Real World Happiness

By Helen Johns and Paul Ormerod

Introduction

Happiness economics has generated an entire new academic industry. Over 10,000 articles have now been published on the concept of happiness, or subjective well-being (the two terms are used interchangeably).

Surveys on the levels of happiness reported by individuals have been carried out over a few decades in most Western countries. The recorded levels of happiness fluctuate from year to year, but in general there is no trend, either up or down. Over the same period, average material standards of living, measured by real gross national product (GNP) per head, have shown a very clear upward trend.

This finding is repeated endlessly and appears to have made an impression on many people. We see the level of happiness over time rumbling along showing no trend. In contrast, there is GNP per head bounding ahead, soaring into the stratosphere. Surely this proves that economic growth is not making us happier?

Time series data show that nations do not get happier over time as they get richer. In contrast, happiness is positively correlated with individual income within a given country at any point in time; the rich generally report greater happiness than the poor. This, the so-called Easterlin paradox, named after the doyen of happiness studies, Richard Easterlin, is also discussed at length in the happiness literature. An implication which is widely drawn is that if we do not get happier as we get richer, this effect must be due to the pernicious psychological effects of inequality.

In conjunction, these findings have been used as the basis for wide reaching policy recommendations. For example, taxation should be more progressive, and indicators of self-reported happiness should be used in formal government policy appraisal, to supplement or even replace economic indicators.

The fact that measured happiness has not increased over decades is viewed by some commentators as indicating a flaw in our society which must be corrected through government intervention. As increasing happiness is a self-evident good, who but the most irredeemable misanthrope could object to such an end?

But scepticism about the use of happiness evidence in policy-making does not mean that the holder of this view is automatically a fanatical believer that economic agents always behave rationally, or that maximising GNP is all that matters. Or, for that matter, that inequality is an irrelevance and we should revive Victorian workhouses for the poor. Neither of us believes any of these things. The question for us is the scientific validity of happiness research, most specifically any findings based on time series.

There are at least two alternative interpretations to the mainstream view that happiness has remained flat over decades because economic growth does not make us happier. First, we could conclude from this flat trend that attempting to improve the human lot through *any* policy – not just through pursuing economic growth - is entirely futile. Second, and alternatively, that happiness data over time shows little movement because it is an extremely insensitive measure of welfare.

We argue that the evidence points to the latter. This can be demonstrated both from empirical reasoning and by examining the mathematical properties of the measure itself.

Above all, we argue that average happiness time series are, by construction, incapable of conveying useful information on the level of overall social wellbeing, and their use should therefore be rejected by policy-makers and social scientists.

Why time series data on happiness tells us nothing

First of all, the lack of correlation over time between measured happiness and the size of the economy, so widely mentioned, is a wholly misleading argument. This lack of correlation extends to a wide range of variables, a fact which attracts far less publicity.

For example, using UK data from 1973 onwards, there is no correlation between self-reported life satisfaction and either real current public expenditure or lower hours of work. In the US, life expectancy for whites rose from 72.0 years in 1972 to 78.0 in 2003. For blacks, the increase was even higher, from 64.6 to 72.7, representing not merely an absolute rise, but a narrowing of the gap with whites. Gender inequality as measured by the median earnings of women compared to men has fallen sharply. In 1972, women earned 58 per cent of men, rising to 75 per cent in 2003. Yet there was no correlation between happiness and any of these improvements.

In fact, dramatic rises in inequality in both the United States and UK had no impact on happiness, as shown for the US in Figure 1.

