AUTHOR CONTACT DETAILS

SOCIAL INDICATORS RESEARCH SUBMISSION

July 22 2016

Bruce Headey

b.headey@unimelb.edu.au

Melbourne Institute of Applied Economic and Social Research

University of Melbourne

Parkville

Australia 3052

Ruud Muffels

School of Social and Behavioral Sciences

Tilburg University

Netherlands

Ruud.j.Muffels@Tilburguniveristy.edu

** program file for results: c:bruce\germany\soepconf2014b.do

TOWARDS A THEORY OF MEDIUM TERM LIFE SATISFACTION: Two-way causation partly explains persistent satisfaction or dissatisfaction

(author details removed)

(Final revision for SIR Sept/Oct 2015)

TOWARDS A THEORY OF MEDIUM TERM LIFE SATISFACTION: Two-way causation partly explains persistent satisfaction or dissatisfaction

Abstract

Long term panel data enable researchers to construct trajectories of Life Satisfaction (LS) for individuals over time. In this paper we analyse the trajectories of respondents (N=3689) in the German Socio-Economic Panel who recorded their LS for 20 consecutive years in 1991-2010. Previous research has shown that at least a quarter of these respondents recorded substantial long term changes in LS (Headey, Muffels and Wagner, 2010, 2013). In this paper, graphs of LS trajectories, and subsequent statistical analysis, show that respondents tend to spend multiple consecutive years above and, in other periods, below their own 20-year mean level of LS. They experience extended 'good times' and extended 'bad times'. These results are contrary to set-point theory which views LS as stable, except for short term fluctuations due to life events.

In the later part of the paper we attempt to move towards a theory of medium term life satisfaction. We estimate structural equation models with two-way causation between LS and variables usually treated as causes of LS, including health, physical exercise, frequency of social activities, and satisfaction with work and leisure. Results are interpreted as showing positive feedback loops between these variables and LS, accounting for extended periods of high or low LS.

The models are based on a modified concept of 'Granger-causation' (Granger, 1969). The main intuition behind Granger-causation is that if x can be shown to be statistically significantly related to y in a model which includes multiple lags of y, then it can be inferred that x is one cause of y.

Keywords: life satisfaction trajectories; set-point theory; two-way causation; positive feedback loops; Granger-causation

INTRODUCTION

Trajectories of LS: towards a theory of medium term change

The availability of long-term panel data means that researchers can now construct individual trajectories of LS over time. The dataset which provides the longest available time series is the German Socio-Economic Panel (SOEP). In this paper we show graphs of the trajectories of individual respondents who recorded their LS for twenty consecutive years from 1991-2010. Visual inspection of the data, and subsequent statistical analysis, make it clear that for many individuals LS is highly volatile. Furthermore, many spend extended periods of time – multiple consecutive years - in which their LS is either well above or well below their own long-term (20-year) mean. They go through 'good times' and 'bad times'. Also, at least a quarter of the sample, as reported in previous papers, appear to record long-term, more or less permanent changes in LS, so that in more recent survey years they are substantially more (or less) satisfied with life than they were in earlier years (Headey, Muffels and Wagner, 2010, 2013). Our results provide further evidence that the *set-point theory of LS* (Brickman and Campbell, 1971; Lykken and Tellegen, 1996), which views individual LS as oscillating around a long-term stable set-point, can no longer be sustained. Set-point theory is primarily a theory of stability. It is suggested that the major challenge for researchers now is to develop a theory which can account for medium and long term change (Sheldon and Lucas, 2014).

Most of the paper involves taking some preliminary steps towards a revised theory. We re-examine relationships between LS and variables usually treated as determinants or causes of LS: health, exercise, frequency of socialising with friends, relatives and neighbours, and satisfaction with work and leisure. In reviews of LS research it is quite common to point out that variables like these could just as well be consequences of LS, or perhaps both causes and consequences (Diener, 1984; Diener, Suh, Lucas and Smith, 1999; Frey and Stutzer, 2002). However, in empirical work the possibility of two-way causation is usually ignored. In the case of the link between health and LS this is plainly a mistake. It has been shown in prospective studies that happy people live longer, which could only happen if happiness promotes better health (Deeg and van Zonneveld, 1989; Headey, Hoehne and Wagner, 2014). Although a priori reasoning is not so compelling in relation to other possible two-way linkages, it is surely plausible to hypothesise that happiness could cause as well as be caused by frequency of exercise, social participation and satisfaction with work and leisure.

The evidence in this paper suggests that, in the LS field, *two-way causation is pervasive*.

Two-way causation – or what might be termed *positive feedback loops* – appears to account for medium term (multi-year) periods of high or low LS. For example, changes in health, occurring in year t, are estimated to produce changes in LS in year t+1. Then, in a positive feedback loop, changes that have occurred in LS are positively associated with changes in health in year t+2, which in turn are associated with LS at t+3.

Estimation of structural equation models with two-way causation presents considerable difficulties and remains a controversial procedure in the social sciences (Wooldridge, 2010; Kline, 2011). However, as we shall see, the models in this paper appear to have very satisfactory fits to the SOEP data on which they are based. It is hoped that they give some preliminary insights into variables and a *feedback process* that contribute towards explaining medium term periods of high or low LS.

Challenges to set-point theory

Set-point theory was developed by psychologists over a period of about thirty-five years, appearing to integrate an increasing range of results and becoming more or less the scientific paradigm within which LS researchers worked (Kuhn, 1962). In its original incarnation it was termed adaptation theory; its central claim being that individuals adapt back to their previous 'baseline' level of LS within a year or two of experiencing a major life event (Brickman and Campbell, 1971; Brickman, Coates and Janoff-Bulman , 1978). The theory was buttressed by finding that most individuals soon revert to baseline after making large financial gains or losses (Easterlin, 1974, 2009; but see Stevenson and Wolfers, 2008). LS was found to be positively related to the personality trait of extroversion and negatively related to neuroticism (Costa and McCrae, 1980). Since these traits are partly genetic and hence fairly stable, they appeared to help account for the presumed stability of LS. Lykken and Tellegen (1996), who coined the term set-point theory, seemed to put the triumphal arch on the theory by comparing the stability of the happiness of monozygotic and dizygotic twins, and inferring that happiness is at least 50 per cent hereditary. Their much quoted conclusion was that, "Trying to be happier may be as futile as trying to be taller".

In the last fifteen years or so, set-point theory has come under increasing challenge. Easterlin (2003), whose work on gains and losses in the financial domain had lent support to the theory, reviewed the effects of changes in non-economic domains, mainly health and family life, and concluded that in these domains adaptation to change is partial, but by no means complete. In particular, evidence about people who developed chronic health problems showed that their LS is permanently lowered (Mehnert, Kraus, Nadler and Boyd, 1990), and research on individuals who got married indicated that some (although by no means all) recorded long term gains in LS (Lucas, Clark, Georgellis and Diener, 2003). It has also been shown that, if individuals experience certain adverse events more than once, particularly unemployment, negative effects on LS are serious and prolonged; full recovery may never occur (Luhmann and Eid, 2009; Luhmann, Hoffman, Eid and Lucas, 2012).

The limitations of set-point theory became clearer as longer and longer term panel data became available. Fujita and Diener (2005), analysing the first seventeen years of SOEP data, found that many respondents had recorded substantial changes in LS. Headey (2008a) and Headey, Muffels and Wagner (2010, 2012), using data from Australian and British household panels as well as SOEP, found that a quarter to a third of survey respondents (but not a majority) recorded substantial gains or losses of LS over periods ranging from a decade to twenty-five years.

Several studies have provided partial explanations of change in LS. Psychologists working in the positive psychology tradition have reported experimental evidence showing the beneficial effects on LS of altruistic actions, religious practice, and activities which play to one's character strengths (Diener and Seligman, 2004; Lyubomirsky, Sheldon and Schkade, 2005; Dunn, Aknin and Norton, 2008; Headey, Schupp, Tucci and Wagner, 2010). On related lines, it has been found that individuals who prioritize altruistic and family goals are on average happier than those who prioritize financial and material goals (Nickerson, Schwarz, Diener and Kahneman, 2003; Headey, 2008b).

In previous papers, replicating results with German, Australian and British panel data, we found that LS is significantly affected by the personality of one's partner, and by behavioural choices relating to social networks and participation in social activities, frequency of physical exercise, and achieving a preferred balance between work and leisure (Headey, Muffels and Wagner, 2010, 2013). The links between these behavioural choices and LS held in within-person (fixed effects) panel regression models, confirming that changes in choices covary over time with changes in LS.

A limitation of most previous attempts to explain changes in LS (including our own) is that, while they may provide explanations for short term change, they do not directly address the issue of how and why LS may remain substantially above (or below) its previous level in the medium or long term, rather than reverting to its previous level, as set-point theory would predict.

A causal model: Granger-causation with positive feedback loops

As already mentioned, we attempt here to account for medium term change with two-way causation models (structural equation models), which include positive feedback loops between LS and variables usually treated as causes of LS. It must be conceded that disentangling two-way causation issues remains problematic, even with long term panel data. Previous researchers have claimed to find two-way causation between LS and domain satisfactions (mainly marriage and job satisfaction), and between LS and quality of social networks, health, housing and volunteering (Headey, Veenhoven and Wearing, 1991; Meier and Stutzer, 2004; Mathison et al, 2007; Nagazato, Schimmack and Oishi, 2011). These studies have been criticised by researchers who claim that structural equation models involving two-way causation can be unstable in the sense that small, reasonably plausible changes in model specification can lead to quite different conclusions about direction of causation (Scherpenzeel and Saris, 1996; see also Wooldridge, 2010).

