Melbourne Institute Working Paper Series

Working Paper No. 4/14

Multidimensionality of Longitudinal Data: Unlocking the Age-Happiness Puzzle

Ning Li

Multidimensionality of Longitudinal Data:
Unlocking the Age-Happiness Puzzle*

Ning Li
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne

Abstract

In social and economic analysis of longitudinal data, the socio-economic variables that are statistically significant in pooled data regressions sometimes become insignificant after individual fixed effects are controlled for. This phenomenon has been observed in the analysis of the relationship between age and happiness. The discrepancy in results between regressions with and without controlling for individual fixed effects is sometimes known as a mystery in the research of age and happiness. This paper points out that cross-sectional information and longitudinal information reflect distinct aspects of the phenomenon under study. In age-happiness studies, cross-sectional information describes whether, in a particular year, people of a certain age are happier than people of other ages. Longitudinal information describes whether people become happier or less happy over the life cycle. The former compares happiness between different people, and the later compares happiness within the same person. Average happiness is U-shaped in age among different cohorts, and simultaneously decreases with age in the life cycle within individuals. Using data on happiness from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, this paper explains what "individual fixed effects are controlled for" means in the context of FE regression, gives insight into the age-happiness puzzle and raises awareness of the multidimensionality of longitudinal data.

Introduction

Happiness is one of the social and economic goals in human lives (Alkire 2010). Age is frequently used in social and economic analysis, either as a variable of primary interest or as a control variable. The relationship between age and happiness has been the subject of considerable research in psychology, economics, sociology, and medicine. But there seems to be little agreement on the nature of the relationship. The majority of studies, such as Clark and Oswald (2006), report that the relationship between age and happiness is U-shaped, with a minimum happiness at the ages between mid-30s and mid-40s. Other studies have found that happiness: increases with age (Mroczek and Kolarz 1998; Diener et al 1999; Carstensen et al. 2000; Charles et al. 2001); is stable over time (Costa et al 1987; Kunzmann et al 2000; Myers 2000, Easterlin 2003;); has an inverse U-shaped relationship with age (Chen 2001; Lang and Heckhausen 2001; Mroczek and Spiro 2005); decreases with age (Goldbeck et al. 2007, Berg et al 2009); or has a wave-shaped relationship with age (Fischer 2009; Frijters and Beatton 2012). These studies used ordinary least squares (OLS) regression or ordered logit or probit regression, mostly on cross-sectional data or pooled longitudinal data, often with the standard socio-economic controls.

With the increasingly available resources in longitudinal data and statistical packages, another group of researchers tried to control for individual fixed effects, including individual fixed personality traits, besides the standard controls. The U-shaped relationship is found less supported (Van Landeghem 2008) or a negative relationship is found (Ree and Alessie 2011; Frijters and Beatton 2012; Kassenboehmera and Haisken-DeNew 2012) when including individual fixed-effects. There are also reports from this group that the U-shaped relationship between age and happiness persistently exhibits in regressions both with and without controlling for individual fixed effects (Clark 2007; Kristoffersen 2013; Cheng et al 2014).

The results from the previous studies appear vary widely. But it is worth noting that segments of a wave shape can look like a U-shape, an increase, or an inverted U-shape within different age ranges. When the shapes and their corresponding ranges are considered together (Gwozdz and Sousa-Poza 2010), the results are not as diverse as they appear to be. For example, the inverse U shape with a peak at age 65 reported by Chen (2001, p.63) is based on people of age 60 or over. This result is largely consistent with the later wave shape reported by other studies. As such, this leaves us with two broad types of results: A) the U shape from

pooled OLS, which may include the increase, the inverse U, and the wave shapes in sub-regions; and B) the declining pattern from FE regressions.

The U shape is intriguing. It says that happiness improves with age after middle life. This seems to contradict with general expectations because health is usually failing with age. It is hence not surprising that, a paper explicitly stating that subjective well-being is U-shaped over the life cycle (Blanchflower and Oswald 2008) received immediate criticism that the U shape is an artifact due to reverse causality (Frijters and Beatton 2012) from the use of inappropriate control variables (Glenn 2009). In the discussion, Glenn (2009) supplied a detailed example to show that when a control variable is constructed in a certain way and is included in the regression of happiness on age, the estimated happiness appears to decrease with age despite of the truth that happiness is constant over time. Glenn's example sets a serious challenge to the reported U shape, but the original authors responded back, in the same volume as Glenn (2009), with a U shape obtained from a regression without any control variables (Blanchflower and Oswald 2009). Upon this point, the debate has become extremely interesting because both sides hold solid evidence that leads to different conclusions. The contradiction between Glenn's example and Blanchflower & Oswald's empirical evidence became a mystery in the research of age-happiness relationship.

Further investigations found the U shape in pooled OLS disappearing and a declining pattern emerging when controlling for individual fixed effects (Van Landeghem 2008; Ree and Alessie 2011, Frijters and Beatton 2012; Kassenboehmer and Haisken-DeNew 2012). Faced with such findings, it appears that on one hand researchers would like to think that the new result obtained with the individual fixed effects control should be closer to the truth. But on the other hand, the large discrepancy between the new and the old results made them hesitant to accept the new result.

An explanation is needed for the discrepancy. Some authors believe that the negative relationship is a misconception (Lacey et al. 2012 p.647), some believe neither the U shape nor the declining pattern can be found without making arbitrary assumptions about the data (Ree and Alessie 2011, p.182), while others suspect that different parameterizations caused the inconsistency between the age-happiness relationships (Wunder et al 2013, p.156-157) . The current understanding is that the difference between the U shape and the declining pattern is a result from unobserved heterogeneity, reverse causality, sample selectivity

(Frijters & Beatton 2012, p.534-536), participants' time-in-panel experience, interviewers' characteristics (Kassenboehmer and Haisken-DeNew 2012, p.235), or limited parameterization (Wunder et al 2013, p.156-157).   This paper points out that a more fundamental reason for the difference between the U shape and the declining pattern is the multidimensionality of longitudinal data.

