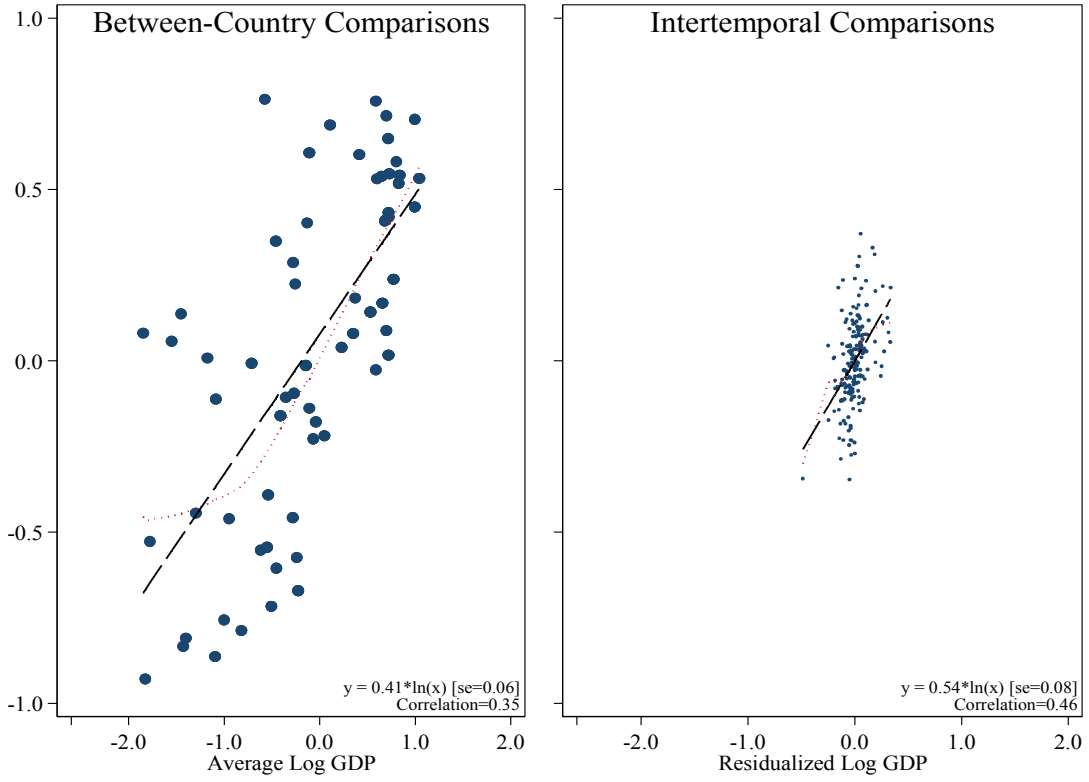
[8] Other scholars have chosen to use these samples; however the estimated changes in well-being reflect both the change in the population being surveyed and any changes occurring in the total population. It is not possible to parse these two effects, even when including dummy variables for non-representative samples, as is done in Easterlin et al (2010).

[9] Research has shown that when respondents are queried about specific well-being in specific domains it impacts their responses to general life-satisfaction questions. (McClendon, M. J. and D. J. O'Brien. 1988)

Thus, to address question order issues it is critical that we control for survey-wave fixed effects, which removes the common effects of these changes across countries.[10]

We start by plotting the relationship between average well-being in a country and per capita GDP for all countries included in the World Values Surveys. The left panel of Figure 2 uses all waves of the sample and shows each country's average well-being and GDP per capita in the time periods they are surveyed. There is a clear relationship between life satisfaction and GDP across countries, and the estimated well-being–GDP gradient is 0.41. We compare this to the time series relationship obtained by estimating equation [1]. Plotting the two graphs on the same scale side-by-side illustrates how little income has grown within countries over the past 25 years relative to the dispersion of income around the world. For the sample of country-years in which we have multiple nationally representative satisfaction data from the World Values Survey

Figure 2: World Values Survey Satisfaction and GDP, Between Countries and Over Time



[10] Easterlin et al (2010) focus on the life satisfaction question rather than happiness as they argue that "there is reason to believe the WVS happiness data are biased upward due to a statistical artifact" (p. 22465). We too focus on life satisfaction; however the statistical issue they highlight can easily be adjusted for by including wave fixed effects in the regression.

94.3% of the variance in log GDP per capita is between-country variation, and only 5.7% is intertemporal variation. Of that intertemporal variation, 65% is common across countries (and hence is accounted for by wave fixed effects), while the remaining 35% identifies our models with country and wave fixed effects.

Thus, our earlier claim that the range of variation of GDP when comparing the same country over time is substantially more limited than that when comparing rich and poor countries is visually apparent in this figure. But equally, it is also clear that in the World Values Survey the slope of the relationship between well-being and income over time is similar to that seen across countries. We test this more formally in in Table 1.

The first row of Panel A of Table 1 reports the between-country estimates shown in the left panel and the second row reports the time series result, β from equation (1)—the specification which includes country and wave fixed effects. This estimate indicates by how much in standard deviations well-being moves from its country average when GDP (in percent) moves from its country average. The coefficient from the time series regression is 0.54 and is precisely estimated. The next few rows show the test of two hypothesis. First, we test to see whether we can reject that the time series coefficient is 0, which we can with 99% confidence. Next we test to see whether we can reject that the time series coefficient is equal to the between-country coefficient. We cannot reject the hypothesis that the estimated gradients are the same.

In panel B, we replace GDP with our estimate of trend GDP. These estimates are quantitatively similar and our results are the same: we can reject that the time series coefficient is 0 and we cannot reject the hypothesis that the time series gradient is the same as the between country gradient.

We next report results using the happiness question in the World Values Survey in Column 2. In both the between country and the time series the estimated gradient is smaller than that seen using the satisfaction question. We find a precisely estimated coefficient of 0.32 when looking between countries and an imprecisely estimated coefficient of 0.16 in the time series. Turning to our two hypothesis tests, we see that we can neither reject that the time series relationship is zero nor that it is the same as the between country estimate. Replacing our measure of GDP with trend GDP in Panel B increases the precision of our time series estimate using the happiness question and we become able to reject the null of a 0 gradient and we remain unable to reject the null that the between country and time series estimates are the same.

Table 1: Cross-sectional and panel regressions of subjective well-being on GDP per capita

Specification:	(1) WVS Satisfaction 1981-2008	(2) WVS Happy 1981-2008	(3) Eurobarometer Satisfaction 1973-2009	(4) ISSP Happy 1991-2008	(5) Gallup Ladder 2005-mid2011	(6) Pew Ladder 2002-2010	(7) Latinobarometro Satisfaction 2001-2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Analyzing the link between Well-being and Log(GDP per capita)							
Between-country variation: $\overline{Well-being_c} = \beta^{between}\overline{Log(GDP\ per\ capita_c)} + \epsilon_c$							
$\beta^{between}$	0.46***	0.32***	1.01***	0.19**	0.34***	0.28***	0.22*
	(0.08)	(0 .08)	(0.29)	(0.09)	(0.02)	(0.05)	(0.12)
Within-country time series variation: $Well-being_{ct} = \beta^{within}\log(GDP\ per\ capita_{ct}) + Country_c + Time_t + \epsilon_{ct}$							
β^{within}	0.54***	0.16	0.17***	0.55***	0.37**	0.56*	0.52
	(0.08)	(0.10)	(0.05)	(0.17)	(0.18)	(0.29)	(0.37)
Hypothesis tests							
Test: $\beta^{within} = 0$	p<0.01	p=0.11	p<0.01	p<0.01	p=0.04	p=0.07	p=0.18
Test: $\beta^{within} = \beta_{between}$	p=0.56	p=0.17	p<0.01	p=0.06	p=0.85	p=0.35	p=0.44
Panel B: Analyzing the link between Well-being and *Trend* Log(GDP per capita)							
Between-country variation: $\overline{Well-being_c} = \beta^{between}\overline{Log(GDP\ per\ capita_c)} + \epsilon_c$							
$\beta^{between}$	0.47***	0.32***	1.01***	0.20**	0.34***	0.29***	0.22*
	(0.08)	(0.08)	(0.29)	(0.09)	(0.02)	(0.05)	(0.12)
Within-country time series variation: $Well-being_{ct} = \beta^{within}\log(GDP\ per\ capita_{ct}) + Country_c + Time_t + \epsilon_{ct}$							
β^{within}	0.52***	0.17*	0.16***	0.63***	0.19	0.70**	0.61
	(0.08)	(0.10)	(0.05)	(0.20)	(0.22)	(0.29)	(0.46)
Hypothesis tests							
Test: $\beta^{within} = 0$	p<0.01	p=0.09	p<0.01	p<0.01	p=0.38	p=0.02	p=0.20
Test: $\beta^{within} = \beta_{between}$	p=0.68	p=0.20	p<0.01	p=0.05	p=0.52	p=0.16	p=0.41

Notes: Each cell shows the coefficient on ln(GDP) obtained from regressing subjective well-being on log GDP and other variables. Each row is a different specification and each column is a different measure of well-being (typically a different dataset). In Panel A we use ln(GDP) in our specifications, in Panel B we replace ln(GDP) with trend ln(GDP) measured using an HP filter. In the first row of each panel we regress average well-being on average ln(GDP). In the second row of each panel, we regress SWB against log(GDP) and country and survey wave fixed effects.