In this paper we suggest that more stable and better fitting models of LS can be estimated by conceptualising relationships in terms of *Granger-causation* (Granger, 1969; Granger and Newbold, 1974; Pearl, 2009). The intuitions behind Nobel Laureate, Clive Granger's concept of causation, are that (1) causes must precede effects and (2) if earlier values of an *x* variable (a presumed cause) can be shown to be statistically significantly related to later values of a *y* variable (a presumed effect) in equations which include multiple lagged versions of *y*, then it may be inferred that *x* is one cause of *y*. Granger's view was that, by including multiple lags of *y*, the researcher would usually take care of most of the variance due to omitted variables ('unobserved heterogeneity'), and so would avoid biased estimates and statistically significant autocorrelated error terms that indicate longitudinal bias (Granger, 1969; Pearl, 2009; Wooldridge, 2010; Beck and Katz, 2011; Wilkins, 2014).

We suggest extending Granger's approach to two-way causation and panel data, and being willing to infer that *x* and *y* may cause each other in models of the kind depicted in Figure 1. This model shows possible relationships between health (the *x* variable) and LS (the *y* variable). It is based on five waves of panel data, in line with the models to be estimated in this paper.

INSERT FIGURE 1 HERE

The causal paths of main interest in Figure 1 – usually referred to in the causal modelling literature as *cross-lagged paths* - are those with BU? and TD? above them. In referring to bottom-up (BU) and top-down (TD) paths, we are using terminology coined by Diener (Diener, 1984; Diener et al, 1999). He labels models in which LS is viewed as being caused by a weighted combination of *x* variables as BU models, and models in which evaluations of specific aspects of life are viewed as consequences (or spin-offs) of overall LS as TD models.

The curved arrows marked SP? in Figure 1 are a further attempt (in addition to Granger lags) to allow for the possibility that covariance between health and LS could be wholly or partly spurious. These curved arrows link the error terms (residuals) of each pair of *x* and *y* equations, and should be routinely included in structural equation panel models (Kessler and Greenberg, 1981; Finkel, 1996; Wooldridge, 2010).

'Extra' Granger lags are shown by dotted paths running along the top and bottom of Figure 1. As well as helping to avoid biased estimates of the coefficients of main interest, these extra lags can be given a substantive interpretation in LS research. If 'extra' lags of LS are significant in LS equations, it means that individual happiness at time t depends not just on happiness at t-1, but is also affected in a cumulative way by happiness in earlier periods. Evidence about the effects of lagged *x* variables on *x* at time t can be given a similar substantive interpretation.

In summary, four main outcomes are possible when we estimate Granger-style models as shown in Figure 1. One possibility is that the covariance between a particular *x* variable and LS will be found to be wholly spurious. Secondly, it is possible that only BU paths from *x* to LS will prove to be statistically significant. Alternatively, only TD paths from LS to *x* may be significant. If both BU and TD paths prove to be significant, then the inference will be that changes in *x* and changes in LS cause and reinforce each other in a positive feedback loop.

Estimation of two-way causation raises issues relating to model identification and assessment of model fit. These more detailed methodological issues are taken up in the next section.

METHODS

Data: The German Socio-Economic Panel Survey (SOEP)

SOEP began in 1984 in West Germany with a sample of 12541 respondents (Frick, Schupp and Wagner, 2007). Interviews have been conducted annually ever since. Everyone in the household aged 16 and over is interviewed. The cross-sectional representativeness of the panel is maintained by interviewing 'split-offs' and their new families. So when a young person leaves home ('splits off') to marry and set up a new family, the entire new family becomes part of the panel. The sample was extended to East Germany in 1990, shortly after the Berlin Wall came down, and since then has also been boosted by the addition of new immigrant samples, a special sample of the rich, and recruitment of new respondents partly to increase numbers in 'policy groups'. There are now over 60,000 respondents on file, including some grandchildren as well as children of the original respondents. The main topics covered in the annual questionnaire are family, income and labour force dynamics. Questions on LS, domain satisfactions (job satisfaction, satisfaction with leisure etc), health, social participation and exercise have been included every year.

The sub-sample used in this paper comprises respondents who provided life satisfaction ratings every year from 1991-2010 (N=3689). In previous papers, critiquing set-point theory, we have restricted the sample to respondents age 25-64, because the theory is usually said to apply most clearly to prime age adults (Headey, Muffels and Wagner, 2010, 2013). However, results in this and previous papers are almost the same whether the age restriction is in place or not. It seemed more sensible in this paper to obtain a larger sample by including respondents of all ages.[1]

Measures

Life Satisfaction (LS)

LS is measured in SOEP on a 0-10 scale (mean=7.001 s.d.=1.830), the end-points of which are labelled 'totally dissatisfied' and 'totally satisfied'. Single item measures of LS are plainly not as reliable or valid as multi-item measures, but are internationally widely used in household panel surveys and have been reviewed as acceptably reliable and valid (Diener et al, 1999).

Variables potentially implicated in two-way causation with LS

Health

An internationally widely used, self-assessed health measure has been included in SOEP since the early 1990s. Respondents rate their own health on a 1-5 scale ('bad' to 'very good'). Despite its simplicity, this scale has been assessed as reasonably valid in that it correlates satisfactorily with physician ratings (Schwarze, Andersen and Silke, 2000). Its correlation with LS is 0.381.

[1] The sample size would have been almost halved to1873 if the age restriction had not been lifted.

Exercise

SOEP respondents are asked annually about how frequently they engage in active sport or exercise. The response scale runs from 0 ('not at all') to 5 ('every day'). The correlation of this item with LS is 0.149. Frequency of exercise is usually found to be associated with LS, and also with reduction of depression and increased longevity (Gremeaux et al, 2012). Subjective measures, like the one in SOEP, are only moderately reliable; they correlate only moderately with objective measures of the kind provided by an accelerator or pedometer (Mackay, Oliver and Schofield, 2011). Women, particularly, are prone to overstate the amount they exercise.

Frequency of socialising with friends, relatives or neighbours (social participation)

The social participation index used here combines two correlated items about frequency of 'meeting with friends, relatives or neighbours' and 'helping out friends, relatives or neighbours'.[2] The response scale has just three points: 'every week', 'every month' and 'seldom or never'.[3] The correlation of this index with LS is 0.133.

Satisfaction with work and leisure

Job satisfaction and satisfaction with leisure are both measured by single questions asked on the same 0-10 ('totally dissatisfied' to 'totally satisfied') scale as LS. The correlation of LS with job satisfaction is 0.444 and with leisure satisfaction it is 0.348.

Exogenous variables included as 'controls' and also to assist with model identification

Additional exogenous variables are included in the panel models, both as standard 'controls' and also to assist with model identification. Standard 'controls' are: gender (female=1 male=0), age, age-squared, age-cubed, partner status (partnered=1 not partnered=0), years of education, household net income, unemployed (unemployed=1 other=0), disability status (disability=1 other=0).[4] It is also desirable to control for personality traits known to be correlated with LS (Lucas, 2008). The traits measured in SOEP are the so-called Big Five, which many psychologists regard as adequately describing normal or non-psychotic personality: neuroticism, extroversion, openness to experience, agreeableness and

[2] The correlation between these two items varies from year to year, but is usually around 0.3.
[3] 'Seldom' or 'never' have been included as separate categories in more recent waves of SOEP.
[4] Disability status was not included in models in which the self-rated health scale was the *x* variable.

conscientiousness (Costa and McCrae, 1991).[5] Since the traits are partly genetic (Lucas, 2008), it clearly makes sense to treat them as exogenous and causally antecedent to LS and the *x* variables in our models.

In any panel survey, what are called 'panel conditioning effects' are a possible source of bias. That is, panel members might tend to change their answers over time – and answer differently from the way non-panel members would answer - as a consequence just of being panel members. There is evidence in the SOEP data that panel members, in their first few years of responding, tend to report higher LS scores than when they have been in the panel for a good many years (Frijters, Haisken-DeNew and Shields, 2004). This could be due to 'social desirability bias'; a desire to look good and appear to be a happy person, which is stronger in the first few years of responding than in later years. Or it could be due to a 'learning effect'; learning to use the middle points of the 0-10, rather than the extremes and particularly the top end.

To compensate for these possible sources of bias, we include in all equations a variable which measures the number of years in which each panel member has already responded to survey questions.

Estimation of structural equation models with two-way causation

Structural equation modelling, rather than OLS regression analysis, is required whenever the aim is to estimate a set of equations, rather than a single equation, and especially when two-way causal links are involved.[6] The structural equations in this article are estimated using maximum likelihood analysis.[7] Maximum likelihood coefficients and their associated standard errors can be given the same interpretation as regression coefficients. However, assessing the 'goodness of fit' of structural models is more complicated than for regression models. It is necessary to assess the overall fit between estimates for several equations and

[5] The Big Five items have been included in SOEP every 4 years since 2005. In the case of respondents who have answered more than once, we have averaged their results and imputed the average for missing years. So we treat adult personality traits as fixed characteristics, as most psychologists would.

[6] OLS regression analysis is essentially a single equation technique. Regression estimates derived from multi-equation systems are likely to be biased, due to correlations between explanatory variables and error terms in some or all equations. A key assumption of OLS regression is that such correlations are zero.