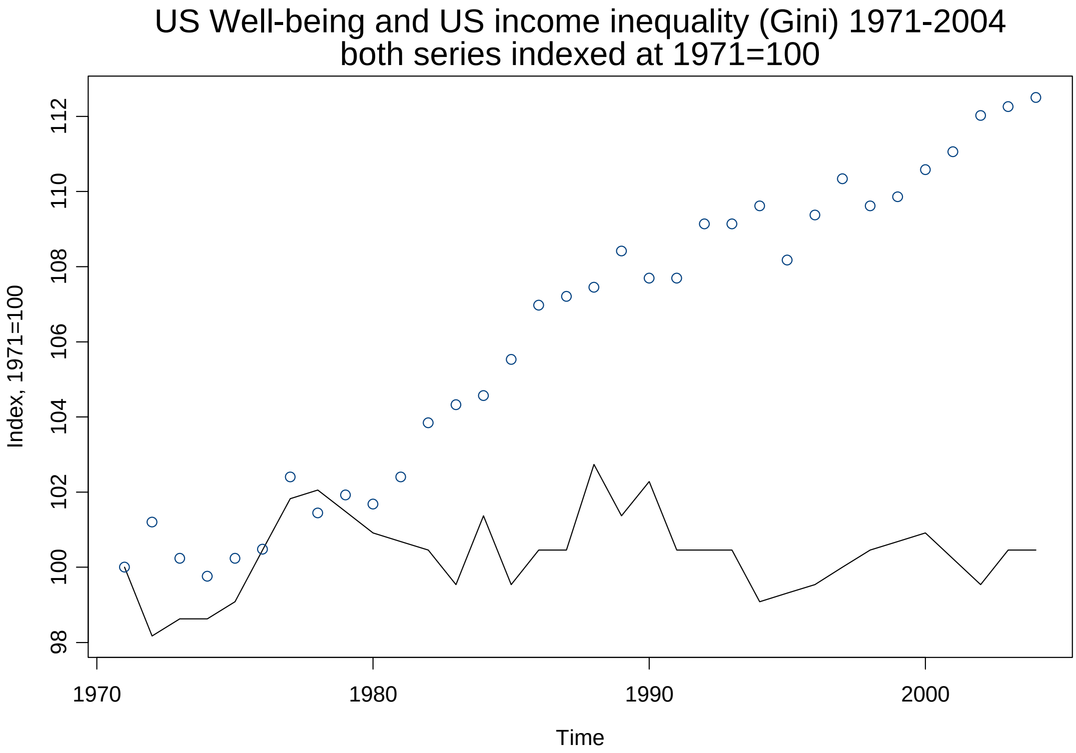

Figure 1: US well-being and Gini coefficient 1971-2004, both indexed at 1971=100. Solid line is well-being, dotted line is the Gini coefficient

This chart seems to rather undermine the emphasis which many happiness advocates place on the adverse effects of inequality on happiness. Most emphatically, this is not to say that inequality cannot possibly have adverse effects on individuals; there are much more soundly based scientific findings which show this in areas such as health, for example. But this is a clear case where the concept of well-being confuses rather than clarifies the issue.

Wide publicity has been given in the UK to the apparent large rise in the number of depressed people in the population. Indeed, the government has taken note and is investing large sums to try to deal with this problem. However, the UK happiness data show no signs of reflecting the claimed increase in depression. This is surely something which, if it is correct, must show up in the happiness data.

So there is no correlation in time series data between reported happiness levels and a whole series of factors which might reasonably be thought to affect well-being: income, public spending, longevity, gender equality, income inequality – even the incidence of depression in a population.

Indeed, if we were to attach any import to this evidence, we would be forced to conclude that measured happiness shows that sixty years' economic and political labours of all descriptions since World War Two have made no difference to the welfare of the citizens of the Western world.

However, in examining the reasons why average happiness is flat it is important to examine the way in which happiness is measured. People are asked to register their level of happiness on a scale of *n* categories (e.g. 1 = 'not happy', 2 = 'fairly happy' or 3 = 'very happy'). These numbers are then averaged over the population to gain an overall happiness score. Discrete categories mean that people have to undergo large discrete change in their happiness in order for this to be registered by the indicator; and once they have reached the top category they officially can't experience any further increase in their happiness. As a consequence, noticeable changes in average happiness can only come about through substantial numbers of people moving category.

As a general rule, if the happiness of 1% of the population (net) increases enough for them to place themselves in the next category, the average happiness score increases by 0.01. For example, happiness surveys on a 3-category scale in the US typically yield an average happiness of about 2.2. In order for the measure to undergo a 10% increase, 22% of the population would have to undergo a substantial enough increase in their happiness for them to shunted up to the next category.

It is very difficult to think of a set of circumstances in which 22% of the population would find themselves moving from, say, 'fairly' to 'very' happy over the space of a few years, particularly as genes and formative experiences play a large role in determining someone's happiness. It is therefore not surprising that we observe average happiness to be sluggish compared to other social or economic indicators such as GNP.

Furthermore, *by construction*, the happiness data can exhibit no indefinite trend. As individuals answer a survey in which they are asked to state their own level of happiness on an n-point scale, the data is therefore bounded between one and n. Over any particular short period of time, an apparent trend either up or down might exist, but by definition it cannot persist. In contrast, at least as it is presently defined, real GNP can exhibit no upper bound. Indeed, for the past 200 years it has shown a persistent trend increase.