In sum, using the World Value Surveys we find a positive gradient between well-being and income over time and we are unable to reject either that the between country and time series estimates are the same. Equally, we are unable to reject a gradient of 0.3-0.4, the estimated within country cross-section gradient established in previous work (Stevenson and Wolfers, 2008).

Eurobarometer

The Eurobarometer has been run since 1973 and covers the countries of the European Union. Each survey consists of a sample of approximately 1,000 per country typically surveyed twice a year. As the European Union has grown so too has the Eurobarometer; in 1973, nine countries were surveyed and by 2009, the latest wave for which we have data, the sample included 27 European Union member states and three candidate countries. Thus the Eurobarometer is also an unbalanced sample in which progressively poorer countries were added over time.

The Eurobarometer includes a question assessing life satisfaction in most waves: "On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?"[11] The survey also briefly included (from 1975-79 and 1982-86) a direct question about happiness: "Taking all things together, how would you say things are these days—would you say you're very happy, fairly happy, or not too happy these days?". Given the short period in which the happiness question was included, we focus on life satisfaction. For the purposes of our analysis, we keep West Germany separate from East Germany.

Column 3 reports regression results for life satisfaction measured in the Eurobarometer. The first row shows that the between country gradient is quite large—estimated at 1.01 with a 95% confidence interval that would allow a gradient as small as 0.7 or as large as 1.26. This is much larger than the estimated gradient in other datasets and one can easily reject a coefficient as small as 0 or even as small as 0.3-0.4, the estimated within-country gradient seen in most countries and in most datasets. Turning to the time series in row (2), we see that the estimated gradient is much lower at 0.17, however it is statistically significantly different from zero. In the Eurobarometer data we can reject both the hypothesis that the estimated gradient is zero and that it is the same as that estimated between countries. Moreover, we can reject the hypothesis that

[11] The life satisfaction question was not asked in 1974 and 1996.

the time series gradient is in the range of 0.3 to 0.4. Replacing ln(GDP) with our estimate of trend ln(GDP) in Panel B yields nearly identical estimates.

International Social Survey Program

We next turn to the International Social Survey Program, an international collaboration that has released surveys since 1984 on varying topics. The survey included a happiness question: "If you were to consider your life in general, how happy or unhappy would you say you are, on the whole?" in 1991, 1998, 2001, 2007, and 2008. While we have data for 39 countries, only 13 have data for all 5 years.

The estimate of the gradient between countries is a bit smaller than we see in other datasets at 0.19. However, turning to the panel regression we see that the point estimate of the time series gradient is larger, at 0.55, and we can reject a null of no relationship between well-being and income. We can also reject the hypothesis that the between country and time series gradients are the same; the time series gradient is statistically significantly bigger than the between country estimate. However, we cannot reject the hypothesis that the estimated time series gradient is in the range of 0.3 and 0.4. Replacing ln(GDP) with our estimate of trend ln(GDP) increases the point estimates of both the between country and time series gradients very slightly, but the qualitative conclusions, and the results of our hypotheses tests, are the same.

Gallup World Poll

The most ambitious cross-country surveys of subjective well-being are being done by Gallup which begin its World Poll in 2005-06 with 132 countries and has surveyed an increasing number of countries in each year since. The survey was designed to measure subjective well-being consistently with similar questions asked in each country and a nationally representative sample of citizens aged 15 and older for each country. The survey asks a wealth of subjective well-being questions, but the most holistic question of life satisfaction or happiness is a ladder question "Here is a ladder representing the "ladder of life." Let's suppose the top of the ladder represents the best possible life for you; and the bottom, the worst possible life for you. On which step of the ladder do you feel you personally stand at the present time?" We include data through the middle of 2011 and we use multiple years of observations for 141 countries.

Gallup thus provides the most comprehensive and precise estimate of the between country subjective well-being–income gradient which we report in the first row of Column 5. The estimated between-country gradient is 0.34 and the estimated time series gradient is 0.37.

Both are precisely estimated and thus we can reject the hypothesis that the time series gradient is 0. Moreover, we are unable to reject the hypothesis that the time series gradient is equal to the estimated between country gradient.

However, the short time period of the Gallup data and the fact that it covers the period of the global financial crisis means that much of the variation in GDP is cyclical. When we replace ln(GDP) with our measure of trend ln(GDP) the estimated coefficient on the time series is slightly smaller and is statistically insignificant. Thus, while we remain unable to reject the hypothesis that the time series and between country gradients are the same, we are equally unable to reject the null hypothesis of no relationship between well-being and income over time.

Pew Global Attitudes Survey

The Pew Global Attitudes Survey has been conducted every year since 2002, but has only asked a subjective well-being question in three waves: in 2002, 2007, and 2010. The question asked is the same ladder of life question asked by Gallup: "Here is a ladder representing the "ladder of life." Let's suppose the top of the ladder represents the best possible life for you; and the bottom, the worst possible life for you. On which step of the ladder do you feel you personally stand at the present time?" Each year a different number of countries are surveyed and the 2002 and 2007 waves include the largest number of countries, 44 and 47 respectively. In 2010, 22 countries were survey. Altogether, there are 39 countries with more than one observation, 21 of which have three. However, among these countries, many of the samples are explicitly non-representative and thus we exclude them.[12] That leaves us with 29 countries, 15 of which have three observations. The countries include a sampling of both upper, middle, lower-income countries across the Americas, Western Europe, Eastern Europe, the Middle East, Asia, and Africa.

Column 6 of Table 2 reports regression results for the Pew Global Attitudes Survey with the first row showing a precisely estimated between country gradient of 0.28. Turning to the panel data—and here we are really taking short differences since for about half the countries we only have data for two points in time—the estimated gradient is 0.56. Testing our two hypothesis we find that we can reject the null hypothesis of no relationship and we are unable to reject the hypothesis that the estimated gradient in the time series is the same as that estimated between countries, nor can we reject coefficients of 0.3-0.4.

[12] Pew also excludes the non-representative countries from their trend analysis.

Replacing our measure of GDP with our estimate of trend GDP we find that the coefficient estimates sharpen. However, while the point estimate in the time series is nearly twice that of the between country estimate, as in Panel A we are unable to reject the hypothesis that the two coefficients are the same and we are able to reject the null of no relationship.