[7] ML estimates are usually consistent and asymptotically normal under the (not very restrictive) assumption of *conditional normality* (StataCorp, 2013). Only paths or covariances linking conditioning (i.e. control) variables may not be consistent and asymptotically normal (even then, the main problem lies just with estimates of standard errors). These paths are not usually of substantive interest. Substantive interest lies in paths (1) linking exogenous with endogenous variables and (2) between endogenous variables.

the input data for the model (a correlation or variance-covariance matrix).[8] Several measures of fit are conventionally used. The root mean squared error of approximation (RMSEA) and the standardized root mean residual (SRMR) are directly based on comparing differences (residuals) between the actual input matrix and the matrix implied by model estimates. It has become conventional to regard an RMSEA under 0.05 and an SRMR under 0.08 as satisfactory (Bentler, 1990; Browne and Cudeck, 1993).

More complicated assessments are provided by the Comparative Fit Index (CFI) and the Tucker-Lewis index (TLI). The CFI is based on a likelihood ratio (LR) chi-square test and takes account of the contribution of each estimate in the model to overall goodness of fit. The TLI is also derived from an LR chi-square test, and is particularly useful because it rewards parsimony and penalises models including explanatory variables which account for little variance, even if statistically significant. CFI and TLI fits above 0.90 used to be regarded as satisfactory, but some recent reviews recommend 0.95 (Bentler, 1990; Browne and Cudeck, 1993).

In addition to these composite measures of model fit, it is important to check for possible autocorrelation among residuals. As previously explained, autocorrelated errors are usually a symptom of omitted variables bias. We used the new STATA 13 module for structural equation modelling to generate the results reported here (StataCorp, 2013). This package offers a range of estimators, including maximum likelihood, and includes the tests of goodness of fit described above. Statistically significant autocorrelated errors are routinely reported.

Models involving two-way causation can be unstable in the sense that they would never reach equilibrium and would ‘blow up’ if sufficient iterations were run (Finkel, 1995; StataCorp, 2013).[9] In view of this, Bentler and Freeman (1983) developed a test of model stability, which is used here.

Strictly speaking, maximum likelihood estimation of structural equations requires an assumption that endogenous variables are measured on an interval or ratio scale. In fact, all

[8] From a mathematical standpoint, a model can be viewed as a set of constraints - or a set of restricted paths - limiting the possibilities of simply reproducing the input data. Attempts by a researcher to improve his/her model involve modifying these constraints to improve model fit…subject to the theory/hypotheses underlying the model.

[9] Model ‘stability’ is here used as a technical term. Previously we used the term in a different sense to refer to Scherpenzeel and Saris’s (1996) claim that two-way causation models of LS are unstable because apparently small differences in model specification can lead to substantially different estimates.

the endogenous variables in our equations (LS and the *x* variables) are measured on fairly long ordinal scales. It has become routine in research on LS to treat the data as interval-level. Andrews and Withey (1976) were the first researchers to show that, substantively, results using interval-level statistics were much the same as those using ordinal statistics. Most researchers since have followed their lead. An important practical reason for making interval scale assumptions in structural equation modelling is that, although equations can be estimated for models with ordinal or binary endogenous variables, few measures of model fit are available, so it is often unclear whether one model is statistically preferable to another.[10]

Model identification

Issues of model identification always arise in estimating models with possible two-way causal links. The basic question is whether there are there sufficient independent pieces of information (variances and covariances) in the input matrix to enable all free parameters in one's model to be estimated. If there are not enough independent pieces of information, the model is said to 'under-identified' and no mathematically correct solution is possible. In models with only one-way causation, identification is not a problem. But special steps are always needed to achieve identification of two-way causation models. There are three main approaches, all of which are implemented in this paper:-

(1) Exogenous variables may serve as instrumental variables. In our models some of the exogenous variables described earlier – specifically, gender and personality traits – are fixed characteristics (i.e. fixed intra-individually) which we link to wave 1 measures of *x* and LS, but not to later waves. The rationale for omitting links to later waves is that we would not expect fixed characteristics to influence later measures, net of their effects on wave 1 variables (Kessler and Greenberg, 1981). Equivalently, we may say that there is no reason to expect fixed exogenous characteristics to be associated with *changes* over time in *x* and LS.

(2) In previous research lagged versions of *x* and LS were commonly used as instruments to identify equations (Headey, Veenhoven and Wearing, 1991; Scherpenzeel and Saris, 1996; Meier and Stutzer, 2004; Mathison et al, 2007; Nagazato, Schimmack and Oishi, 2011). However, in the models presented in this paper, the inclusion of multiple Granger-style lags increases rather than reduces the number of free parameters that need to be estimated. Our models are nevertheless still identified, due to inclusion of fixed exogenous variables. Further, in later computer runs we modified our models by removing some Granger-style lags – specifically longer term lagged effects of *x* on LS, and LS on *x*. Consequently, in our final models, this second approach to model identification came back into play.

[10] Another limitation is that covariances between the error terms of equations cannot be estimated, so it becomes difficult to assess whether relationships are spurious.

(3) Equality constraints may be imposed. That is, sets of coefficients may be fixed (programmed) to be equal to each other, so reducing the number of free parameters that need to be estimated (Kessler and Greenberg, 1981; Finkel, 1995). No equality constraints were imposed in the initial models estimated for this paper. However, both a priori reasoning and initial computer runs indicated that, empirically, some causal links appeared to be almost exactly the same in consecutive waves of data. So equality constraints were added in later runs, and in some cases improved model fit (see the Results section).

Because we make use of all three approaches to achieving identification in panel data models with two-way causation, our models are 'over-identified'. That is, they actually contain many more bits of information than are required to estimate parameters. This may seem like overkill, but it is ideal for maximum likelihood estimation; the maximum likelihood estimator is designed to find the best solution among the range of solutions available.

A final point relating to identification: the STATA 13 structural equation module includes built-in checks for identification; models that are not identified are either rejected outright, or else the iterative model-fitting procedure fails to converge.[11]

Time lags

Another important issue to consider in causal modelling is the *time lag* between changes in an explanatory variable and changes in the outcome variable. A 'common sense' view, which underlies Granger-causation and most other accounts of causation, is that causes must precede effects. In Figure 1 (above) it is implicitly assumed that changes in health affect changes in LS after some discrete time lag, and that changes in LS, if they affect health at all, also take effect after a time lag. However, in considering our data, it is plausible to believe that the effects of x variables on LS might well be (a) continuous and/or (b) simultaneous or almost simultaneous. For example, consider possible two-way effects between health and LS. It is reasonable to suppose that these effects (if statistically significant) are continuous and perhaps almost simultaneous.

It can be shown mathematically that, if effects are continuous, cross-lagged models of the kind shown in Figure 1 yield unbiased and consistent estimates of the relevant coefficients (Coleman, 1968; Tuma and Hannan, 1984; Finkel, 1995; Beck and Katz, 2011). The point is that, if effects are continuous, the time points at which panel survey measurements are taken

[11] When convergence fails, exactly the same Chi-square LR is produced repeatedly, model iteration after iteration.

are arbitrary, and the task of the researcher is to recover the coefficients of change, using calculus (i.e. the branch of mathematics which deals with continuous change):

1.1 $dy_t /dt= c_o + c_1x_t + c_2y_t$

1.2 $dx_t /d_t = c_3 + c_4y_t + c5x_t$

These two differential equations state that instantaneous rates of change (the *c* coefficients) in *x* and *y* are dependent on each other over time (Coleman, 1968; Tuma and Hannan, 1984; Finkel, 1995). Additional mathematics shows that these equations are equivalent to the standard structural equations (2.1 and 2.2 below) for two-way cross-lagged causation between *x* and *y*, except that in the structural equations random error terms are added to take account of omitted variables.

2.1 $y_t =a_1 +b_1x_{t-1} + b_2y_{t-1} + e_1$

2.2 $x_t=a_2 + b_3y_{t-1} + b_4x_{t-1} + e_2$

These structural equations can be estimated to provide unbiased and consistent estimates of the continuous effect of *x* on *y* and *y* on *x* (Finkel, 1995).

Further modelling issues arise because the changes we are considering may be not just continuous, but also simultaneous. In Granger-causation it is assumed that causes precede effects, so strictly speaking, Granger's approach cannot be extended to simultaneous causation (Pearl, 2009). However, the word 'simultaneous' should not be taken absolutely literally in this context. The causal influences in question could be almost simultaneous, or more practically, could take effect at an interval considerably shorter than the one-year period between panel surveys (Finkel, 1995). It might be thought that the way to deal with this possibility would be just to add simultaneous causal links to the model in Figure 1. Unfortunately, models with both simultaneous and cross-lagged links cannot usually, in practice, be estimated. This is due to problems both of under-identification and multicollinearity (Kessler and Greenberg, 1981; Finkel, 1995).[12] A practical, fall-back procedure that can be followed if, as we found here, a model of this type cannot be estimated

[12] Kessler and Greenberg (1981, chap. 3) describe how a multi-wave panel model of this kind may in principle be identified by using *equality constraints* (i.e. by constraining sets of parameters to be equal to each other). In practice, this approach to identification only succeeds if relationships among variables are quite far from equilibrium (i.e. relationships differ substantially from wave to wave of the panel data). In panel surveys of life satisfaction, it is reasonably clear that relationships between LS and causal variables of interest are much the same from wave to wave. Estimation of models with both simultaneous and cross-lagged links is also likely to run into problems because of multicollinearity; the effects on LS of *x* at time t and *x* at t-1 are likely to be too highly correlated for estimates to be reliable.

is to estimate a model with *only* simultaneous links, and then compare results with those from a model with *only* cross-lagged links, or by extension, Granger-links (Greenberg and Kessler, 1982). If the signs of coefficients of interest in both models are the same, and the coefficients are also statistically significant, then the researcher can feel reassured that results are consistent with each other, although if the 'true' causal (time) lag is misspecified in both

models, coefficients are likely to be under-estimates. This procedure was followed in the present paper and, as reported in detail below, the two-way estimates of main interest were always consistent. More to the point, the Granger-style models were in all cases a much better fit to the data than models with only cross-lagged links or only simultaneous links.