The presence of a sizeable difference in estimates between regressions with and without controlling for individual fixed effects is not unique to the study of happiness. It is often the case that the socio-economic variables that were statistically significant in pooled data regressions become insignificant or much less significant when individual fixed effects are controlled for.

This paper attempts to provide a general explanation for this phenomenon. Longitudinal data contain rich information that can be extracted in many ways. Cross-sectional and longitudinal information reflect distinct aspects, answer different questions. Execution of OLS in cross-sectional data extracts information about different people, while execution of FE regressions extracts information about the same person. Different ages in cross-sectional OLS come from different people, while different ages in FE regressions are time of years experienced by the same person. A lack of clarification between these meanings of "age" causes confusion. Cross-sectional and longitudinal estimates are numerically un-related and conceptually different. OLS regression of happiness on age in cross-sectional data produces average happiness among different cohorts of people. In contrast, the FE regression of happiness on age shows yearly change in happiness within the same person.

Section 2 justifies the interpretation of OLS in cross-sectional data as cohort effects. Section 3 identifies change in happiness over time, describes a simple way to achieve this, and discusses the exact meaning of the FE regression. Section 4 explains the conflict between Glenn's (2009) example and Blanchflower and Oswald's (2009) empirical evidence. Section 5 concludes the paper. Simple summaries, regression results, and graphs are used for illustration.

Age Difference in Cross-Sectional Data is Cohort Difference

In order to convey the idea, a longitudinal data set is needed. Without loss of generality, the HILDA data set will be used.

The HILDA Survey is a household panel survey with a focus on income, employment and family, but also extends to cover many other topics including health and subjective well-being. Each year a question in the Person Questionnaire asks ''All things considered, how satisfied are you with your life?" Respondents are instructed to select a rating between 0 (totally dissatisfied) and 10 (totally satisfied) to indicate the degree of satisfaction. Detailed information about HILDA data can be found in Summerfield et al (2013). The data used in this paper are drawn from the balanced panel of 7229 persons followed up from 2001 to 2011 in the first eleven waves of the survey. Each person's age and the overall life satisfaction (called happiness in this paper) will be analyzed. To avoid distraction, no control variables will be used. The dependent variable in all analyses is happiness, rescaled to 0–100. The sole independent variable is age, in various forms.

Average Happiness Observed in a Cohort

At a fixed point in time, different ages must come from people born in different years. Therefore, age differences in cross-sectional data are cohort differences.

Use a person's age in the first wave to indicate the cohort that the person belongs to. Cohort 30, for example, consists of all people aged 30 in Wave 1. For each cohort, the average of happiness can be calculated from all members. Comparison of the average happiness between cohorts yields cohort effect on happiness. Strictly speaking, any difference in happiness between two cohorts is a combined effect from cohort and period because the older cohort experienced a longer time period than did the younger cohort. When period effect is negligible, the difference in average happiness from cross-sectional data is attributed to cohort difference. The dots in Figure 1 indicate average happiness in Wave 1 for each cohort versus the cohort's age in Wave 1. A U shape over the age range 15 to 89 is evident.

Figure 1: Average Happiness versus Age in Cross-Sectional Data (Wave 1)
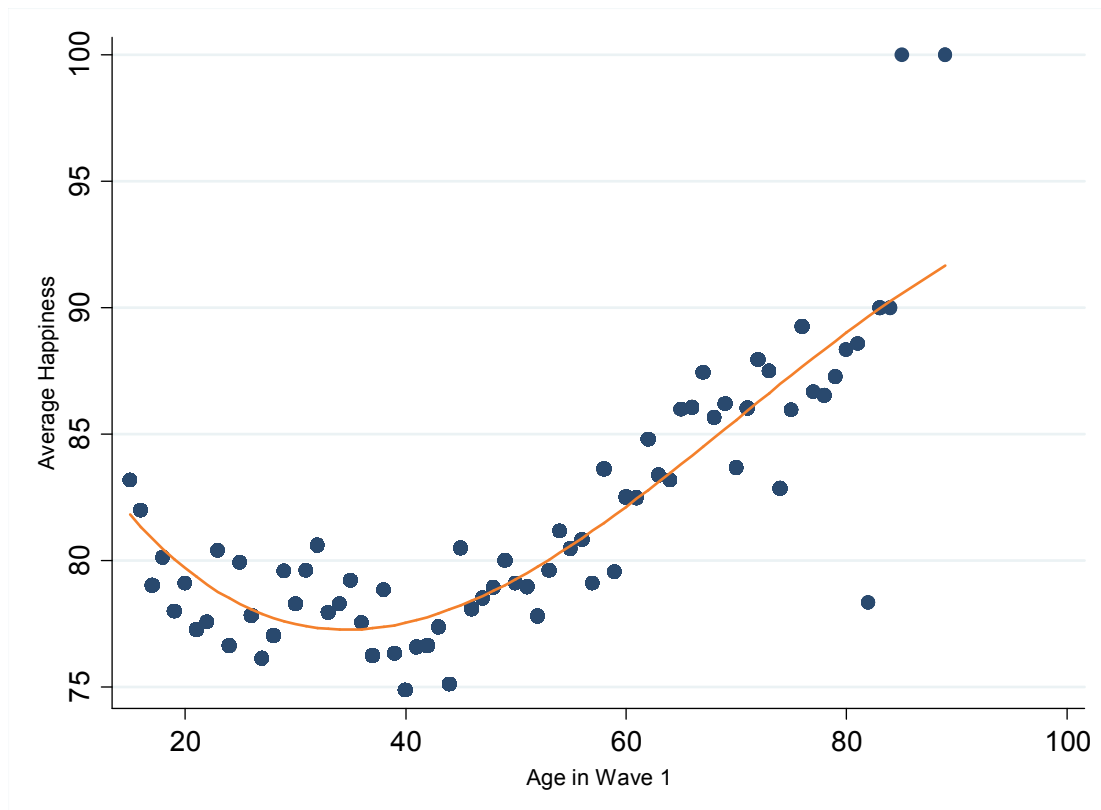
Notes: Results presented in the paper can be replicated using the programs provided in the Supplementary Material.