Latinbarometro

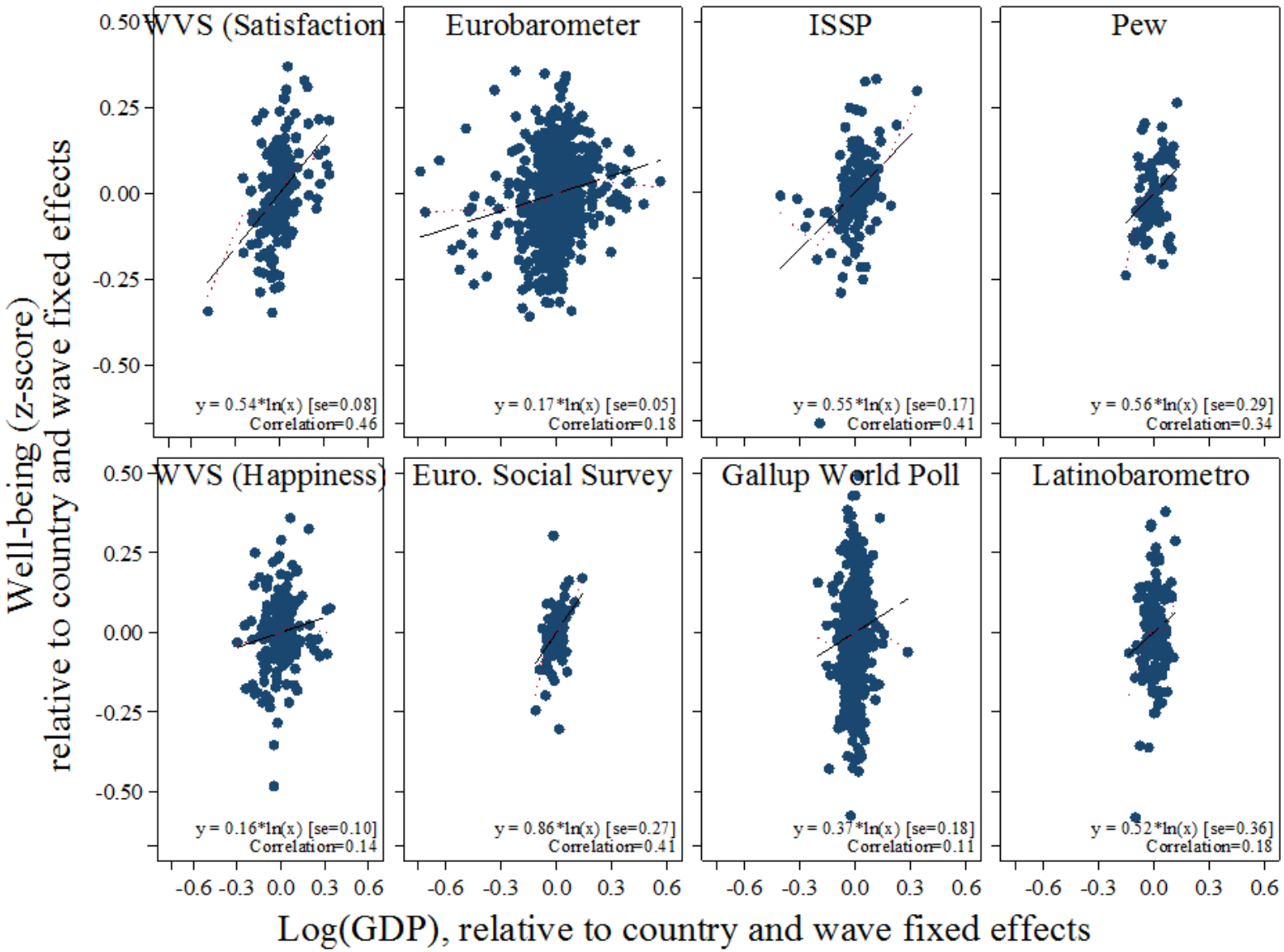
Finally, we use the Latinbarometro, a survey started in 1995 in eight countries with representative samples of around 1,000 respondents per country. It expanded the following year to include 17 Latin American countries. Since then, the survey has expanded to 18 with the addition of the Dominican Republic in 2004. The Latinbarometro asks: "In general, would you say that you are satisfied with your life? Would you say that you are very satisfied, fairly satisfied, not very satisfied or not satisfied at all?"

With less variation in the GDP of the countries surveyed it is not surprising that the between country gradient is the least precisely estimated of all of our datasets, but it is statistically significantly different from zero at the 10 percent level and with a point estimate of 0.22, we can't reject a coefficient of .3 to .4. Turning to the panel data we also see that the estimated time series gradient is 0.57, but it is imprecisely estimated. Thus, using the Latinbarometro data we can neither reject a null hypothesis of no relationship between well-being and income, nor can we reject a hypothesis that the time series gradient is the same as the between country gradient. Finally, we are also unable to reject an estimated gradient of 0.3-0.4. Replacing our measure of GDP with our estimate of trend GDP has little impact on the results.

Summary of Cross-Country Panel Data

Figure 3 summarizes our findings thus far across all of our datasets by plotting the change in subjective well-being and the change in GDP (both relative to survey-wave and country fixed effects). We also show the OLS fit and the nonparametric fit. The figure makes three points. First, our results in all datasets are driven by the positive relationship between well-being and income, over time, for the mass of countries. Second, there are substantial differences across datasets in the amount of residual variation in log GDP. The Eurobarometer and the World Values Survey, which cover several decades, have a great deal of variation. The ISSP, which covers 17 years, has more variation, while Pew, Latinobarometro, and the Gallup World Poll all have less. Third, the estimated relationship between well-being and GDP does not appear to depend on whether the dataset features a longer or shorter panel, or more or less

Figure 3


WVS (Satisfaction
Eurobarometer
ISSP
Pew
WVS (Happiness)
Euro. Social Survey
Gallup World Poll
Latinobarometro
y = 0.54*ln(x) [se=0.08]
Correlation=0.46
y = 0.17*ln(x) [se=0.05]
Correlation=0.18
y = 0.55*ln(x) [se=0.17]
Correlation=0.41
y = 0.56*ln(x) [se=0.29]
Correlation=0.34
y = 0.16*ln(x) [se=0.10]
Correlation=0.14
y = 0.86*ln(x) [se=0.27]
Correlation=0.41
y = 0.37*ln(x) [se=0.18]
Correlation=0.11
y = 0.52*ln(x) [se=0.36]
Correlation=0.18
0.50
0.25
0.00
-0.25
-0.50
-0.6 -0.3 0.0 0.3 0.6
Well-being (z-score)
relative to country and wave fixed effects
Log(GDP), relative to country and wave fixed effects

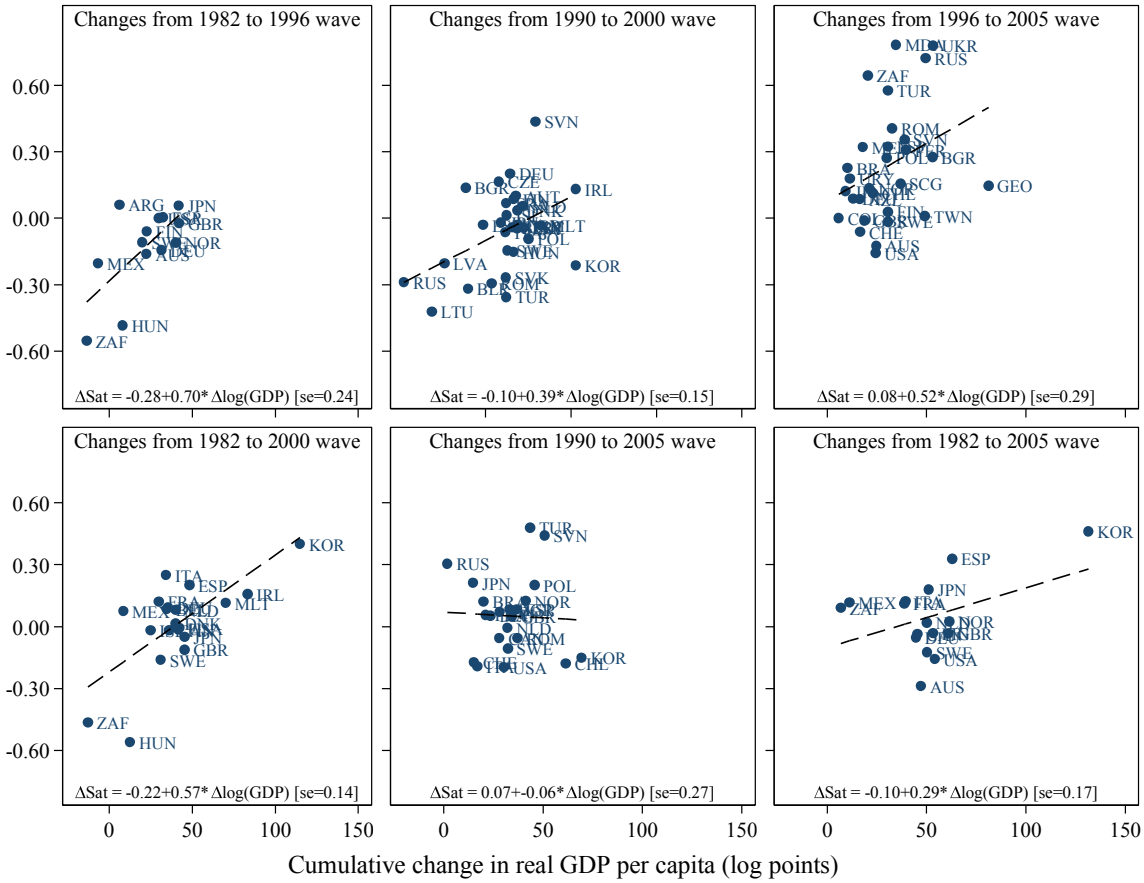
residual variation in GDP. Thus we interpret these estimates as suggesting that through time, variation in well-being tracks the log of per capita GDP, at about the magnitude we would expect from cross-country comparisons. In two surveys we are able to reject the hypothesis that the between country gradient is the same as the time series gradient. However in both of these cases the hypothesis of a null relationship between well-being and income in the time series is also rejected. In one case—the Eurobarometer—the estimated time series gradient is smaller than the between country gradient. In the second case—ISSP—the estimated time series gradient is larger than the between country gradient. In two of our seven assessments the data are sufficiently imprecise that we are unable to reject either the hypothesis that the two gradients are the same or a null of zero.