The periods covered by our five-wave panel models are consecutive and overlapping: 1991-95, 1992-1996…up to 2006-10. The reason for using all available five-year periods, instead of just four non-overlapping periods (1991-95, 1996-2000, 2001-05 and 2006-10) is to obtain more reliable results due to larger sample numbers. An assumption which has to be met for this decision to be sensible is that relationships among variables do not change much within the overall time period. Inspection of bivariate correlations within and across waves suggests that this assumption is plausible.

RESULTS

Individual trajectories of LS

It is suggested that a revised theory of LS needs to improve on set-point theory, in part by accounting for typical trajectories of change.

Here – side-by-side – are three pairs of LS trajectories. On the left-hand side (Figures 2.1, 2.3 and 2.5) are *stylised trajectories*, based on set-point theory, for individuals in the top, middle and bottom thirds of an imaginary LS distribution. Next to them (Figures 2.2, 2.4 and 2.6) are the trajectories of three SOEP respondents, selected because their actual trajectories in 1991-2010 are quite typical of individuals in the top, middle and bottom thirds of the LS distribution in SOEP. Of course, no three individuals can perfectly represent a large sample, but these three are typical in having 20-year means and standard deviations of LS which are very close to average for 'their' third of the distribution.[13]

[13] It is accepted, of course, that differing 20-year trajectories can have the same mean and standard deviation. Our inspection of a very large number of trajectories indicated that few panel members recorded steady 20-year declines or steady increases in LS. The majority recorded trajectories with multi-year periods of both relatively

INSERT FIGURES 2.1-2.6 HERE

The stylised cases have obvious set-points and just ‘record’ occasional upward and downward fluctuations in LS, presumably due to favourable and adverse life events. In sharp contrast, the trajectories of the ‘real’ cases are volatile and it is far from clear that they have a set-point. Of course, one could calculate their individual mean (or median) ratings for the 20-year period, but the evidence does not suggest that they have fixed set-points to which they keep reverting.

It is worth looking at the ‘real’ cases in more detail to gain some understanding of why they differ from set-point theory ‘predictions’. Figure 2.2 graphs the trajectory of a married woman age 34-53 in 1991-2010, who recorded a mean LS rating in this period of 8.1 with a standard deviation of 1.1; typical for individuals in the top third of the distribution. She rated above the sample mean on neuroticism and below the mean on extroversion; the opposite of what set-point theory predicts for a happy person. She was unemployed for one year, but her husband was in work and her LS did not decline. Figure 2.4 reports the trajectory of a woman age 49-68 in 1991-2010, who recorded a mean LS rating of 7.0 with a standard deviation of 1.2; typical for individuals in the middle third of the distribution. She rated below the sample mean on both neuroticism and extroversion. In 1991, when first interviewed, she was divorced. She married in 2001, but her LS remained unchanged at that time. In 2008 she was widowed, her LS declined, but then rose again by 2010. Figure 2.6 shows the LS changes of a married man age 40-59. His highly volatile LS trajectory (mean=5.4 s.d.=1.7) is typical of individuals in the bottom third of the distribution. His high rating on neuroticism and low rating on extroversion partly account for his low LS mean. His objective circumstances do not appear to have been particularly unfavourable. He had an above average education and income and was never unemployed, nor divorced.

We graphed the LS trajectories of many individuals in the SOEP panel. The evidence led us to hypothesize that, instead of recording short term fluctuations around a set-point, many people spend several consecutive years above… and then several years below their own long

high and relatively low LS, as shown in Figures 2.2, 2.4 and 2.6. The statistical analysis below (Table 1) confirms this point.

term mean. This hypothesis was tested by a series of within-person (fixed effects) panel regression analyses (see Table 1). In these analyses the independent variable was LS in year t. The dependent variables were successively (i) LS in years t+1 and t-1 (ii) LS in years t+2 and t-2 (iii) LS in years t+3 and t-3 and (iv) LS in years t+4 and t-4. In order to interpret the results in Table 1, it is crucial to appreciate that, in these fixed effects analyses, coefficients are calculated by relating annual deviations from each individual's own (20-year) grand mean on the dependent variable to annual deviations from his/her grand mean(s) on independent variable(s).

INSERT TABLE 1 HERE

The coefficients in Table 1 indicate that if an individual is above his/her own long term mean of LS in a particular year, then h/she is more rather than less likely to be above it in each (but, given the size of the coefficients, not all) of the three years beforehand and the three years afterwards (i.e. six years in total).[14] Similarly, if a person is below his/her grand mean in year t, he/she is likely to be below it for three years before and after.

These apparently innocuous regression results have non-trivial implications for LS theory. They show that LS tends to change in medium term spurts or waves rather than fluctuating short term around a stable set-point.

Structural equation models with 2-way causation: positive feedback loops between LS and x variables

In the rest of the paper we attempt preliminary explanations of why these medium term changes occur. We indicate that there appears to be two-way causation between LS and a range of *x* variables, so that positive feedback loops sometimes tend to perpetuate periods of above average LS, and sometimes perpetuate below average periods. Our initial modelling led us to conjecture that the *x* variables might also tend to change in multi-year periods, rather than short term oscillations. It turned out that they do. Fixed-effects equations like those in Table 1 indicate that health, frequency of exercise and satisfaction with leisure tend to change

[14] Another way of making the same point: among individuals who were above their own 1991-2010 grand mean of LS in any particular year, just over 25% remained above it for all of the next four years. By chance only 6.25% would have done so.

in eight-year periods, social participation in six-year periods, and job satisfaction in four-year periods.

In analysing the five-wave panel data (see Figure 1 above), we first estimated relationships between LS and each particular *x* variable, deploying a Granger-style model with multiple lags. We then modified this model, removing statistically insignificant links and imposing equality constraints where appropriate. We then compared our final Granger-style model - both substantive causal estimates and measures of model fit - with results from one-way causal models, and also from cross-lagged and simultaneous causation models of the kind deployed in previous studies which have investigated two-way causation.

Health and LS

Our first step was to estimate a 'full' Granger-style model; that is, a model in which all lags of both self-assessed health and of LS were included in equations. It was immediately clear that all lags of the health measure were statistically significant in health equations, and that all lags of LS were significant in LS equations. This already suggested that the Granger approach was going to greatly improve model fit, compared with previous approaches. It was also clear, however, that the model could be pruned. While health lagged by one year had a statistically significant effect on LS, and LS lagged by one year had a significant effect on health, little additional variance was accounted for by 'extra' (2^{nd}, 3^{rd}) cross-lags. No extra cross-lags of LS were significant, and although some two-year cross-lags of health were significant at the 0.05 level, their effects were substantively trivial.[15] So we dropped extra cross-lags from our final model.

We also experimented with imposing equality constraints on the cross-lagged BU and TD links of main interest, marked =a and =b in Figure 1.[16] The rationale for imposing these constraints was that inspection of correlation matrices indicated that relationships between health and LS were similar within each wave of data, and also between consecutive waves. So it seemed reasonable to hypothesize that 'true' relationships were, in fact, constant or almost constant over time. In the event, a model with equality constraints was neither clearly

[15] With a large sample size, it is obvious that many statistically significant but substantively trivial effects are likely to be found.

[16] In the final run of this model the equality constraints on the BU and TD estimates for the equations for $Exercise_{t+1}$ and LS_{t+1} were dropped. The reason is that these are not 'Granger' equations in that no 'extra' (2^{nd}, 3^{rd}) lags are available. Consequently, as Granger would predict, the estimates of the BU and TD links from these equations are considerably higher than from the equations with multiple lags, and are probably biased (Granger and Newbold, 1974).

a better, nor clearly a worse fit than a model without these constraints. It depended on which measures of fit one chose to place faith in. The LR Chi-square test and the CFI diagnosed a somewhat worse fit with the equality constraints in place, but the RMSEA and the TLI measures, which reward model parsimony, diagnosed an improved fit. On grounds of parsimony, we elected to treat the Figure 1 model with equality constraints as our preferred model.

A crucial point is that all reasonable variations of a Granger-style model, including the 'full' model with all available lags, indicated two-way causation between health and LS. All had similarly close fits to the data.

Table 2 gives both metric and standardized maximum likelihood estimates for the BU, TD and SP relationships marked in Figure 1. The standardized results are particularly useful because they enable us to compare effect sizes. (Self-assessed health and LS are measured on ordinal scales of different but arbitrary lengths). Table 2 also provides a fairly comprehensive set of measures of model fit.

INSERT TABLE 2 HERE

Positive feedback loops are found between self-assessed health and LS. The size of the standardized BU path from Health to LS (0.079 $p<0.001$) is greater than the TD path from LS to Health (0.057 $p<0.001$). The relative size of the two-way links is approximately the same for men and women, and for older and younger people.[17] The SP link, reflecting the effects of omitted variables, is 0.233 ($p<0.001$); much larger than either the BU or TD links.