Age Dummies and Fitted Values

OLS estimates can be linked to average happiness in the observed data through dummy variables. Happiness is regressed on age dummies in Wave 1 data. Each dummy corresponds to a specific age value. The total number of dummy variables is equal to the total number of distinct ages, so is the total number of regression coefficients (including the constant). Given age a, the associated dummy variable takes the value 1 if age is a and takes zero otherwise. After the regression, average happiness can be estimated by the fitted regression equation. These fitted values can be plotted versus age to visualize the relationship between the two variables. The OLS results can be produced using the program given in the supplementary.

A direct comparison of the average happiness in each cohort between the observed and the estimated confirms that in a given cohort (and hence given the cohort's age in Wave 1), the fitted values from regression of happiness on age dummies are identical to the average

7

happiness in the observed data. That is, regression using age dummies in cross-sectional data shows the effect of different birth years on happiness.

OLS Regression Using Age Polynomials

While regression using age dummies fits the data exactly, a polynomial form is more commonly used for its simplicity and a higher power in prediction. When polynomials are used, the number of regression coefficients is often less than the number of distinct age values, and the coefficients no longer directly correspond to the observed averages. But the fitted values from polynomial regressions do have a similar interpretation as that from the regression using age dummies.

The smooth line in Figure 1 shows the average happiness in each cohort from the regression fitting of happiness on age, age squared and age cubed in Wave 1 data. Given a cohort, the value in the fitted curve approximates the observed average happiness (shown as a dot) in that cohort. The precision of the approximation depends on the polynomial terms. In general, the higher the order of the polynomials used, the better the approximation. That is, the fitted values from OLS regression of happiness on age based on single wave data or cross-sectional data either exactly equal (for age dummies) or approximately equal (for age polynomials) the average happiness in each of the cohorts. When period effect can be ignored, the fitted curve from cross-sectional regression displays cohort effect on happiness at a fixed point in time.

Regressions from Pooled Data and from Single Wave Data

When multiple waves of data are available, it has almost become a standard procedure in happiness research to run OLS regression on pooled data. A single age value in the pooled data corresponds to multiple birth cohorts. A pooled data regression alone does not take into account of the age distinctions between multiple cohorts.

Moreover, results from regressions by pooling longitudinal data depend on the variance structure in the data. In a longitudinal data set, the average value of repeated measures from a person is called the within-person mean. This value represents the typical score of the person during the survey period. The variability in the set of typical scores from all people indicates

the extent to which one person differs from the others. The standard deviation of this variability is termed as the between-person standard deviation (Between SD).

A person's score on any specific occasion can be compared to his or her within-person mean. This deviation reflects the degree to which a person's behaviour in a particular time differs from his or her long-term performance. The set of deviations from all people in the data can be formed and the standard deviation of the set can be calculated. This value is called the within-person standard deviation (Within SD). A small Within SD indicates that every people's fluctuation around his or her typical value are similar, either all small or all large. A large Within SD indicates that some persons' variability around his or her typical score differ a lot from other people's. Because no part of the Within SD is captured by the Between SD and vice versa, the two represent independent sources of variation in longitudinal data.

Figure 2: Average Happiness and Associated 95% Prediction Intervals
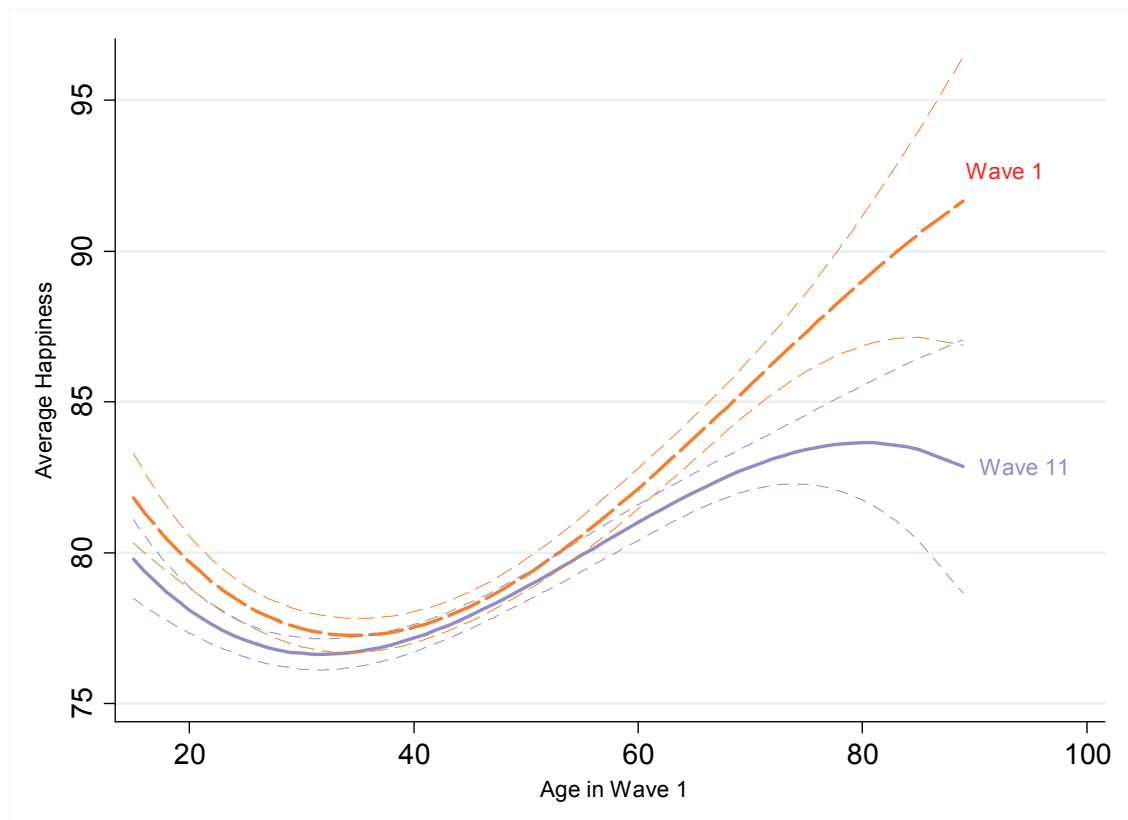
Notes: The fitted values were obtained by regressing happiness on age, age squared and age cubed separately using Wave 1 and Wave 11 data. The thin dashed lines are point-wise 95% predictive intervals of the fitted values. The thick dashed and solid lines are the fitted curves from Wave 1 and Wave 11 data, respectively.