This means that we have to exercise extreme caution in drawing any inferences from the correlation, or rather the lack of it, between time series data on well being and real GNP. From a statistical perspective, *any* calculation of a correlation between a variable which exhibits a trend and one which does not is fraught with inherent problems. (In technical terms, by definition time series happiness data is integrated of order zero, and GDP is integrated of order one).

The difficulties due to the inherent properties of the time-series happiness data would make it problematic were it to be used in policy. If a time series measure of well-being were to become used as a basis for policy, governments would succumb to an irresistible urge to try to influence its level. In such circumstances, it would be essential that the data should contain real information. Unfortunately, this is not the case.

Time series happiness data is in general indistinguishable from a purely random series. The autocorrelation function is flat and has no statistically significant individual values. In turn, this implies that it not possible to carry out systematically accurate forecasts of this variable[1].

Furthermore, we do not know what variables have influenced in a systematic way the movements in well being over the past. Note that even if we did, this would still not imply that the series could be successfully predicted. The variables which exercised a systematic influence would themselves have to be capable of being predicted.

So what causes variation in the happiness time series? One of the present authors, Helen Johns, has carried out original analysis which shows that the variations which we observe in measured happiness are completely consistent with the view that they are simply fluctuations based on sampling error. Her short mathematical paper is available on request (general.hj@googlemail.com), and here we try to give a flavour of the analysis. The particular difficulty in explaining the results is that the happiness index is based on discrete categories (0, 1,…, n), so the sampling error probability distribution associated with it is also discrete. This makes the analysis mathematically complicated.

The time series happiness data is based on surveys. For, example , the American survey is based on a survey of around 1,500 people. This is a sufficiently large sample to be reasonably representative of the population as whole. But by the very fact of being a sample, it is not, except by the purest of coincidences, exactly the same as the population as a whole.

So we will observe fluctuations from survey to survey which arise simply because of sampling error. How big are these compared to the fluctuations which we actually observe?

A way of conveying the algebraic results is to examine numerical results for the sampling error distribution associated with the kinds of sample sizes and population characteristics which are typical of happiness surveys.

The standard US survey, for example, asks people to place themselves in one of three categories (1 = not very happy, 2 = fairly happy, 3 = very happy). As already noted, this is usually conducted over a sample of about 1500 people. The results of US surveys seem to indicate that the proportions of the population in each category are roughly 12%, 55% and 33% respectively. The actual happiness score which would pertain in the absence of sampling error under these conditions is 2.21.

The probability distribution of the happiness value for a sample of 1500 drawn from a population with these 12%, 55%, 33% probabilities of being in each category is shown in Figure 2.

---

[1] If the series exhibited long memory, this would not necessarily be the case. But *many* more data points are required before it could be established whether the data exhibit long memory.