This Granger-style health model has a satisfactory fit to the data with a CFI of 0.975 and a TLI of 0.964; both above the conventional 'close fit' level of 0.95. The RMEA is 0.034 (below the standard cut-off of 0.05) and the SRMR is also 0.033 (below the 0.08 cut-off). No statistically significant autocorrelated errors were found. Model stability is satisfactory; all eigenvalues are within the unit circle (Bentler and Freeman, 1983).

It is crucial to compare the fit of the Granger-style model to all reasonable alternatives. All the other models we consider are 'nested' versions of the larger Granger-style model, so

[17] Results for sub-sets of the population are not printed here; available from the authors.

model fit can be directly compared (Bentler, 1990). A model with only one-way causation from health to LS, which includes 'extra' Granger lags and equality constraints, yields exactly the same standardized estimate of the BU link from health to LS as the Granger model, but is a much worse fit to the data: CFI=0.897, TLI=0.862, RMSEA=0.061 and SRMR=0.039. A one-way TD model with causation running only from LS to health also gives the same estimate of substantive interest as the Granger model, but is an even worse fit: CFI=0.892, TLI=0.854, RMSEA=0.068 and SRMR=0.049.

Equally to the point, the Granger model is a closer fit to the data than either of the two-way causation models commonly deployed in previous LS research, namely cross-lagged models without 'extra' Granger lags and simultaneous causation models without extra lags. A cross-lagged model (which includes equality constraints equivalent to those in the Granger model) gives the following fit readings: CFI=0.895, TLI=0.859, RMSEA=0.067 and SRMR=0.060. Readings for a simultaneous causation model are: CFI=0.858, TLI=0.812, RMSEA=0.077 and SRMR=0.071.

Despite fitting the data less well, it is important to record that substantive estimates for BU and TD links in both the cross-lagged and simultaneous causation models are consistent with, although somewhat larger than estimates for our preferred model. The cross-lagged model gives a standardized BU estimate of 0.094 ($p<0.001$) and a TD estimate of 0.085 ($p<0.001$). The simultaneous causation model yields standardized BU and TD estimates of 0.185 ($p<0.001$) and 0.151 ($p<0.001$) respectively. The reason for these somewhat higher estimates is presumably that one consequence of omitting extra 'Granger' lags of outcome variables is to give an upward bias to estimates of the BU and TD links of main interest (Granger and Newbold, 1974).

Additional sensitivity tests were performed to check whether the equality constraints in the model are justified. The results were slightly ambiguous. It transpired that, if all constraints were removed, estimates of the BU and TD links remained within 0.01 (standardized) of the estimates in Table 2. Lagrange multiplier tests indicated that model fit would be just slightly improved (but only at the 0.05 significance level) by removing just one pair of equality constraints. On grounds of theory and parsimony, we elected not to make this change and to retain the model shown in Figure 1.

In summary, finding significant two-way links between health and LS is compatible with the interpretation that feedback loops are operating, so that having better or worse health results in more or less LS, which in turn results in better or worse health, and so on.

Exercise and LS

A full multiple-lag Granger model for links between frequency of exercise and LS indicated that both first and second lags of BU and TD relationships were statistically significant. However, the effect sizes of the second lags were small, suggesting just lingering effects. Again, as with the health model, inspection of the input correlation matrix made it clear that relationships within and between waves of data were quite similar over time. So we again opted for a model in which, as in Figure 1, the cross-lagged causal links of main interest were constrained to be equal. As before, all Granger lags of outcome variables were included.

INSERT TABLE 3 HERE

Two-way causal links are found between exercise and LS. Increased frequency of exercise increases LS. Then, in a positive feedback loop, enhanced LS leads to more frequent exercise. A comparison of the standardized coefficients indicates that the strength of the BU link from exercise to LS (0.034 $p<0.001$) is greater than the TD link from LS to exercise (0.014 $p<0.001$). The two-way links are again approximately the same for men and women, and for older and younger people. SP links, reflecting the effects of omitted variables, are also statistically significant (0.027 $p<0.001$).

The Granger-style model has a satisfactory fit to the data with a CFI of 0.988 and a TLI of 0.982. The RMEA of 0.026 is below the standard cut-off of 0.05, and the SRMR of 0.022 is also satisfactory. No significant autocorrelated errors are reported. Model stability is again achieved; all eigenvalues being within the unit circle (Bentler and Freeman, 1983).

The two-way Granger model is a somewhat closer fit to the input data than a one-way causation model (exercise->LS) and than either a cross-lagged or a simultaneous causation model. A one-way BU model (with 'extra' Granger lags and equality constraints) has these fit readings: CFI=0.979, TLI=0.972, RMSEA=0.025 and SRMR=0.016. A cross-lagged model without 'extra' Granger lags gives the following fit readings: CFI=0.955, TLI=0.940, RMSEA=0.047 and SRMR=0.032. A simultaneous causation model (without 'extra' Granger lags) has these fit readings: CFI=0.921, TLI=0.895, RMSEA=0.062 and SRMR=0.049. A feature of the simultaneous model is that it has negative correlated error terms. These are almost certainly a sign of poor model specification, since it is hard to envisage any omitted variable which could somehow correlate positively with exercise and negatively with LS, or vice-versa (Finkel, 1995).

Although they fit the data less well, estimates for BU and TD links in both the cross-lagged and simultaneous causation models are consistent with, although somewhat larger than estimates for our preferred model. The cross-lagged model gives a standardized BU estimate of 0.038 ($p<0.001$) and a TD estimate of 0.018 ($p<0.001$). The simultaneous causation model yields standardized BU and TD estimates of 0.062 ($p<0.001$) and 0.030 ($p<0.001$) respectively.

Lagrange multiplier tests of the equality constraints in the Granger-style model indicated that all the equality constraints were justified; that is, model fit would not be improved by removing any of them.

Frequency of socialising/social participation and LS

A full Granger model for links between LS and the social participation index indicated that both first and second cross-lags of BU and TD relationships were statistically significant. However, the effect sizes of the second lags were very small, so they were omitted.

INSERT TABLE 4 HERE

A modest degree of two-way causation is found between social participation and LS. The standardized BU link is 0.032 ($p<0.001$), while the TD link is 0.030 ($p<001$).[18] The SP link is also significant at 0.026 ($p<0.001$). The two-way links can again be interpreted as indicating positive feedback loops. Active social participation enhances LS, and enhanced LS reinforces social participation. Results are much the same for men and women, and for older and younger people.

[18] Again, no statistically significant differences were found between men and women, or older and younger people.

The measures of fit again do not give an unambiguous message as between the full Granger-style model and the reduced (Figure 1) model; measures which reward parsimony indicate that the reduced model is marginally preferable, whereas measures which give less weight to parsimony suggest that the full model is a slightly better fit. BU and TD estimates for both models are approximately the same, except that in the full model the BU and TD variance accounted for is partitioned between first and second lags of the explanatory variables, rather than being entirely attributed to first lags.

Alternative models - one-way causation, cross-lagged and simultaneous models – fit the data less well. The cross-lagged model has these fit readings: CFI= 0.931, TLI=0.907, RMSEA=0.051 and SRMR=0.033. Readings for the simultaneous causation model are: CFI=0.885, TLI=0.848, RMSEA=0.065 and SRMR=0.050.

A multivariate two-way causal model: links between exercise, social participation, health and LS

So far we have only assessed two-way causation between LS and *x* variables, taking one *x* at a time. However, it is a plausible hypothesis that health, exercise and social participation exert combined and perhaps more or less simultaneous effects on LS. If so, they should be entered at the same step in a causal model. It seemed possible that, in a multivariate model, one or more of these variables, would be shown not to have significant reciprocal effects with LS. Table 5 gives results for our preferred model with imposed equality constraints.

INSERT TABLE 5 HERE

It appears that health, exercise and active social participation all have statistically significant reciprocal effects with LS. They can combine to change happiness, and changes in happiness then affect subsequent health and subsequent frequency of exercise and social participation.

It is often the case that large structural equation models are a poor fit to the data. It is pleasing that this relatively large model is a satisfactory fit: CFI=0.982, TLI=0.975, RMSEA=0.027 and SRMR=0.033.[19]

[19] As was the case for some of the models with only one *x* variable, just one pair of imposed equality constraints in this multivariate model was diagnosed as not strictly justified; the LR test result would be improved if they

Exercise and health reinforce each other

As well as forming feedback loops with LS, exercise and self-assessed health reinforce each other. In a model with a Figure 1 structure, the standardized effect of exercise on health is 0.044 ($p<0.001$) and the effect of health on exercise is 0.019 ($p<0.001$). These reciprocal effects presumably also contribute to keeping some individuals above their long term average LS for consecutive years, and keeping others below their long term level.

were dropped. Again, however, the measures of fit which reward parsimony – the TLI and RMSEA - provide countervailing evidence in favour of retaining the constraints.

Job satisfaction, satisfaction with leisure and LS

In a final model we assess the effects on LS of job satisfaction and satisfaction with leisure. An assumption here is that job satisfaction and satisfaction with leisure should enter into a model of LS at a later causal step than the three *x* variables in our previous multivariate model. This assumption is in line with most research on LS in which domain satisfactions are treated as consequences of conditions like health, and also of variables measuring behavioural choices like social participation and exercise. Again, the estimates given in Table 6 are from our preferred model with constrained equalities.