V. Long Differences

We now turn to analyzing our various datasets in first differences, rather than in their panel form. We do this for four main reasons. First, it aids transparency, as changes in well-being are easy to plot against changes in per capita GDP. Second, regressions in first differences remain appropriate even if one is concerned that per capita GDP has a unit root. Third, if one is concerned that our panel regressions—which use all the available data—are dominated by business-cycle movements, then one simple response is to analyze changes over periods long enough that business-cycles account for little of the variation. And fourth, this allows us to assess the sensitivity of our findings to possible adaptation, which would yield much smaller effects in the long-run. The cost, of course, is a loss in statistical power, relative to the panel regressions which exploit all of the variation over time within a dataset.

Another advantage of first differencing is that it differences out factors that have a common effect on all countries—which includes the effects of changing survey design, which is a major concern with the World Values Survey. This allows clean comparisons between countries which are represented in the same two waves of the World Values Survey. Because it is such an unbalanced panel, comparisons between different waves yield very different samples of countries, and sample sizes. In order to focus on long differences, Figure 4 only shows changes between waves that were separated by at least one other wave. Each panel also reports the results of the corresponding first-difference regression.

Figure 4: Medium and Long Differences in the World Values Survey



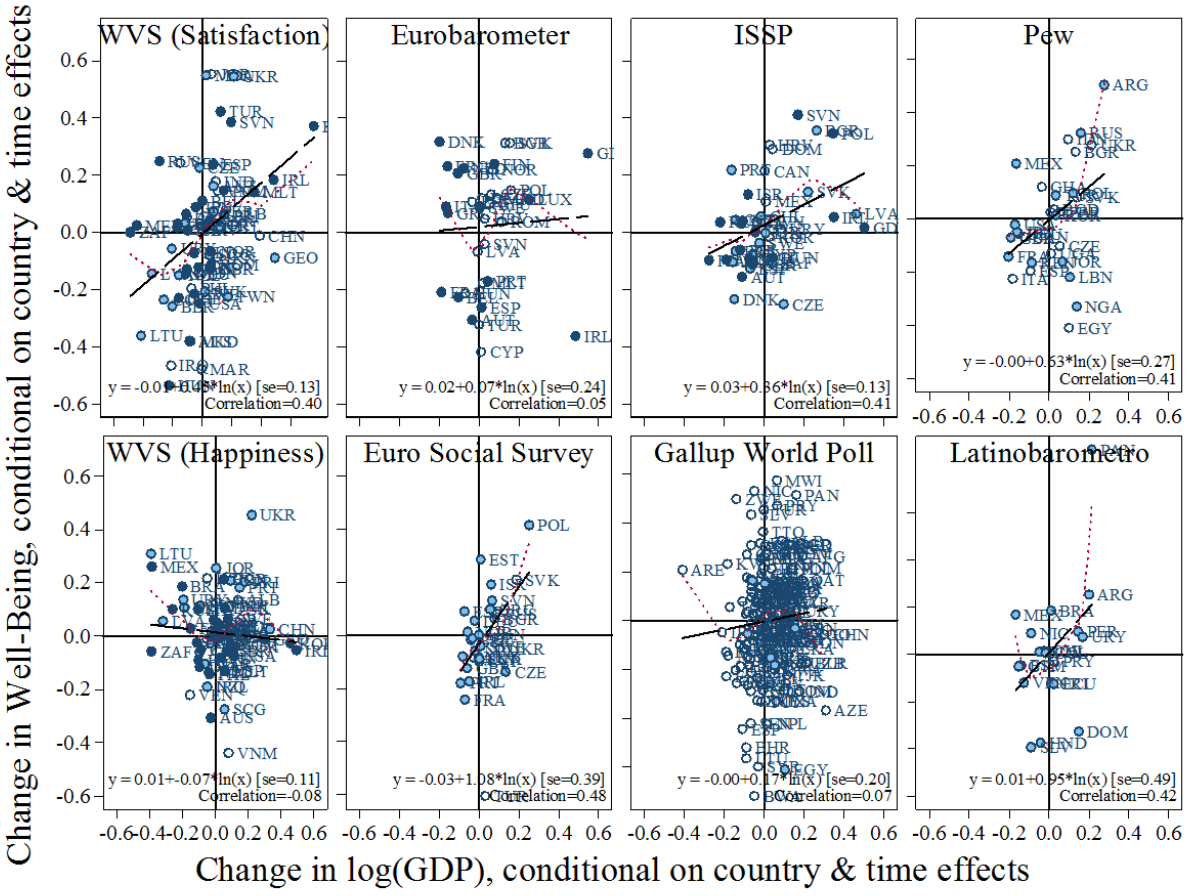
Notes: The figure shows changes in life satisfaction and log(GDP per capita) between each pairing of non-adjacent waves of the World Values Survey. Changes are shown whenever the same sampling scheme is used in each wave (Appendix X provides further details).

In five of the six panels, these long difference regressions show that countries which experienced larger changes in per capita GDP also experienced large changes in well-being. Moreover, the estimated well-being–income gradient in each panel is roughly similar, and none are statistically significantly different from the panel regression results for the whole sample, shown in Table 1. The only panel in which the slope is not clearly positive is between the 1990 and 2005 waves of the World Values survey, and there simply is not much variation in GDP growth among the 22 countries represented in this panel, and so this is also a very imprecise estimate. In none of the panels are we able to reject that the estimated gradient is between 0.3 and 0.4. A precision-weighted average of the coefficient estimates in all six panels yields an estimated well-being–income gradient of 0.43. In further regressions (not shown), we control for

unemployment, inflation and the output gap, and the corresponding precision-weighted average is 0.47.

Unfortunately, the very longest difference—shown in the final panel of Figure 4—only involves the sixteen countries surveyed with similar sampling frames in both the first and most recent waves. Thus, we also provide an alternative approach: As in section IV, we regress both well-being and log(GDP) on country and wave fixed effects, and analyze the residuals—that is, we analyze each variable purged of factors common to each wave, including changes in survey methods. Thus, we are left with the variation that was used to identify our earlier panel regressions. But in order to ensure that we focus only on the low frequency variation, we pair the first observation for each country with the most recent, allowing us to focus on the longest differences possible, while retaining the broadest possible sample. These results are shown in

Figure 5: Long Differences



Notes: For each dataset, we begin by regressing well-being on country and year fixed effects, and for each country, we compare the first and last observation of these residuals. Hollow circles

reflect differences over periods of less than six years; shaded dots are periods of 6-12 years, and solid dots are differences spanning 12 or more years. The reported regressions include all countries.

the top left panel of Figure 5. The estimated well-being–income gradient is both statistically significantly different from zero, and comparable in magnitude to that estimated in both the cross-country cross-section, and in our earlier panel results.

We also repeat this latter exercise for the Eurobarometer, the Pew Global Attitudes Survey, and the ISSP data, and the results are also shown in Figure 5. While we include results in the figure for the Gallup World Poll and the Latinbarometer, since we have few years of data these results are not particularly useful for considering long differences.