In cross-sectional data, only the Between SD is in action. Regressions from cross-sectional data are purely determined by the between-person variability. In pooled multi-wave longitudinal data, regression results are similar to that from cross-sectional data when the Between SD dominates. If the Within SD dominates, the pooled regression may not preserve the features that the single wave data generate. For happiness, the Within SD decreases with the number of waves in the survey (Cummins et al. 2013). The Between SD and Within SD are 10.93 and 9.78 respectively in HILDA balanced panel. That is, if two persons were selected at random from the data, the difference in happiness between them is expected to be 11.7% larger than the difference in happiness between two randomly selected years for a single person. The regression curve from pooled data therefore preserves the U shape, before age 80. After age 80 the shape shows a declining tail that is not evident in Wave 1 data. This feature is inherited mainly from Wave 11 data.

Age Difference in FE Regressions is Time of Years Experienced by Fixed Cohorts

When the goal is to identify change in happiness over time within the same person or within fixed cohorts, multiple waves of data must be used.

Longitudinal Comparison at Wave Aggregate Level

Change in happiness between the beginning and end of a survey may be identified by running separate regressions to the first and last waves of data and then comparing the results. In Figure 2 the fitted curves of a third-order polynomial regression of happiness on age, separately conducted in Wave 1 and Wave 11 data, are aligned by persons' age in Wave 1. Each value in the horizontal axis corresponds to a specific cohort. The points on the two curves represent estimated average happiness in Wave 1 and 11 respectively, for that cohort. The distance between the two points is therefore the difference in average happiness that has occurred between waves 1 and 11. Or, the ageing effect on happiness when the period effect can be ignored. The following observations can be made from Figure 2:

- The size of change in average happiness depends on cohort
- The minimum and maximum changes occurred in Cohort 44 and Cohort 89, respectively
- The change is statistically insignificant in cohorts 20 to 62, and is statistically significant otherwise.

The above procedure compares the average happiness in each cohort between waves 1 and 11. For a given cohort, only the wave-level average for the cohort is used. The method hence also works for quasi-panel data where a series of cross-sectional data are collected over time. Gerlach and Stephan (1996) and Easterline's (2006) treatments in obtaining ageing effect fall into this category. This approach is intuitively appealing and has the advantage of showing interactions between the change and cohort.

Longitudinal Comparison at the Person Level

The above approach may be altered by swapping the order of the two operations. That is, differentiate happiness scores of each person first and then analyse the differences. Taking the difference between repeated measures from the same person eliminates noise in the data caused by time-invariant individual heterogeneity. For example, scores reported by a pessimistic person may be consistently lower than those reported by others, but the difference between such two low scores is identical to the difference as if the reporting behaviour of this person were the same as others. It is under this consideration that the FE model incorporates the idea of a matched data design into the general regression framework to make a powerful comparison tool. An essential ingredient in FE regressions is the deviations from the time means, for all variables involved. The FE regression of y on x can be thought of as a short name for the OLS regression of the deviations in y on the deviations in x.

Single Cohort Case—

To understand how a FE regression works numerically, it is helpful to consider data from a single age cohort. Without loss of generality, Cohort 65 is used for illustration. The FE regression of happiness on age dummies is run to the data from the cohort. The coefficients, the average happiness, and the yearly change in average happiness estimated from the FE regression are given in Table 1.

The last two columns in the table are summaries from the observed data, without using any regressions. The cohort consists of 87 persons aged 65 in Wave 1. The average happiness of the cohort in Wave 1, 85.98, is the first entry in Column 5. In Wave 2 the cohort is 66 years old, and the average happiness of the cohort in this age is 86.44, shown as the second entry in

11

Column 5. When the cohort ages from 65 to 66, its average happiness increased by 0.46 point. This yearly change is the first entry in the last column of the table. Other entries in the last column are calculated similarly.