INSERT TABLE 6 HERE

Job satisfaction has a significant BU effect on LS; the standardized effect being 0.078 ($p<0.001$). However, the TD and SP estimates are considerably higher at 0.139 ($p<0.001$) and 0.273 ($p<0.001$) respectively. Satisfaction with leisure also makes a significant contribution to LS (BU=0.046 $p<0.001$), and LS has a reciprocal effect on satisfaction with leisure (TD=0.056 $p<0.001$). The results suggest that, although there are feedback loops between these domain satisfactions and LS, the main direction of causation may be top-down rather than bottom-up.

DISCUSSION

Granger-style models fit the data well and two-way causation is pervasive

Modified Granger-style models provide a coherent, consistent account of relationships between LS and *x* variables of interest. It appears that in this area of research *two-way causation is pervasive*. Granger-style models have a satisfactory fit to the data, and indeed are a much closer fit than alternative one-way causation, cross-lagged and simultaneous causation models of the kind deployed in previous research. We have nevertheless found that estimates from cross-lagged and simultaneous causation models are broadly consistent with estimates from Granger-style models, at least in the sense that they also indicate two-way causation.

Two-way causation models may partly explain medium term changes in LS…why many people experience extended periods of satisfaction or dissatisfaction

The main current challenge for LS researchers is to develop a theory of change. Set-point theory is purely a theory of stability and does not account for evidence from panel surveys which shows that the LS trajectories of many respondents are subject to medium or long term change, and certainly do not fluctuate around a stable set-point.

The suggestion in this paper is that *positive feedback loops* between LS and associated variables help to explain why people go through extended 'good times' and 'bad times'. If our models are valid, they provide a preliminary explanation of why an individual might enjoy an extended period of time with a level of LS above his/her own long term average, and then perhaps spend several consecutive years below this average.

The feedback loops in the models should not be over-interpreted. They might appear to imply that the same people keep getting happier and happier, while others get steadily more miserable. However, since they are not fixed effects (within-person) models, they do not strictly imply that (Allison, 2005; Bollen, 2010). Rather they indicate how changes (positive or negative changes) may be maintained for an extended period of time. Psychological adaptation mechanisms doubtless come into play and partly counteract the effects of both positive and negative changes. So, for example, one would expect that a change in exercise levels, or a change in health, would have a greater effect on LS in the short term than in the longer term. People partially adapt to almost all changes, but adaptation is often far from complete (Wilson and Gilbert, 2008).

From this perspective one key issue for future research is to explain why some individuals record more or less permanent changes, and 'lock in' at a level of LS substantially different from what they experienced earlier in life. This paper has focussed on medium term change. So far as we know, there is virtually no empirical research accounting for long term change.

References

Allison, P.D. (2005) *Fixed Effects Regression Methods For Longitudinal Data Using SAS.* Cary, N.C., SAS Institute.

Andrews, F.M. and Withey, S.B. (1976) *Social Indicators of Well-Being.* New York, Plenum.

Beck, N. and Katz, J.N. (2011) Modelling dynamics in time-series cross-section political economy data, *Annual Review of Political Science*, 14, 331-52.

Bentler, P.M. (1990) Comparative fit indices in structural models, *Psychological Bulletin,* 107, 238-46.

Bentler, P.M. and Freeman, E.H. (1983) Tests for stability in linear structural equation systems, *Psychometrika,* 143-45.

Bollen, K.A. (2010) A general panel model with random and fixed effects: A structural equation approach, *Social Forces*, 89, 1-34.

Brickman, P.D. and Campbell, D.T. (1971) ‘Hedonic relativism and planning the good society’ in M.H. Appley ed. *Adaptation Level Theory.* New York, Academic Press, pp. 287-302.

Browne, M.W. and Cudeck, R. (1993) ‘Alternative ways of assessing model fit’ in K.A. Bollen and J.S. Long eds. *Testing Structural Equation Models.* Newbury Park, Ca., Sage, pp. 136-62.

Clark, A.E, Diener, E., Georgellis, Y. and Lucas, R.E.(2008) Leads and lags in life satisfaction: A test of the baseline hypothesis, *Economic Journal,* 118, 222-43.

Coleman, J.S. (1968) ‘The mathematical study of change’ in H. Blalock and A. Blalock eds. *Methodology In Social Research*. New York, McGraw Hill. pp. 428-78.

Costa, P.T. and McCrae, R.R. (1980) Influences of extroversion and neuroticism on subjective well-being, *Journal of Personality and Social Psychology,* 38, 668-78.

Costa, P.T. and McCrae, R.R. (1991) *NEO PI-R.* PAR, Odessa, Fla.

Deeg, D. and van Zonneveld (1989) ‘Does happiness lengthen life?’ In R.Veenhoven ed. How Harmful Is Happiness? Rotterdam, Erasmus University Press, pp. 29-43.

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

Diener, E., Suh, E.M., Lucas, R.E. and Smith, H.L. (1999). Subjective well-being: Three decades of progress, *Psychological Bulletin*, 25, 276-302.

Diener, E. and Seligman, M.E.P. (2004) Beyond money: Toward an economy of well-being, *Psychological Science in the Public Interest*, 5, 1-31.

Easterlin, R.A. (1974) ‘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*. New York, Academic Press, pp. 89-125.

Easterlin, R.A. (2003) Explaining happiness, *Proceedings of the National Academy of Sciences*, 100.19 (May 23) 11176-11183. doi: 10.1073/pnas.1633144100.

Easterlin, R.A. and Angelescu, L. (2009) Happiness and growth the world over: Time series evidence on the happiness-income paradox. Discussion Paper No. 4060 (Institute for the Study of Labor, Bonn). Available at http://ftp.iza.org/dp4060.pdf.

Finkel, S.E. (1995) *Causal Analysis With Panel Data*. Thousand Oaks, Ca., Sage.

Frey, B.S. and Stutzer, A. (2002) What can economists learn from happiness research? *Journal of Economic Literature*, 40, 402-35.

Frick, J.R., Schupp, J. and Wagner, G.G. (2007) Enhancing the power of the German Socio-Economic Panel Study (SOEP) – evolution, scope and enhancements, *Schmollers Jahrbuch*, 127, 139-69.

Fujita, F. and Diener (2005) Life satisfaction set-point: Stability and change.

Journal of Personality and Social Psychology, 88, 158-64.

Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, 37, 424–38.

Granger, C. W. J. and Newbold, P. (1974) Spurious regressions in econometrics, *Journal of Econometrics*, 2.2, 111–120.

Greenberg, D.F. and Kessler, R.C. (1982) Equilibrium and identification in linear panel models, *Sociological Research and Methods*, 10, 435-51.

Gremeaux, V., Gayda, M., Lepers, R, Sosner, P., Juneau, M. and Nigam, A. (2012) Exercise and longevity, *Maturitas*, 73, 312-17.

Headey, B.W. (2008a) The set-point theory of well-being: Negative results and consequent revisions, *Social Indicators Research*, 85, 389-403.

Headey, B.W. (2008b) Life goals matter to happiness: A revision of set-point theory, *Social Indicators Research,* 86, 213-31.

Headey, B.W., Hoehne, G. and Wagner G.G. (2014) Does religion make you healthier and longer lived? Evidence for Germany, *Social Indicators Research*, 119, 1335-61.

Headey, B.W., Muffels, R.J.A. and Wagner, G.G. (2010) Long-running German panel survey shows that personal and economic choices, not just genes, matter for happiness, *Proceedings of the National Academy of Sciences*, 107.42, 17922-17926 (Oct. 19).

Headey, B.W., Muffels, R.J.A. and Wagner, G.G. (2013) Choices which change life satisfaction: Similar results for Australia, Britain and Germany, *Social Indicators Research*, 112, 725-48.

Headey, B.W., Schupp, J., Tucci, I. and Wagner, G.G. (2010) Authentic happiness theory supported by impact of religion on life satisfaction: A longitudinal analysis with data for Germany, *Journal of Positive Psychology*, 5, 73-82.

Headey, B.W., Veenhoven, R. and Wearing, A.J. (1991) Top-down versus bottom-up theories of subjective well-being, *Social Indicators Research*, 24, 81-100.

Kessler, R.C. and Greenberg, D.F. (1981) *Linear Panel Analysis.* New York, Academic Press.

Kline, R.B. (2010) *Principles and Practice of Structural Equation Modelling.*

New York, Guilford Press, 3rd edition.

Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press.

Kuskova,V.A. (2011) A longitudinal analysis of the relationship between life satisfaction and employee volunteerism, *Academy Of Management Proceedings*,1, 1-6.

Lucas, R.E., Clark, A.E., Georgellis, Y. and Diener, E. (2003) Reexamining adaptation and the set point model of happiness: Reactions to change in marital status, *Journal of Personality and Social Psychology*, 84, 527-39.

Lucas, R.E. (2008) Personality and subjective well-being in M. Eid and R.J. Larsen eds. *The Science of Subjective Well-Being.* New York, Guilford Press, pp. 171-94.

Luhmann, M. and Eid, M. (2009) Does it really feel the same? Changes in life satisfaction following repeated life events, *Journal of Personality and Social Psychology*, 97, 363-81.

Luhmann, M., Hoffman, W., Eid, M. and Lucas, R.E. (2102) Subjective well-being and adaptation to life events: A meta-analysis, *Journal of Personality and Social Psychology*, 102, 592-615.

Lykken, D. and Tellegen, A. (1996) Happiness is a stochastic phenomenon, *Psychological Science, 7*, 186-89.

Lyubomirsky, S., Sheldon, K.M. and Schkade, D. (2005) Pursuing happiness: The architecture of sustainable change, *Review of General Psychology*, 9, 111-31.