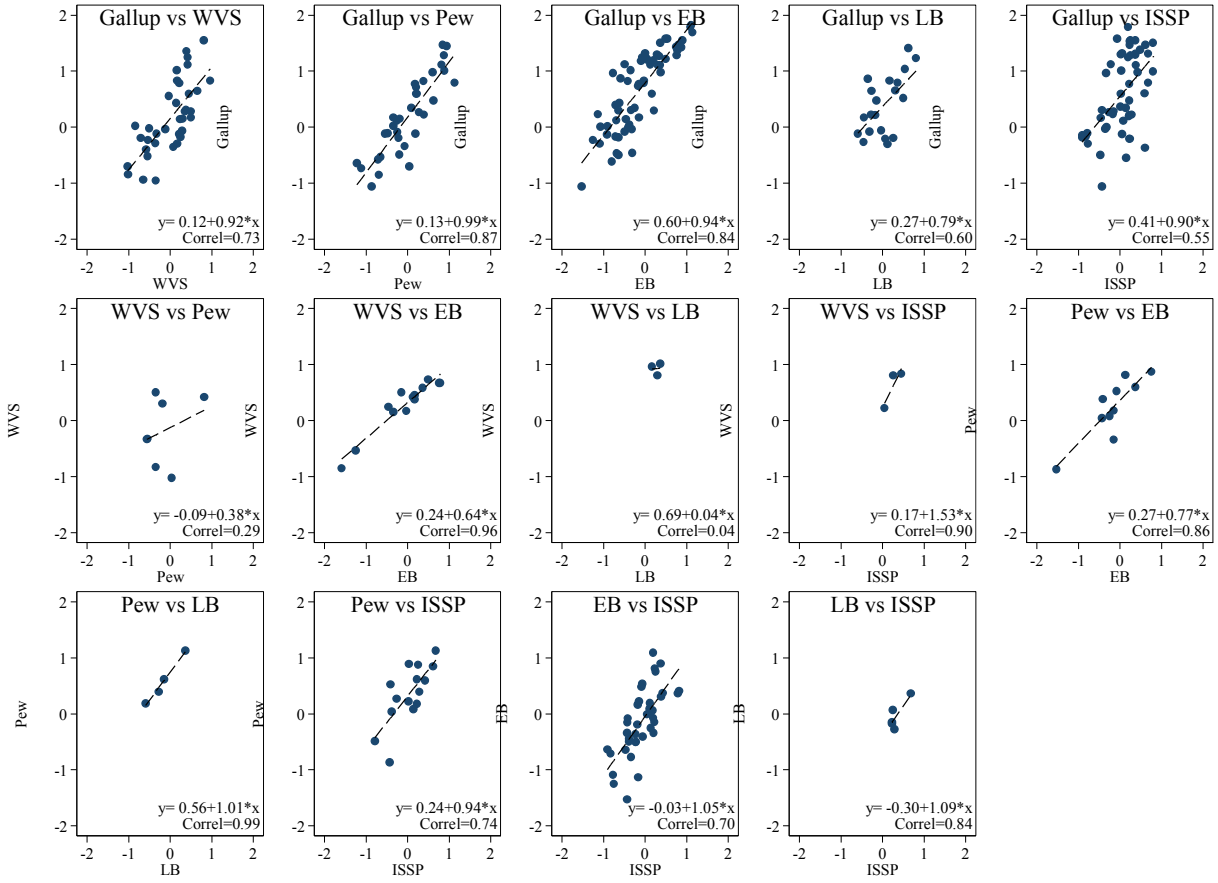
In three of our four datasets, there is a clear relationship between long-run changes in well-being and long-run changes in GDP. The exception is the Eurobarometer, which yields more imprecise regression estimates and a much noisier picture.

VI. Combining Data: A Panel of Panels

Thus far, our analysis has been somewhat piecemeal, in that we analyze the results from each of our datasets separately. These results are the most transparent and require fewer assumptions about comparability across surveys. But if we are willing to assume that each survey provides comparable data about well-being, then the six cross-national datasets can be usefully combined.

There are several issues to be concerned with. First, it could be that results from different survey organizations yield different average values. Fortunately, this is easy to deal with, by including survey fixed effects in all of our regressions. Second, while our normalization attempts to ensure that all survey responses are re-scaled into a comparable metric (the standard deviation of the within-country cross-section), these may still differ, and so different surveys may yield different variation across countries. Third, it may be that different surveys—which do ask different questions—actually reveal different constructs. In order to examine these last two concerns, we plot comparisons across surveys of average well-being of each country over whatever measurements were taken from 2005-08 in Figure 6.

Figure 6: Well-Being in Six Different Surveys, 2005-08



As this figure shows, each of these surveys yields quite similar results when compared across the most recent data. Our final concern is that even if each survey yields similar variation between countries, they may be differentially responsive to time series changes. Unfortunately, because each survey has such different historical coverage, this remains an untested assumption. With these caveats we now proceed to the analysis.

Recall from equation [1], that when we analyze the results from a specific survey, s, we are running a panel regression of the form:

$$Well-being_{c,t} = \beta^s \log(GDP_{c,t}) + \sum_t Wave_t + \sum_c Country_c + \epsilon_{c,t} \qquad [2]$$

If we are to "stack" the six datasets that we use and run the same regression, we get:

$$Well-being_{sct} = \bar{\beta} Log(GDP_{ct}) + (\beta^s - \bar{\beta}) \log(GDP_{c,t}) \qquad [3]$$
$$+ \sum_{s,t} Survey * Wave_t + \sum_{s,c} Survey * Country_c + \epsilon_{cts}$$

As long as the second term—which describes heterogeneity in the effects of GDP—is uncorrelated with log(GDP), then we can simply estimate:

$$Well - being_{sct} = \bar{\beta} Log(GDP_{ct}) + \sum_{s,t} Survey * Wave_t +$$

$$\sum_{sc} Survey * Country_c + \epsilon_{cts} \quad [4]$$

Thus, the estimated $\bar{\beta}$ is our most precise estimate of the well-being–income gradient, estimated using international panel data. As before, we continue to cluster our standard errors by country, so that these extra observations do not artificially inflate the precision of our estimates.

Table 2: Panel of Panels Regressions of Subjective Well-Being on Per-Capita GDP

Regression of SWB on ln(GDP) and indicated controls	*Coefficient on ln(GDP)*
1. **Country*dataset and year*dataset fixed effects**	0.33 (0.09) [0.15, 0.51]
2. **Control for output gap, plus country*dataset and year*dataset fixed effects**	0.21 (0.07) [0.08, 0.34]
3. **Control for inflation and unemployment, plus country*dataset and year*dataset fixed effects**	0.28 (0.06) [0.16, 0.40]
4. **Control for all macroeconomic indicators, plus country*dataset and year*dataset fixed effects**	0.23 (0.06) [0.16, 0.40]
5. **Control for all macroeconomic indicators, country, and year fixed effects, all interacted with dataset fixed effects**	0.22 (0.09) [0.04, 0.40]
Range of Years	1973-2011
Number of countries	159
Number of years	43
Observations	2124

Notes: The table shows the regression on log GDP from a series of regressions of standardized subjective well-being on log GDP and other variables. The sample is our super panel, created by pooling all of the data sets used in table 1. Each row is a separate regression of subjective well-being on ln(GDP) plus the indicated controls. The output gap is estimated from an HP filter of log(GDP). Inflation is included as a three-piece linear spline . Unemployment enters separately for OECD and non-OECD countries. In these regressions, there are also indicators for "missing inflation" and "missing unemployment". Robust standard errors, clustered on countries, are in parentheses.