Table 1: Relationship between the Fixed-Effects Regression of Happiness on Age Dummies and Observed Data (Cohort 65)

Age	Coefficient	Estimated Average Happiness	Yearly Change	Mean	Yearly Change
		Fixed-Effects Estimates		Observed Data	
65				85.98	
66	0.46	86.44		86.44	0.46
67	1.72	87.70	1.26	87.70	1.26
68	-0.58	85.40	-2.30	85.40	-2.30
69	1.03	87.01	1.61	87.01	1.61
70	-1.15	84.83	-2.18	84.83	-2.18
71	-2.41	83.56	-1.26	83.56	-1.26
72	-3.33	82.64	-0.92	82.64	-0.92
73	-1.95	84.02	1.38	84.02	1.38
74	-1.38	84.60	0.58	84.60	0.58
75	-3.79	82.18	-2.41	82.18	-2.41
Constant	85.98				

 A direct comparison between Column 4 and the last column of Table 1 shows that the yearly change in average happiness estimated by the FE regression are identical to the yearly change in the observed data. In other words, the yearly change in average happiness in the observed data can be equally found from the FE regression of happiness on age dummies.

Because a single cohort is involved here, the age difference in the FE regression is the number of years experienced by the cohort, i.e., ageing. The estimated average happiness from the FE regression, when scattered over age, displays the ageing process from one year to the next.

Figure 3: Estimated Average Happiness from the FE Regression of Age Polynomials
Approximate Average Happiness in the Observed Data
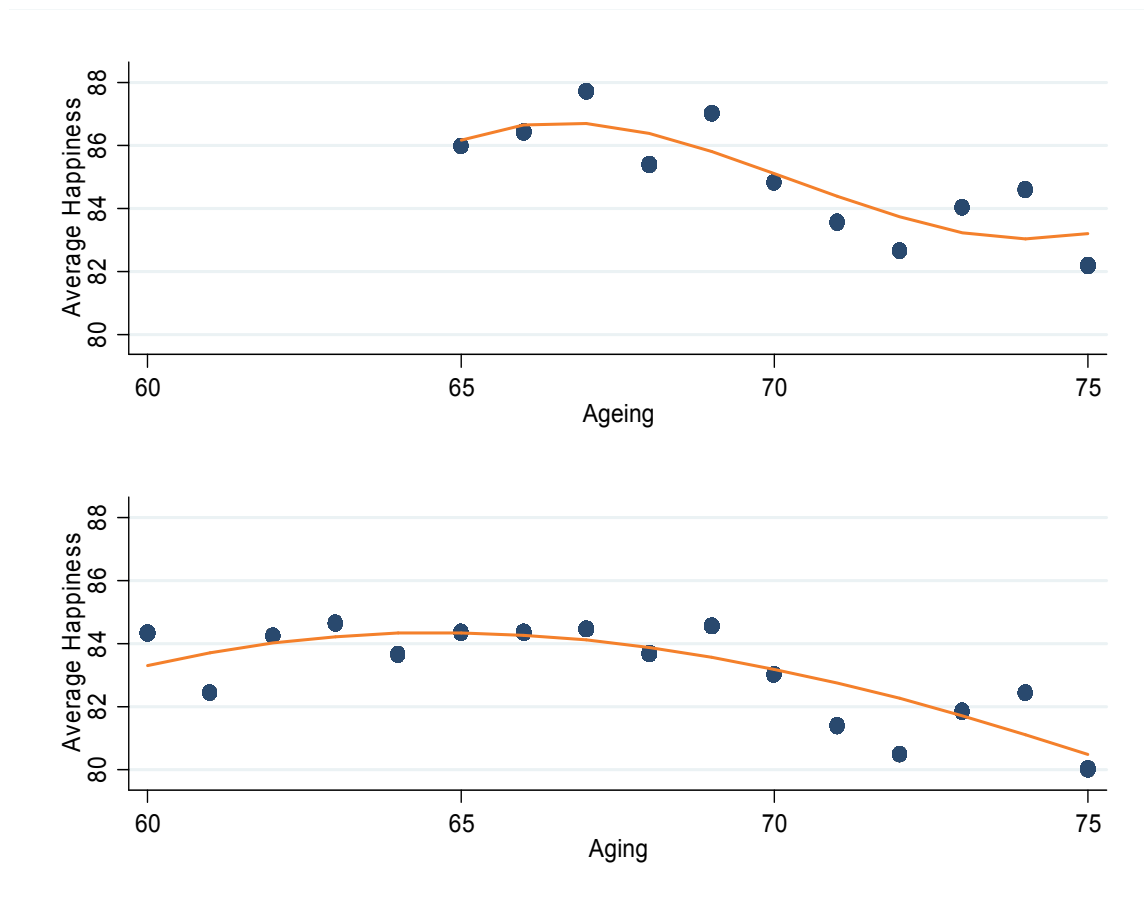
Notes: The upper graph is from a single cohort data, Cohort 65here. The dots and line in the graph are respectively the observed and estimated average happiness. The estimation is from the FE regression of happiness on age, age squared and age cubed. The lower graph is from Cohort 60 and Cohort 65. In the graph the dots are average happiness at each age from both cohorts. The line represents average happiness estimated by the FE regression of happiness on age, age squared and age cubed using data from both cohorts.

When age polynomials replace age dummies, the interpretation of FE regressions is similar. The yearly change in average happiness estimated from the regression now approximately equal the yearly change in the observed data, and the fitted curve forms a smooth line over age. The dots in Figure 3 represent the observed average happiness. The lines represent the average happiness estimated from the FE regression on age polynomials. It is evident that the lines approximate the observed average and capture trends in the data.

Multiple Cohort Case—

When two or more cohorts present in data, which is usually the case, and when age dummies are used, the yearly change in average happiness estimated from the FE regression is a weighted average of the observed yearly changes in each of the cohorts. The weights are proportional to the relative cohort size.