Mackay, L.M., Oliver, M. and Schofield, G.M. (2011) Demographic variations in discrepancies between objective and subjective measures of physical activity, *Open Journal of Preventive Medicine*, 1, 13-19. Available at http://www.scirp.org/journal/OJPM/

Mathison, L., Andersen, M.H., Veenstra, M, Wahl, A.K., Hanestad, B.R. and Fosse, E. (2007) Quality of life can both influence and be an outcome of general health perceptions after heart surgery, *Health and Quality of Life Outcomes*, 5:27 doi:1186/1477-7525-5-27.

Mehnert, T., Kraus, H.H., Nadler, R. and Boyd, M. (1990) Correlates of life satisfaction in those with a disabling condition, *Rehabilitation Psychology*, 35, 3-17.

Meier, S. and Stutzer, A. (2004) Is volunteering rewarding in itself? IZA Discussion Paper No 1045 (March) IZA, Bonn.

Nagazato, N., Schimmack, U., and Oishi, S. (2011) Effect of changes in living conditions on well-being: A prospective top-down bottom-up model, *Social Indicators Research*, 100, 115-35.

Nickerson, C., Schwarz, N., Diener, E. and Kahneman, D. (2003) Zeroing in on the dark side of the American dream: A closer look at the negative consequences of the goal for financial success, *Psychological Science*, 14, 531-36.

Oswald, A.J. and Wu, S. (2010) Objective confirmation of subjective measures of human well-being: Evidence from the U.S.A., *Science*, 327, 576-79. See also comment on pp. 534-35.

Pearl, J. *Causality: Models, Reasoning and Inference*. Cambridge, Cambridge University Press, 2009.

Scherpenzeel, A. and Saris, W.E. (1996) Causal direction in a model of life satisfaction: The top-down/bottom-up controversy, *Social Indicators Research*, 38, 161-180.

Schwarze, J., Andersen, H. and Silke, A. (2000) Self-assessed health and changes in self-assessed health as predictors of mortality – first evidence from the German panel data, DIW Discussion Paper No. 203, Berlin, DIW.

Sheldon, K.M. and Lucas, R.E. eds. (2014) *The Stability of Happiness*. Amsterdam, Elsevier.

StataCorp (2013) *Structural Equation Modelling Reference Manual, Release 13*. College Station Texas, Stata Press.

Stevenson, B. and Wolfers, J. (2008) Economic growth and subjective well-being: Reassessing the Easterlin Paradox, Brookings Papers on Economic Activity, 39, 1-102.

Thoits, P.A. and Hewitt, L.N. (2001) Volunteer work and well-being, *Journal of Health and Social Behavior*, 42, 115-31.

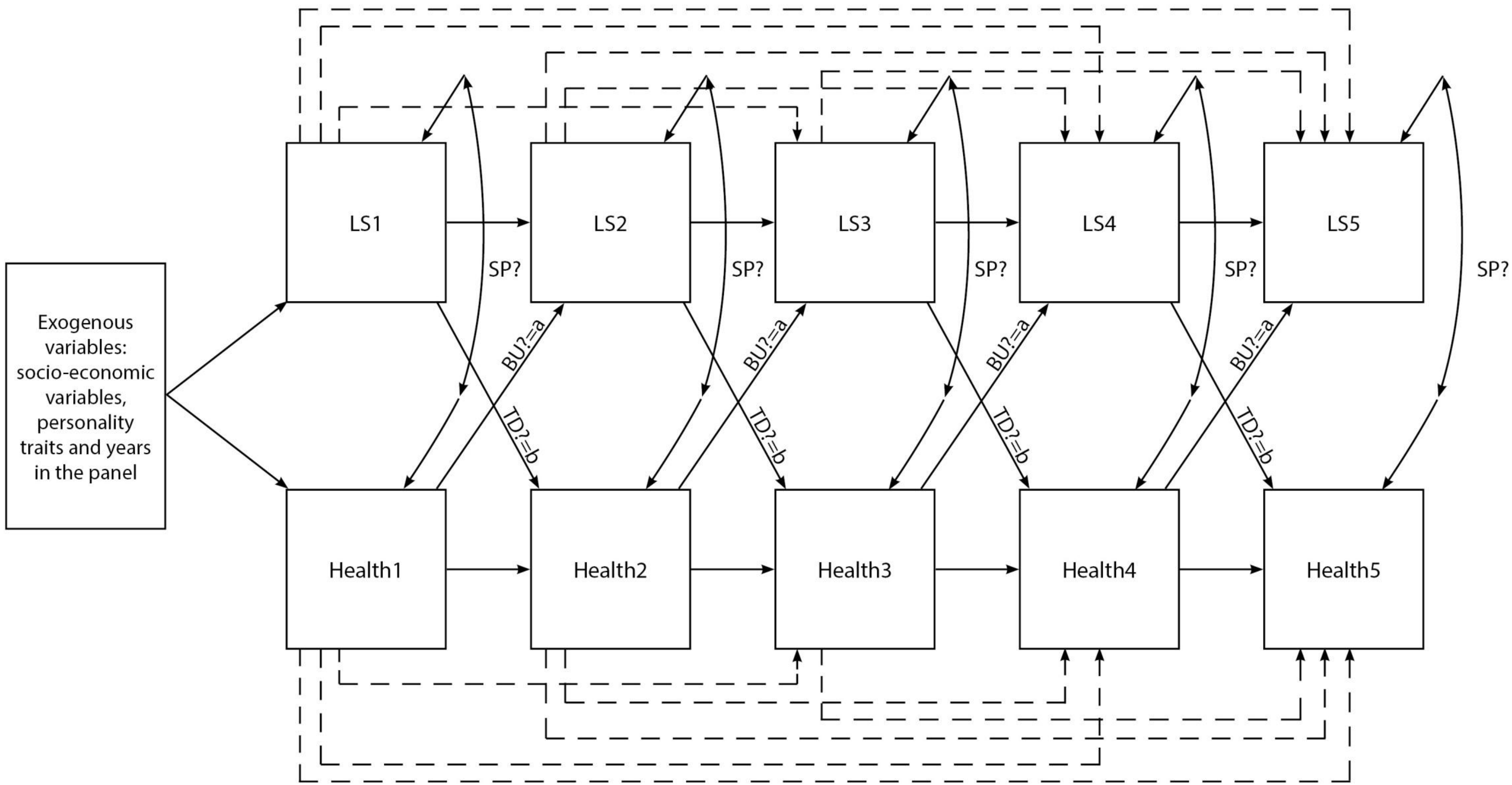
Tuma, N. and Hannan, M. (1984) *Social Dynamics*. New York, Academic Press.

Wilkins, A.S. (2014) To lag or not to lag: Re-evaluating the use of lagged dependent variables in regression analysis. Working Paper, Stanford University Department of Political Science. Downloaded July 4, 2014.

Wilson, T.D. and Gilbert, T.D. (2008) Explaining away: A model of affective adaptation, *Perspectives on Psychological Science*, 3, 370-86.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, Mass., 2nd edition.

Figure 1: Two-way causal links between self-rated health and LS: A 5-wave panel model with Granger-causation* and imposed equality constraints



Key: BU? = bottom-up effects? TD? = top-down effects? SP? = spurious effects?

BU effects between consectutive waves were constrained equal to each other (=a), as were TD effects (=b)

* 'Extra' Granger-lags are shown by dotted arrows.

Individual Trajectories of LS: Bar Charts

Fig 2.1 - Stylised Case: high stable LS due to high E, low N

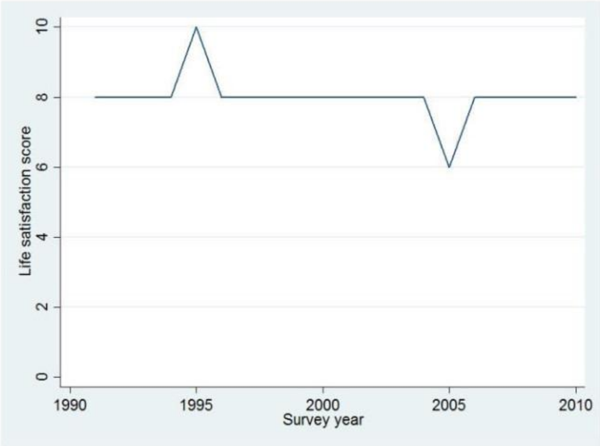
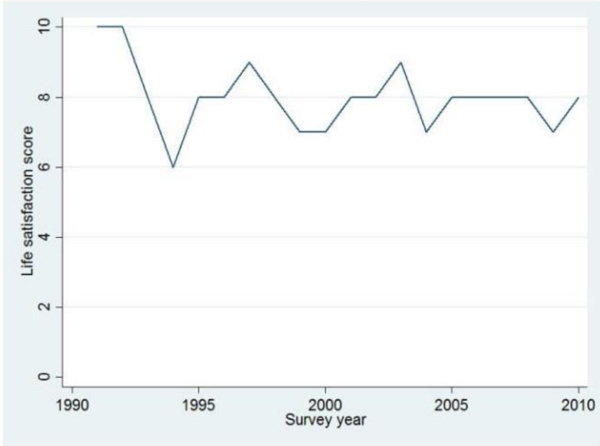