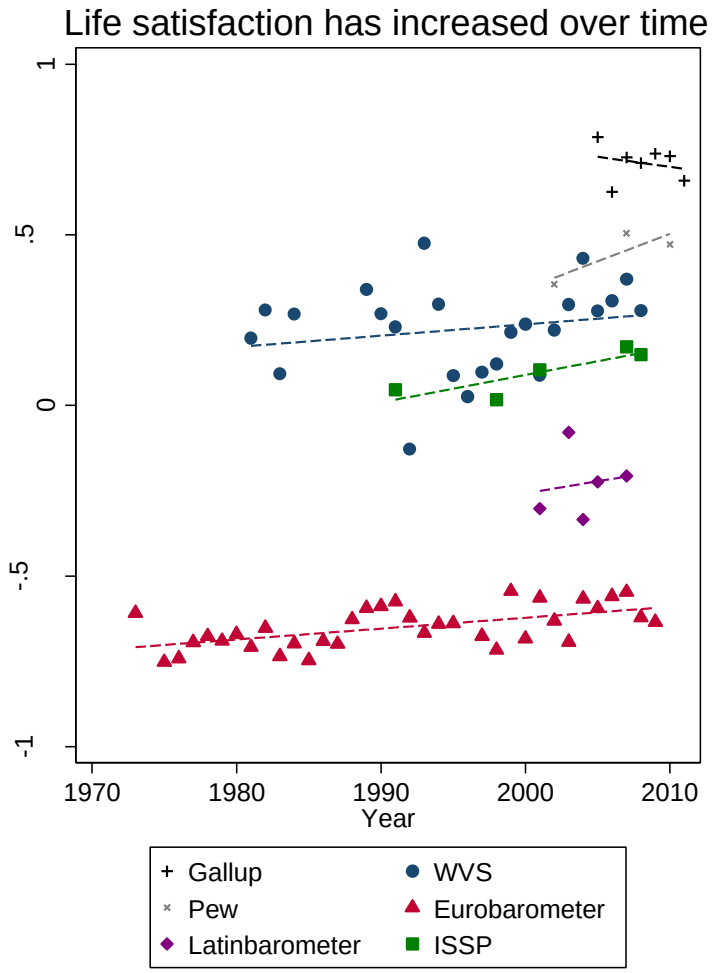
In Table 2, we present our estimates using our panel of panels. In the first row, for comparison, we present the average between-country regression coefficient, obtained by regressing average well-being (for a given country in a given dataset) on average log(GDP) and dataset fixed effects. In row (2), we add country-by-dataset and survey-by-dataset controls. This estimate is essentially a precision-weighted average of the estimates in Table 1. Consistent with our earlier findings, the panel estimate is higher than the between-country estimate, although not statistically significantly so. The 95% confidence interval rules out values of the well-being–income gradient less than 0.16, or about 2/3 the between-country estimate. Adding macro controls changes this conclusion only slightly. When we add controls for the output gap, the coefficient falls to 0.21, and when we replace controls for the output gap with controls for unemployment and inflation, the coefficient is 0.28. Including all of our controls for the macroeconomy yields an estimated gradient of 0.23. Finally, we interact all of our controls with dataset fixed effects and yield a point estimate of 0.22.

In all of our specifications we can reject a null of no relationship between well-being and income in the time series. Moreover, in all of our specifications we fail to reject a null that the estimated gradient is the same in the time series as that seen between countries. Similarly, we are unable to reject a null that the magnitude of the relationship between well-being and income in the time series is between 0.3 and 0.4.

VII. Has Global Well-Being Grown?

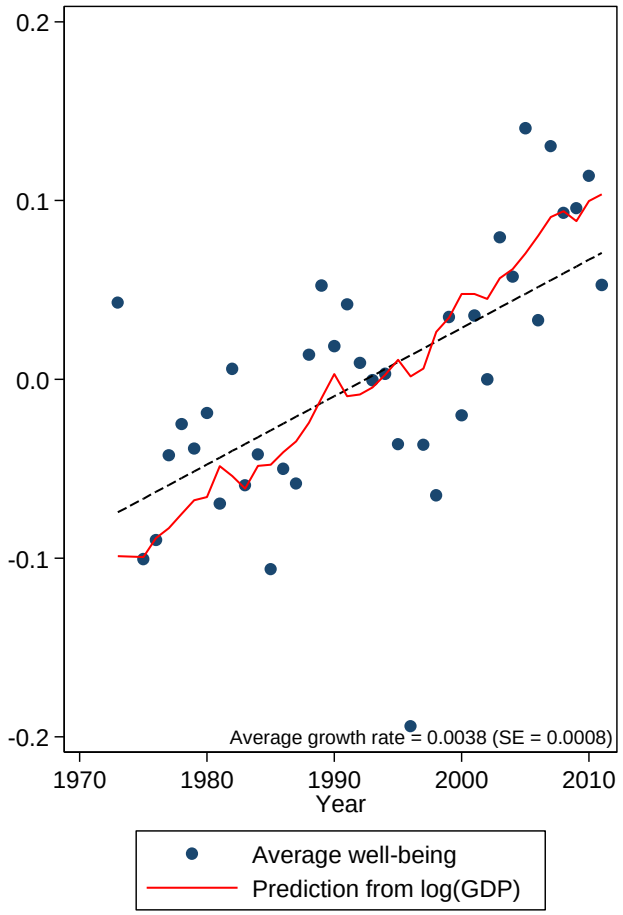
In all of our specifications thus far we have partialled out changes that are common to the world and look for deviations from country and year trends. We now ask whether average well-being in the world has increased since 1973 when our data begin. Although the world has gotten vastly richer over this time period, and we find a clear and positive relationship between well-being and income, that does not necessarily imply that global well-being has grown. First, as we have emphasized elsewhere, there are other factors that impact well-being and thus societal changes, such as increases in pollution, war, and temperature, might have offset the gains from rising GDP. Second, thus far we have shown that rises in GDP that are relatively greater or smaller than the average change in GDP impact well-being, and this would be equally true if income relative to a global reference point determines well-being.

Figure 7



To measure changes in average well-being, we plot in Figure 7 average well-being in each dataset and year over time. To obtain the figure, we regressed well-being on country and year fixed effects in each dataset; we plot the year fixed effects. Five of our datasets show a clear increase in well-being. The exception to this is the Gallup World Poll which covers the period from 2006-11 and finds that world well-being has declined. It is worth noting that the decline only occurs when data for 2011 is included. The slopes of the line suggest a global increase in well-being of around 0.0035 of a standard deviation per year, so that over the past four decades well-being around the world has improved by about 0.14 of a standard deviation.

Figure 8



In Figure 8, we combine estimates from all of our datasets to create average well-being in the world for each year. While the datasets in Figure 7 all show different levels of happiness—and thus highlighting the importance of survey specific effects—combining the surveys should reduce such measurement error. This average does not include country fixed effects, so its level is not informative; indeed we re-center it to have mean zero. We have also plotted the prediction of yearly well-being from log GDP per capita. To obtain this prediction, we regress log GDP per capita on country, dataset, and year fixed effects. We then multiply the year fixed effects by our measure of the between country gradient between well-being and income. (We also re-center this series.)

The figure shows a clear increase in well-being over time. Average well-being appears to have increased by about 0.15 standard deviations over time. The two series are remarkably correlated (ρ=0.66), suggesting that much of the growth in well-being over the last 40 years can

be explained by growth in income. The figure shows considerable variation in well-being around the general increase. Some of this no doubt reflects measurement error, but looking at year-to-year changes, it is possible to see the impact of global recessions on global well-being. For example, there is a clear decline in well-being during the Asian Financial crisis and its aftermath in the 1990s, and well-being ticked downwards from 2007 to 2008 and 2009. These fluctuations are surprising not so much because well-being tracks income, but because the well-being response to recessions vastly exceeds the income response. (See also Wolfers, 2003; Di Tella, MacCulloch and Oswald, 2003).

VIII. Reconciliation with the Literature

In a recent paper, Easterlin et al. (2010) have argued that there is no long-term relationship between growth in well-being and growth in GDP. As their conclusions are very different from ours, we now turn to reconciling our different findings. We begin by describing in detail Easterlin et al.'s methodology.[13] Two factors explain the difference between Easterlin et al.'s results and our own: first, they include many countries with non-representative sampling frames in their analysis; and second, they do not sufficiently account for differences in measured well-being across different survey instruments.

We present our replication and reconciliation in Table 3. In the first row, we report Easterlin et al. (2010)'s published estimate of the long-term relationship between the growth rate of well-being and the growth-rate of GDP. This number is not directly comparable to our estimate of the well-being–gradient for two reasons: first, they measure well-being on a 1-10 scale, and second, they express the growth rate of GDP on a 0-100 scale (i.e. a doubling of GDP is a growth rate of 100, not 1). Throughout this section, we analyze well-being on a 1-10 scale—as they do—rather than our standardized scale. But we change the scale to make it easier assess the point estimates and to make it more comparable to the estimates presented throughout our paper.