Without loss of generality, two cohorts are used to illustrate the multi-cohort situation. Suppose the data come from Cohort 60 and Cohort 65, of size 122 and 87 respectively. Persons aged between 60 and 64 only come from Cohort 60, because people in Cohort 65 were already 65 years old at the beginning of the survey. Similarly, persons aged 71 to 75 come from Cohort 65 only. For ages between 65 and 70, data come from both cohorts.

Table 2 displays three panels: Fixed-Effects Estimates, Observed Average Happiness, and Observed Change. Entries in the Observed Average Happiness column are mean scores from the data without using any regressions.   Successive values under this column are subtracted to obtain the Observed Changes. For each cohort, the observed yearly difference shows how happiness has changed from one year to the next within the cohort. When both cohorts present at a given age, the observed yearly changes in both cohorts are weighted according to their relative cohort size to produce an overall yearly change. This quantity is entered in the last column of the table. For example, the overall change in average happiness under age 67 is calculated as follows. The observed yearly change −0.74 in Cohort 60 and the observed yearly change 1.26 in Cohort 65 are multiplied by corresponding cohort proportions (122/229) and (87/229) to form the overall difference 0.10.   When a single cohort presents at a given age, the overall yearly change is simply the yearly change in that cohort.

The panel Fixed-Effects Estimates in Table 2 refers to the FE regression of happiness on age dummies, using data from both cohorts. The column Estimated Average Happiness are obtained as the sum of the regression constant and each coefficient. The next column, Yearly Change, displays the difference in the estimated average happiness between successive ages.

Table 2: Relationship between Fixed-Effects Estimates and Observed Data (Two Cohorts)

Age	Fixed-Effects Estimates			Observed Average Happiness		Observed Yearly Change		
	Coefficient	Average Happiness	Yearly Change	Cohort 60	Cohort 65	Cohort 60	Cohort 65	Overall
60				82.79				
61	-1.89	82.44	-1.89	80.90		-1.89		-1.89
62	-0.08	84.25	1.80	82.71		1.80		1.80
63	0.33	84.66	0.41	83.12		0.41		0.41
64	-0.66	83.67	-0.98	82.13		-0.98		-0.98
65	0.03	84.36	0.68	83.20	85.98	1.07		1.07
66	0.03	84.36	0.00	82.87	86.44	-0.33	0.46	0.00
67	0.12	84.45	0.10	82.13	87.70	-0.74	1.26	0.10
68	-0.64	83.69	-0.77	82.46	85.40	0.33	-2.30	-0.77
69	0.22	84.55	0.86	82.79	87.01	0.33	1.61	0.86
70	-1.31	83.02	-1.53	81.72	84.83	-1.07	-2.18	-1.53
71	-2.92	81.41	-1.61		83.56		-1.26	1.07
72	-3.84	80.49	-0.92		82.64		-0.92	-0.92
73	-2.46	81.86	1.38		84.02		1.38	1.38
74	-1.89	82.44	0.58		84.60		0.58	0.58
75	-4.30	80.03	-2.41		82.18		-2.41	-2.41
Const.	84.33							
N				122.00	87.00			
Weights						0.58	0.42	

A direct comparison between the columns estimated Yearly Change (Column 4) and the Overall yearly change in observed data (the last column) in Table 2 shows that the yearly change estimated by the FE regression are identical to the weighted yearly change in the observed data. A comparison between the observed average happiness (dots in Figure 3 lower graph) and the fitted regression curve (line in the same graph) shows that the yearly change in happiness estimated by the FE regression of happiness on age polynomials approximately equal the overall yearly change in the observed data. That is, *the yearly change in average happiness estimated by the FE regression of happiness in age is either exactly equal to (for age dummies) or approximately equal to (for age polynomials) the overall yearly change in average happiness in the observed data over all cohorts, weighted according to cohort size.*

The comparison at wave aggregate level from two waves of a survey produces the change in average happiness between the two waves of the survey for each cohort. Conversely, the FE regression generates an overall change in average happiness weighted over all cohorts for every single survey year.

The Root of the Mystery: A Graph with an Ambiguous Axis

Happiness is regressed on age, age squared and age cubed using the entire data (wave 1 to 11) controlling for individual fixed effects via the FE regression. The estimated average happiness, displayed as the dashed line in Figure 4, shows that on average happiness decreases by about 0.3 points per year from age 15 to 89, or decreases by 0.5% per year within each person. The dashed line also shows that the average happiness barely changed over time between the ages 40 and 63, for a given person. During the life course of a person, the average happiness decreases at a decelerating rate from age 15 to 40, then stabilizes until age 63, after which it gradually decreases at an accelerating rate.