Fig 2.2 - 'Real case: typical mean and s.d. of LS for a person in top third of distribution[a]



a. 20-year mean = 8.1 s.d. = 1.1

Fig 2.3 - Stylised case: medium stable LS due to mean ratings on E and N

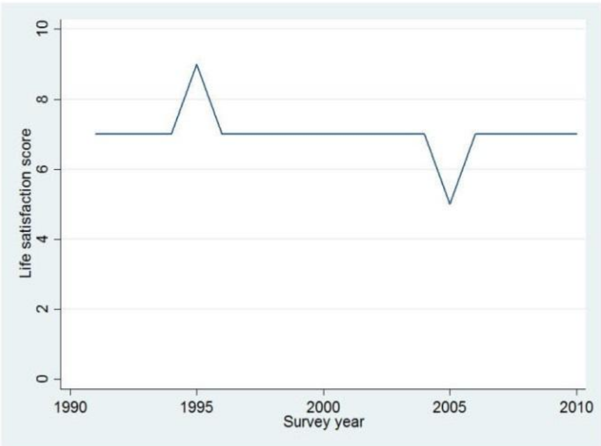
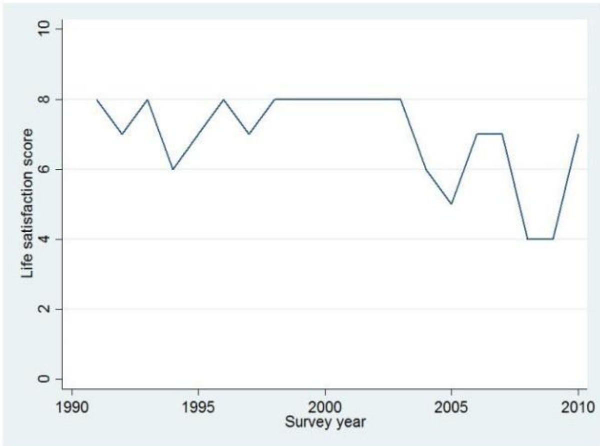


Fig 2.4 - 'Real' case: typical mean and s.d. of LS for a person in middle third of distribution[a]



a. 20-year mean = 7.0 s.d. = 1.2

Fig 2.5 - Stylised case: low stable LS due to low E and high N

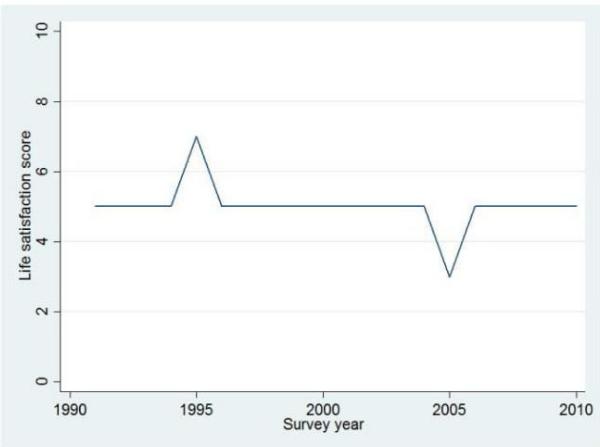
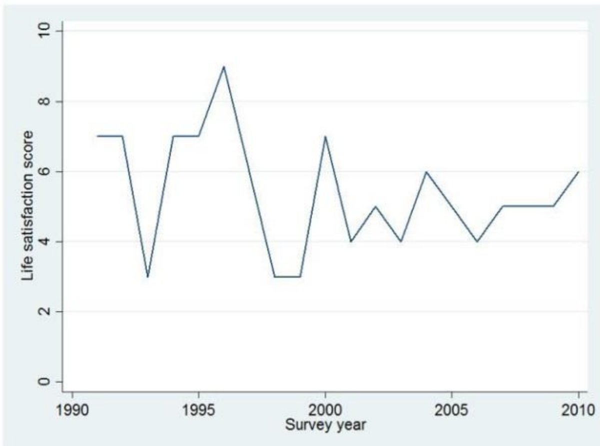


Fig 2.6 - 'Real' case: typical mean and s.d. of LS for a person in bottom third of distribution[a]



a. 20-year mean = 5.4 s.d. = 1.7

Table 1

Within-Person (Fixed Effects) Regressions Of LS_{t+1}, LS_{t-1} etc on LS_t: Metric Coefficients with Standard Errors in Parentheses (N= 59024)[ab]

Dependent variable	Independent variable LS_t
LS_{t+1}	0. 213*** (0.004)
LS_{t-1}	0.216*** (0.004)
LS_{t+2}	0.114 *** (0.004 -)
LS_{t-2}	0.116 (0.004)
LS_{t+3}	0.053 *** (0.004)
LS_{t-3}	0.054 *** (0.004 -)
LS_{t+4}	0.001 [ns] (0.004)
LS_{t-4}	0.001 [ns] (0.004)

a. N = person-years.
b. Controls: age, age-squared.

*** significant at the 0.001 level

ns not significant at the 0.05 level

Table 2

Two-Way Causal Links Between Self-Assessed Health and LS (N= 56729)[a]

	Metric ML estimates	*Standardized ML estimates*	*Measures of model fit*[b]	
$Health_t$ -->LS_{t+1} (BU link)	0.155 *** (0.004)	**0.079 *** (0.002)**	LR Chi-square (df=128)	8406.45***
LS_t ->$Health_{t+1}$ (TD link)	0.029 *** (0.001)	**0.057 *** (0.002)**	RMSEA	0.034
Correlated error (SP link)	0.182 *** (0.003)	**0.233 *** (0.004)**	SRMR	0.033
			CFI	0.975
			TLI	0.964

a. N=person years

b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level

Table 3

Two-Way Causal Links Between Exercise and LS (N=55653)[a]

	Metric ML estimates	*Standardized ML estimates*	*Measures of model fit*[b]	
$Exercise_t$ -->LS_{t+1} (BU link)	0.047 *** (0.003)	**0.034 *** (0.002)**	LR Chi-square (df=136)	5102.34***
LS_t ->$Exercise_{t+1}$ (TD link)	0.010 *** (0.001)	**0.014 *** (0.001)**	RMSEA	0.026
Correlated error (SP link)	0.024 *** (0.004)	**0.027 *** (0.004)**	SRMR	0.022
			CFI	0.988
			TLI	0.982

a. N=person-years.

b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level

Table 4

Two-Way Causal Links Between Social Participation and LS (N=56593)[a]

	Metric ML estimates	*Standardized ML estimates*	*Measures of model fit*[b]	
Social $participation_t$ -->LS_{t+1} (BU link)	0.126 *** (0.007)	**0.032 *** (0.002)**	LR Chi-square (df=136)	5455.33***
LS_t ->Social $participation_{t+1}$ (TD link)	0.008 *** (0.000)	**0.030 *** (0.002)**	RMSEA	0.026
Correlated error (SP link)	0.010 *** (0.002)	**0.026 *** (0.004)**	SRMR	0.021
			CFI	0.983
			TLI	0.975

a. N=person years
b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).
*** significant at the 0.001 level

Table 5:

Two-Way Causal Links: LS Related to Health, Exercise and Social Participation (N=55512)[a]

	Metric ML estimates	*Standardized ML estimates*	*Measures of model fit*[b]	
Health			LR Chi-square (df=332)	13732.53***
$Health_t$ ->LS_{t+1}	0.148 *** (0.004)	**0.078 *** (0.002)**	RMSEA	0.027
LS_t ->$Health_{t+1}$	0.029 *** (0.001)	**0.057 *** (0.002)**	SRMR	0.033
Correlated error (SP link)	0.182 *** (0.003)	**0.234 *** (0.004)**	CFI	0.982
$Exercise_t$			TLI	0.975
$Exercise_t$ --> LS_{t+1}	0.016 *** (0.003)	**0.011 *** (0.002)**		
LS_t ->$Exercise_{t+1}$	0.010 *** (0.001)	**0.014 *** (0.001)**		
Correlated error (SP link)	0.022 *** (0.004)	**0.025 *** (0.004)**		
Social participation				
Social $participation_t$ -->LS_{t+1}	0.097*** (0.010)	**0.022 *** (0.002)**		
LS_t->Social $participation_{t+1}$	0.005*** (0.000)	**0.014 *** (0.001)**		
Correlated error	0.007*** (0.002)	**0.024 *** (0.004)**		

a. N=person years. b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983). *** significant at the 0.001 level

Table 6

Two-Way Causal Links: LS related to Job Satisfaction and Satisfaction with Leisure (N=27291)[a]

	Metric ML estimates[a]	*Standardized ML estimates*[a]	*Measures of model fit*[b]	
Job satisfaction				
Job satis.$_t$ ->LS$_{t+1}$ (BU link)	0.065 *** (0.002)	**0.078 *** (0.003)**	LR Chi-square (df=232)	3051.49***
LS $_t$ -> Job satis.$_{t+1}$ (TD link)	0.172 *** (0.004)	**0.139 *** (0.003)**	RMSEA	0.021
Correlated error (SP link)	0.503 *** (0.012)	**0.273 *** (0.006)**	SRMR	0.023
Satisfaction with leisure			CFI	0.986
Leisure satis. $_t$. ->LS $_{t+1}$ (BU link)	0.034 *** (0.002)	**0.046 *** (0.003)**	TLI	0.981
LS->Leisure satis. $_{t+1}$ (TD link)	0.075 *** (0.004)	**0.056 *** (0.003)**		
Correlated error (SP link)	0.307 *** (0.011)	**0.172 *** (0.006)**		

a. N = person-years   b. Model stability=0.000; all eigenvalues inside unit circle (Bentler and Freeman, 1983).

*** significant at the 0.001 level