As we do, Easterlin et al. (2010) emphasize the long-term relationship between well-being and GDP. They focus on the following specification:

[13] For the sake of clarity we provide more detail than Easterlin et al. have provided in print; much of which we learned from our replication efforts and from correspondence with Richard Easterlin and Laura Angelescu, who were helpful in providing the details necessary to replicate their results.

$$Growth\ rate\ of\ satisfaction_c = \alpha + \beta Growth\ rate\ of\ GDP_c + \epsilon_c \quad (1)$$

That is, β measures the long-term relationship between economic growth and growth in well-being. But there are some differences in our approaches. Both the independent and dependent variables in this equation are generated regressors—trend growth rates measured over long time periods of at least 12 years—while we estimate more standard panel data models. To estimate the long-term growth rate of GDP, Easterlin et al. take the growth rate of GDP over the time period for which they have well-being data. This is reasonably straightforward.

But their estimates of trend growth in well-being are worth examining in more detail. First, they combine data from multiple surveys (indexed by s), including the Latinobarometro and the World Values Survey. Second, while different sources measure life satisfaction on different scales, Easterlin et al. (2010) rescale satisfaction scores to a 1-10 scale. Having combined these data and put them on a (hopefully) comparable scale, they then estimate the long-term growth rate of life satisfaction by running the following regression separately for each country in their sample:

$$\begin{aligned} Satisfaction_{cts} \quad (2) \\ = \alpha_c + \delta_c Year_t + \gamma_c I[Satisfaction\ measured\ on\ a\ 1-10\ scale_s] \\ + \epsilon_{cts} \end{aligned}$$

In this equation, δ_c is interpreted as the growth rate of well-being, which is the independent variable in their main analysis. Notice that they take account of differences in well-being measures across surveys, but only in a rather crude way. They allow for differences in the average response to a well-being question—depending on whether it is a 4-point question, or a 1-10 scale—but otherwise assume that there are no other differences across surveys. However, not only are there many other differences in survey questionnaires (the questions differ, the survey mode differs, the question order differs), but these are likely also correlated with the survey year, since the different surveys used are typically non-overlapping—one is used for early years and a second for later years. Moreover, they allow this correction factor to be estimated separately for each country.

Their sample consists of the 37 countries for which they could construct a time series of life satisfaction spanning at least one complete business cycle. Importantly, their sample includes observations based on non-representative sampling frames, while we systematically

exclude these cases.[14] This presents an important problem, because typically, the unrepresentative samples were taken in earlier years (when GDP was lower, and hence it is likely that so too was well-being). Moreover, the non-representative samples were typically of more affluent (and hence happier) sub-groups of the population. This has the effect of making measured well-being look higher in the earlier years of the survey, when GDP was lower. That is, not only do changes in sampling frames make the well-being data non-representative, they create measurement error that is likely correlated with GDP.

We now turn to reconciling our findings, in Table 3. In row 2 of Table 3, we follow Easterlin et al. exactly and, using their data, replicate their published finding exactly (after rescaling the growth rate of GDP to be on a 0-1 scale). In the second row, we keep their specification identical, but now drop all country-years with non-representative sampling frames.

Table 3: Reconciliation with Easterlin et al. (2010)

Specification	Coefficient on growth rate of GDP
1. Easterlin et al. (2010) as published (0-100 scale)	-0.003 (0.0019)
2. Easterlin et al.'s data and specification (replication times 100)	-0.33 (0.20)
3. Easterlin et al.'s data and specification, drop nonrepresentative observations	0.07 (0.16)
4. Easterlin et al's data, drop nonrepresentative observations, include survey fixed effects	0.54 (0.18)
5. Our data, Easterlin's countries and specification	0.00 (0.31)
6. Our data, Easterlin's countries and specification, drop nonrepresentative observations	0.38 (0.26)
7. Our data, all countries with T>=12, drop nonrepresentative observations, Easterlin's specification	0.67 (0.32)

Notes: The numbers reported in row 1 are from Easterlin et al. (2010), figure 2. To obtain the standard error, we divided their point estimate by their reported t-statistic. In rows 2-4, we use aggregate (country-year-dataset) data provided to us by Laura Angelescu. We follow the procedure described in Easterlin et al. (2010), with the indicated modofications. In rows 5-7, we use our own microdata, pooling data from the World Values Survey, Latinobarometro, and Eurobarometer, but follow Easterlin et al.'s methodology. For more details, see the text.

[14] With one exception: when estimating equation (2) for China, they include a dummy for the years in which the sampling frame was urban. In years in which the Latinobarometro is nonrepresentative, they limit attention to people living in places of 100,000 people or more, but this clearly does not fix the problem, which is that these areas are overrepresented anyway. Our Data Appendix provides details of which country-year-datasets have non-representative sampling frames.

The estimated coefficient remains negative, but is about half as big. In row 3, we drop the non-representative observations and add dataset fixed effects—as we do in our main analysis—rather than their γ_c correction for scale differences. This results in a positive, albeit statistically insignificant coefficient, but we are now unable to reject a coefficient similar in magnitude to what is seen in the cross-section. The reason this has such a large impact on the estimated coefficient comes down to six Eastern European countries for which Easterlin et al. (2010) use data from a special 2004 supplement to the Eurobarometer, combined with data from the World Values Survey. Their estimate depends on being able to make a direct comparison of the change in well-being from one survey to the next. The practical effect of including dataset rather than response number fixed effects is that the country*dataset fixed effects effectively mean that we drop these six observations.[15]

There are in fact a lengthy list of differences between Easterlin et al's methodology and our own, despite our common interest in the long-term relationship between well-being and growth. For example, Easterlin et al. (2010) compute GDP slightly differently than we do: they match GDP to SWB not by the survey calendar year but by the most likely relevant period. So if a survey occurred in January of 1990, they use GDP data from 1989, but if it occurred in September 1990, they use 1990 GDP data. Either of these are arguably correct approaches and it turns out that these small differences have little impact on our results.

Finally, while we use data only from datasets spanning many countries, Easterlin et al. combine data from many individual time series. In addition to using the WVS, EB, and LB, they also use data from the General Social Survey for the USA, from the Life in Nation surveys for Japan, and from Ottar Helvik's survey for Norway.

To show that none of these differences are material to the divergence between our results and Easterlin et al.'s , we use our own data to demonstrate the importance of non-representative sampling frames. Our data here consists of pooled data from the World Values Survey, the Latinobarometero, and the standard Eurobarometer, so the survey dummy variable is identical to the question order dummy. In row 5 of Table 3, we follow Easterlin et al.'s *procedure* exactly, selecting the same set of countries and using the same methods to compute the growth rates of GDP and satisfaction, and include the nonrepresentative observations as they do. We too find a

[15] In dropping these observations, we re-calculate the growth rate of GDP to reflect the new sample period for the affected countries.

negative and insignificant value of β. In row 6, we throw out the countries with non-representative sampling frames, and our estimate of β rises to 0.44, close in magnitude and not significantly different from the value we obtain from Easterlin's data. Although close in magnitude to our preferred estimate of β (after accounting for the different scale), this estimate is not significantly different from zero. When we include the full sample of countries for which we have at least 12 years of life satisfaction data, the estimated coefficient rises to 0.72 and is statistically significantly different from 0.

We have shown that the differences between our results and Easterlin's stem from the decision to include controls for survey-specific fixed effects, and to use country-years with non-representative sampling frames. We suspect that it is uncontroversial to claim that analysts should not use non-representative data and changing sample frames to estimate national aggregates such as the growth rate of well-being. It is perhaps less obvious that one should control for survey-specific fixed effects. We have argued already that different surveys will induce different levels of well-being, but there is no reason to expect this measurement error to be correlated with GDP or other variables of interest. In a large sample, ignoring survey fixed effects is unlikely to much affect results. Easterlin et al.'s samples, however, are miniscule. The critical six countries affected by the choice of fixed effects each have time series with four observations. Mis-measured well-being in a single observation can (and in fact does) have an enormous influence on the growth-rate of well-being. This fragility is the cost of their rather unusual two-step procedure, compared with our panel regressions.