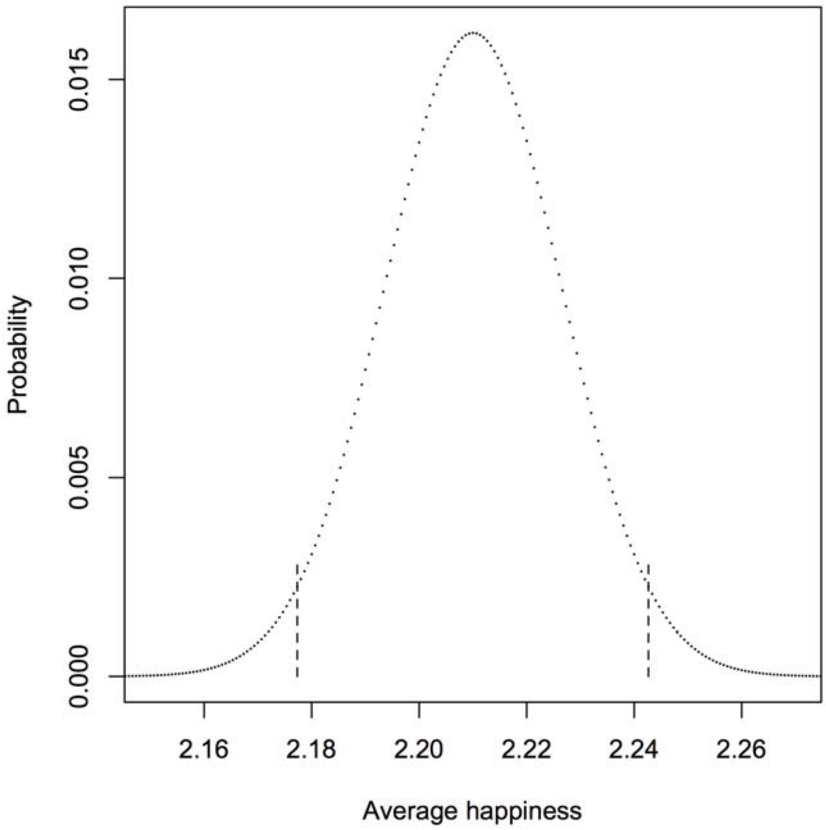

Figure 2: *The probability of obtaining each average happiness value given a sample of 1500 respondents and a probability of being in category 1 of 0.12, in category 2 of 0.55, and category 3 of 0.33. The dashed lines show the 95% confidence limits.*

In other words, Figure 2 shows the following:

- If the measured happiness score across the US populations as a whole really were 2.21, what the probability is of observing not just this but other values in a survey of 1500 people

We can see that 95 per cent of the time the survey will give a happiness level in a range of just under 2.18 to just over 2.24. Now, most of the actually recorded levels of happiness in the US are within this range. In other words, most of the annual movements in recorded happiness which we observe could arise simply from sampling error. And by definition such movements convey no true information.

Another way of looking at the lack of true information in the data is to imagine that something truly wondrous were to happen and that the sum total of human happiness was indeed augmented. A wise and perspicacious policy was implemented which caused a whole 6 million people in the USA (roughly 2% of the population) to undergo such a dramatic change in their personal happiness that they started to describe themselves as “fairly” rather than “not very happy”. The average happiness of the population would now be 2.23 rather than 2.21. The resulting probability distribution is shown in Figure 3, superimposed on that before the change.

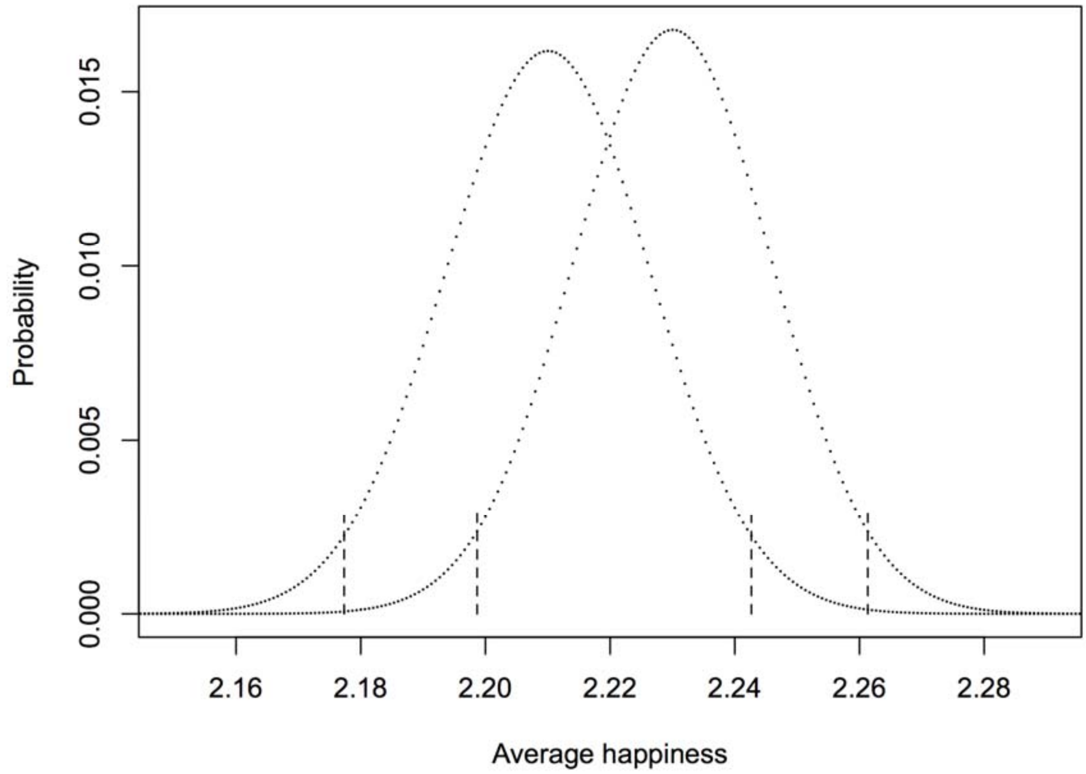

Figure 3: The probability of obtaining each average happiness value given a sample of 1500 respondents and: (left) a probability of being in category 1 of 0.12, in category 2 of 0.55, and category 3 of 0.33; (right) a probability of being in category 1 of 0.10, in category 2 of 0.57, and category 3 of 0.33. The dashed lines show the 95% confidence limits.