Figure 4: Relationship between Happiness and Age from Regressions with and without Controlling for Individual Fixed-Effects (HILDA Balanced Panel)
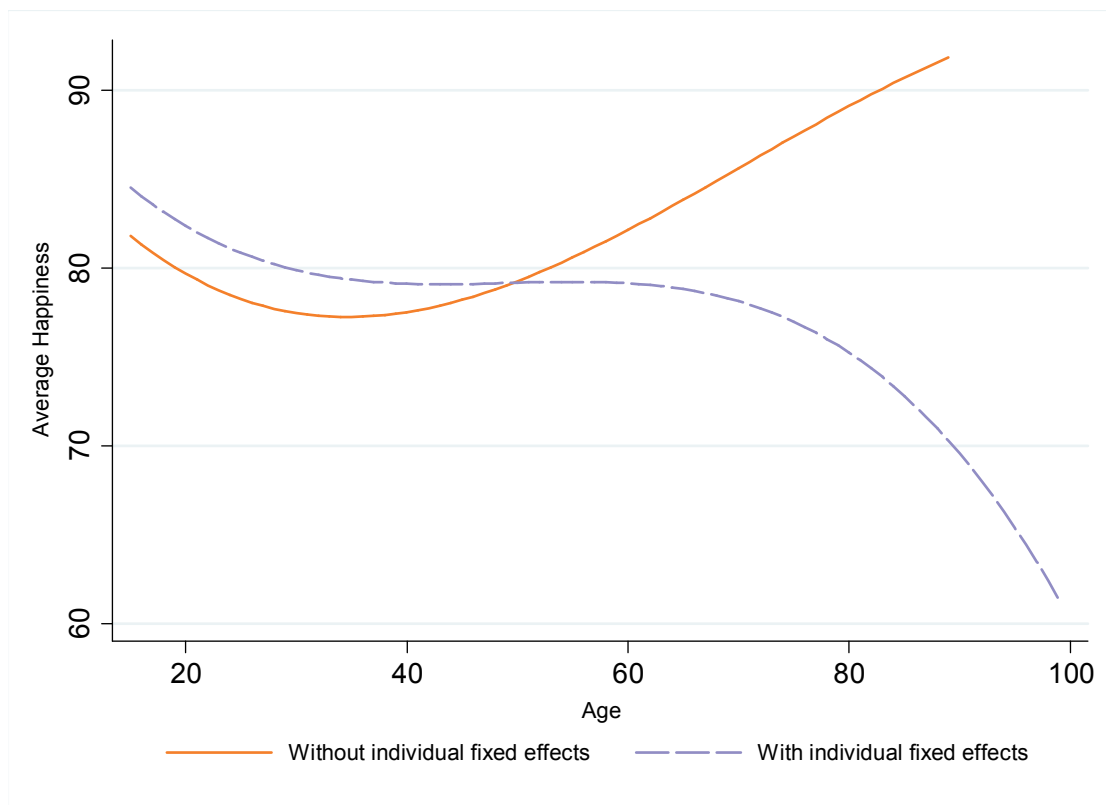
For the purpose of illustration, happiness is also regressed on age, age squared and age cubed using the data from Wave 1 only. The fitted curve from this regression is shown as the solid line in Figure 4.

In previous studies, similar graphs in which one curve was fitted from OLS regression of cross-sectional or pooled data and the other was from the FE regression were used, either explicitly or implicitly, to compare the age-happiness relationship between regressions with and without controlling for individual effects. The difference between the curves is called the age-happiness mystery. Figure 4 is a graphical illustration of the mystery.

Figure 5: Multidimensional Relationship between Age and Happiness
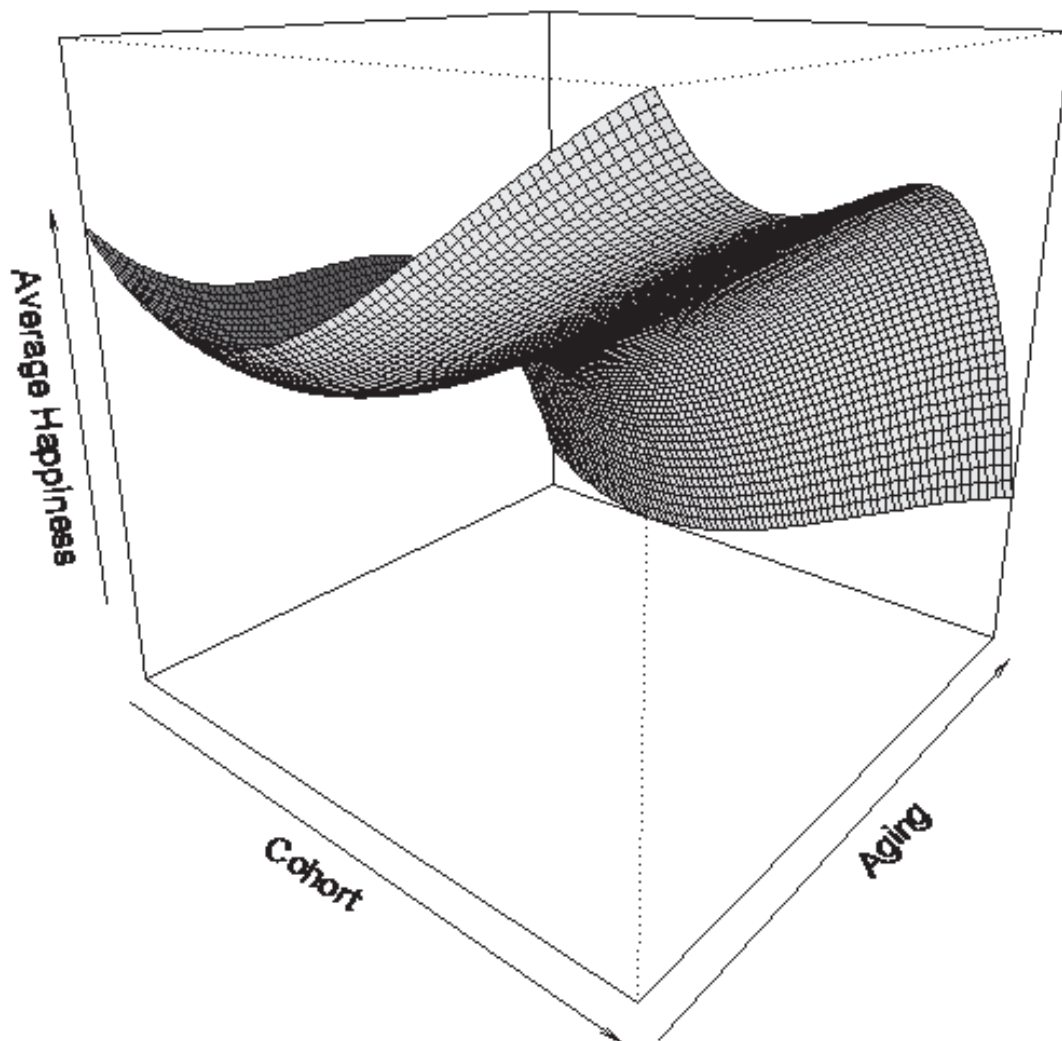
As noted earlier, different ages in cross-sectional data refer to different birth cohorts. In contrast, the age in FE regressions refers to time in years experienced by fixed cohorts. Thus the meaning of the horizontal axis in Figure 5 differs between the curves. It is this ambiguity that caused confusion in the literature. A better graph, where cohort and ageing each occupies an independent dimension, based on Wave 1 OLS and the FE regression, is given in Figure 6. The new graph shows that average happiness is U-shaped among cohorts at a given time and simultaneously decreases with ageing within a given cohort. Based on this graph, in 2001 (HILDA Wave 1) a typical 75 years old is 10% happier than a typical 50 years old. But a 50 year-old person will become 2.9% less happy when he or she turns 75. The least happy people in 2001, Cohort 35, have the lowest average happiness 77.3. By age 89, the average happiness of this cohort will decrease to 60.8.