One final comment is in order. In throwing out non-representative observations, and dummying out data for some countries, we have shortened the time series underlying the estimation of growth rates of well-being and GDP. Easterlin et al.'s focus was on creating sufficiently long time series, so one might worry that what really drives the difference in results between rows 2 and 4 is this shortening. But the results presented using our own data do not suffer from this problem: for example in row 7 we explicitly limit the sample to countries with at least 12 years of well-being data, and here we find the strongest results.

IX. Conclusion

- .

2. References

Acemoglu, D. and D. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Volume 4, Amsterdam: Elsevier-North Holland.

Blanchflower, D. G. and A. J. Oswald. 2004. "Well-Being over Time in Britain and the USA." *Journal of Public Economics*, 88(7-8), pp. 1359-86.

Bernanke, B. S. 2010. "The Economics of Happiness." Speech presented at the University of South Carolina Commencement Ceremony, Columbia, South Carolina.

Brickman, P.; D. Coates and R. Janoff-Bulman. 1978. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology*, 36(8), pp. 917-927.

Cantril, H. 1965. "Evaluating the Probable Reactions to the Landing in North-Africa in 1942 - a Case-Study." *Public Opinion Quarterly*, 29(3), pp. 400-12.

Caporale, G. M.; Y. Georgellis; N. Tsitsianis and Y. P. Yin. 2009. "Income and Happiness across Europe: Do Reference Values Matter?" *Journal of Economic Psychology*, 30(1), pp. 42-51.

Clark, A. E. and A. J. Oswald. 1996. "Satisfaction and Comparison Income." *Journal of Public Economics*, 61(3), pp. 359-81.

Clark, A. E. 2008. "Happiness, Habits and High Rank: Comparisons in Economic and social Life." PSE, Discussion Paper No. 2008-61.

Clark. A. E.; N. Kristensen and N. Westergaard-Nielson. 2009. "Economic Satisfaction and Income Rank in Small Neighborhoods." Journal of the European Economic Association, 7, pp. 519-527.

Clark, A. E.; P. Frijters and M. A. Shields. 2008. "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles." *Journal of Economic Literature*, 46(1), pp. 95-144.

Clark, A. E. 2010. "Where Should we Live?" *Review of Economic Analysis*, 2(1), pp. 20-31.

Deaton, A. and D. Kahneman. 2010. "High income imprvoes evaluation of life but not emotional well-being." *Proceedings of the National Academy of Sciences*, 107(38), pp. 16489-16493.

Deaton, A. 2008. "Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll." *Journal of Economic Perspectives*, 22(2), pp. 53-72.

Deaton, A. 2011. "The Financial Crisis and the Well-Being of Americans." NBER working paper No. W17128.

Di Tella, *Rafael*, Robert MacCulloch, and Andrew Oswald 2001. "Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness." *American Economic Review* 91(1).

Di Tella, R.; R. J. MacCulloch and A. J. Oswald. 2003. "The Macroeconomics of Happiness." *Review of Economics and Statistics*, 85(4), pp. 809-27.

Diener, E.; E. Sandvik; L. Seidlitz and M. Diener. 1993. "The relationship between income and subjective well-being: Relative or absolute?" *Social Indicators Research* 28, pp. 195–223.

Diener, E. and R. Biswas-Diener. 2002. "Will Money Increase Subjective Well-Being?" *Social Indicators Research*, 57(2), pp. 119-69.

Diener, E. and S. Oishi. 2000. "Money and happiness: Income and subjective wellbeing across nations" in E. Diener and E.M. Suh (eds.), *Subjective Well-being across Cultures* (MIT Press, Cambridge, MA).

Diener, E. 1984. "Subjective Well-Being." *Psychological Bulletin*, 95(3), pp. 542-75.

Duncan, O. D. 1975. "Does Money Buy Satisfaction." *Social Indicators Research*, 2(3), pp. 267-74.

Easterlin, R. A. "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," In P. A. David and M. W. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz* (New York and London: Academic Press, 1974).

Easterlin. R. A. 1973. "Does Money Buy Happiness." *Public Interest*, (30), pp. 3-10.

Easterlin, R. A. 2003. "Explaining Happiness." *Proceedings of the National Academy of Sciences*, 100(19), pp. 11176-11183.

Easterlin, R. A. 2005. "Feeding the Illusion of Growth and Happiness: A Reply to Hagerty and Veenhoven." *Social Indicators Research*, 74(3), pp. 429-43.

Easterlin, R. A., and L. Angelescu. 2009. "Happiness and Growth the World Over: Time Series Evidence on the Happiness-Income Paradox." IZA Discussion Paper No. 4060.

Easterlin, R. A.; L. A. Mcvey; M. Switek; O. Sawangfa and J. S. Zweig. 2010. "The Happiness-Income Paradox Revisited." *Proceedings of the National Academy of Sciences*, 107(52), pp. 22463-68.

Easterlin, R. A. 1995. "Will Raising the Incomes of All Increase the Happiness of All." *Journal of Economic Behavior & Organization*, 27(1), pp. 35-47.

Frank, R. H. 2005. "Positional Externalities Cause Large and Preventable Welfare Losses." *American Economic Review*, 95(2), pp. 137-141.

Frey, B. S. and A. Stutzer. 2002. "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, 40(2), pp. 402-35.

Gardner, J. and A. J. Oswald. 2001. "Does Money Buy Happiness? A Longitudinal Study Using Data on Windfalls." Mimeo. Warwick University.

Graham, C. 2008. "Happiness and Health: Lessons - and Questions - for Public Policy." *Health Affairs*, 27(1), pp. 72-87.

Graham, C. and S. Pettinato. 2002. "Frustrated Achievers: Winners, Losers and Subjective Well-Being in New Market Economies." *Journal of Development Studies*, 38(4), pp. 100-40.

Hagerty, M. R. 2000. "Social Comparisons of Income in One's Community: Evidence from National Surveys of Income and Happiness." *Journal of Personality and Social Psychology*, 78(4), pp. 764-71.

Hagerty, M. R. and R. Veenhoven. 2003. "Wealth and Happiness Revisited - Growing National Income Does Go with Greater Happiness." *Social Indicators Research*, 64(1), pp. 1-27.

Inglehart, R.; R. Foa; C. Peterson and C. Welzel. 2008. "Development, Freedom, and Rising Happiness a Global Perspective (1981-2007)." *Perspectives on Psychological Science*, 3(4), pp. 264-85.

Kahneman, D. and R. Thaler. 1991. "Economic Analysis and the Psychology of Utility: Applications to Compensation Policy." *American Economic Review*, 81(2), pp. 341-346.

Layard, R. 2005. *Happiness: Lessons from a New Science*. London: Penguin.

McClendon, M. J. and D. J. O'Brien. 1988. "Question-Order Effects on the Determinants of Subjective Well-Being." *Public Opinion Quarterly*, 52(3), pp. 351-364.

Oswald, A. J. 2010. "Emotional Prosperity and the Stiglitz Commission."IZA Discussion Paper No. 5390, Institute for the Study of Labor.

Oswald, A. J. 1997. "Happiness and Economic Performance." *The Economic Journal*. 107(445). pp 1815-1831.

Sacks, D.; B. Stevenson and J. Wolfers. 2010. "Subjective Well-Being , Income, Economic Development and Growth," in World Bank volume, *Development Challenges in a Post-Crisis World*.

Schyns, P. 2002. "Wealth of Nations, Individual Income and Life Satisfaction in 42 Countries: A Multilevel Approach." *Social Indicators Research*, 60(1-3), pp. 5-40.

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

Stevenson, B. and J. Wolfers. 2010. "Inequality and Subjective Well-Being." Working paper.

Stiglitz, J. E., A. Sen, and J. Fitoussi. 2010. "Report by the Commission on the Measurement of Economic Performance and Social Progress." Technical Report (September).

Strack, F.; L. L. Martin and N. Schwarz. 1988. "Priming and Communication - Social Determinants of Information Use in Judgments of Life Satisfaction." *European Journal of Social Psychology*, 18(5), pp. 429-42.

Veenhoven, R. 2011. "World Database of Happiness Example of a focused 'Findings Archive'," Working Paper Series of the German Council for Social and Economic Data 169, German Council for Social and Economic Data