There is a substantial overlap between the two curves. In fact, if we observed two points next to each other in a US happiness time series with the values 2.21 and 2.23 we could not be particularly confident that they were in fact different and there had been any actual change in the happiness of the population.

The large uncertainty which exists even if millions of people *were* to experience a genuine large increase in their happiness indicates the inherent insensitivity of this measure.

The resultant effect on the happiness time series of any real change could easily be drowned out by statistical noise. Any happiness-increasing policy effect would have to be of long duration, not be offset by countervailing trends in society, and be produced by kind of benefit which is not quickly adapted to, in order for it to be perceptible in the time series.

It is in fact easy to show, as Helen Johns' technical paper does, that the happiness data contains about as much information on the level of overall social well-being as a series of random numbers drawn from an appropriate probability distribution.

This startling finding raises serious questions over the validity of happiness time series and their ability to contribute useful evidence to social science.

Policy implications and concluding remarks

The originators of GNP never insisted that this was the *only* way of measuring an economy. In his Nobel lecture[2], for example, Kuznets specifically discussed the social implications of growth and argued that: 'Many of these are of particular interest, because they are not reflected in the current measures of economic growth; and the increasing realization of this shortcoming of the measures has stimulated lively discussion of the limits and limitations of economic measurement of economic growth'.

Politicians of all parties in all democratic countries already take into account a broad range of factors when they are making decisions. They are not simple GNP maximisers and they do not need an additional measure of 'well-being' to force them to consider policy objectives other than the purely economic.

Indeed, the official British government guidelines on policy appraisal, the Treasury's *Green Book*[3], clearly states that: "wider social and environmental costs and benefits for which there is no market price also need to be brought into any [policy] assessment" and that the inclusion of "non-market impacts is a challenging but important element of appraisal, and should be attempted wherever feasible". The extent to which formal policy processes are weighted towards maximizing GNP have been exaggerated.

Even if this were not so, wellbeing evidence is currently not robust enough to guide policy-making. The British government recently commissioned a group of academics led by Paul Dolan of Imperial College, London, himself a distinguished well-being researcher, to survey the literature. The results are published in the February 2008 *Journal of Economic Psychology*[4]. Here is what they concluded:

"One very firm conclusion that can be drawn from our review is that the existing evidence base [for well-being] is not quite as strong as some people may have suggested….This, in addition to lack of clear evidence on causality, makes it difficult to make clear policy recommendations at this stage."

This message will be very disappointing to many, but the point about causality is a very useful one; some of the conclusions which have been drawn from time series

---

[2] S. Kuznets, (1971), 'Modern Economic Growth: Findings and Reflections', Nobel Prize Lecture available at: www.nobelprize.org/nobel_prizes/economics/laureates/1971/kuznets-lecture.html

[3] Her Majesty's Treasury (2003), *Green Book: Appraisal and Valuation in Central Government*, available on the Treasury website

[4] P Dolan, T Peasgood and M White (2008), 'Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being', *Journal of Economic Psychology*, 29, 94-122

data have relied too heavily on assumptions about the direction of causation. In addition, the more credible results from happiness research seem to come from treating happiness scores as **ordinal** - i.e. using ordered models on the probability of an individual placing him or herself in particular category - rather than as cardinal numbers which can be averaged over entire populations.

Such analysis does produce intuitively sensible results, such as stable family life, being married and good health, contributing to happiness, while chronic pain, divorce and bereavement detract from happiness. These results, while consistent with everyday experience, don't however really tell us anything we didn't know already.

Our inexorable conclusion is therefore that society-wide happiness time series should be abandoned as they don't tell the social scientist anything useful; in addition, the flatness of happiness time series most certainly cannot be pinned on the economic system, and neither do they point to some kind of social aberration in need of government correction.

Average happiness has shown demonstrably stubborn flatness despite vastly differing government styles and levels of inequality, and it is seriously misleading to argue that, armed with the 'insights' of time-series happiness research, government intervention is going to make society measurably happier.

END