Figure 5 unlocks the age-happiness puzzle. The U.S. General Social Surveys data used by Blanchflower & Oswald (2008) are quasi-panel data consisting of a set of random samples cumulated over multiple years, and the between-person variability dominates the pooled data. Although the first part of Blanchflower & Oswald (2008) attempts to adjust for cohort effects and the associated result is more than a simple pooled cross-sectional analysis, the reported U shape is essentially describing a cohort–happiness relationship. In contrast, Glenn's example (2009) refers to the change in average happiness over time experienced by a single cohort, hence an ageing-happiness relationship. The two aspects don't have to be numerically related. Even if they do, they differ conceptually.

Conclusion and Discussion

Cross-sectional information and longitudinal information reflect distinct aspects of the phenomenon under study. In age-happiness studies, cross-sectional information describes whether, in a particular year, people of a certain age are happier than people of other ages. Longitudinal information describes whether people become happier or less happy over the life cycle. The former compares happiness between different people, and the later compares happiness within the same person.

FE models extract longitudinal information in the data. Unlike regression of y on x, the FE regression of y on x internally conducts regression of deviations in y on deviations in x, treating each individual as a dummy.

Controlling for individual fixed effects differs from the usual sense of controlling for any observable variables. Controlling for marital status, for example, in a regression of happiness on age, the model works on the level of happiness, differentiating between marital statuses. In contrast, when controlling for individual fixed effects in FE regression of happiness on age, the model works on the within-person deviations in happiness.

While FE regressions are able to control for individual fixed effects, the FE models aim to estimate the change in response variable over time within persons. The control is only a means to make the comparison more effective.

The fixed-effects model is a standard model that has been studied extensively in statistics. With the implementation of the model in major statistical software packages, the method quickly earned popularity in analysing longitudinal data in many areas. But, the interpretation of the model seems to be less clear. This paper illustrates the meaning of the FE regressions using the variables age and happiness.

In doing so, many assumptions have been made to keep the illustration simple. These assumptions include cardinal happiness scores, a balanced panel, no control variables, and specific cohorts used. However, the interpretations illustrated hold in general. The age dummies and polynomials were used because the dummy variables directly link the regression results to the observed data, thus building the interpretation basis. The quadratic/cubic models were used because they are commonly seen in the literature. The fact that data from Australia were used doesn't prevent the interpretation of FE regressions from being applied to data collected elsewhere. The multidimensionality property of longitudinal data does not exclude the possibility that other sources may influence happiness. In the study of age and happiness, a clear distinction between different meanings of age avoids confusion and gives a fuller picture of the relationship between cohort, ageing and happiness.

REFERENCES

Alkire, Sabina. (2010). Human Development: Definitions, Critiques, and Related Concepts. *United Nations Development Programme, Human Development Reports Research Paper*, 2010/01.

Baltagi, Badi H. (2008). *Econometric analysis of panel data*. 4th Ed. Chichester, UK ; Hoboken, NJ : John Wiley & Sons, (Chapter 2).

Berg, Anne Ingeborg, Lesa Hoffmanb, Linda Hassinga, Gerald McClearnc, and Boo Johanssona. (2009). What Matters, and What Matters Most, for Change in Life Satisfaction in the Oldest-Old? A Study Over 6 Years Among Individuals 80+. *Ageing & Mental Health*, 13(2), 191-201.

Blanchflower, David, and Andrew J. Oswald. (2008). Is Well-Being U-Shaped over the Life Cycle? *Social Science & Medicine,* 66, 1733–1749.

Carstensen, Laura L., Monisha Pasupathi, Ulrich Mayr, and John R. Nesselroade. (2000). Emotional Experience in Everyday Life across the Adult Life Span. *Journal of Personality and Social Psychology*, 79 (4), 644–655.

Charles, Susan. T., Chandra A. Reynolds, and Margaret Gatz (2001). Age-Related Differences and Change in Positive and Negative Affect over 23 Years. *Journal of Personality and Social Psychology*, 80, 136–151.

Chen, Chaonan. (2001). Ageing and Life Satisfaction. *Social Indicators Research*, 54 (1), 57–79.

Cheng, Terence C., Nattavudh Powdthavee, and Andrew J. Oswald. (2014). Longitudinal Evidence for a Midlife Nadir in Human Well-Being: Results from Four Data Sets. *IZA Discussion Paper No.* 7942.

Clark, Andrew. E., and Andrew Oswald. (2006). The Curved Relationship between Subjective Well-Being and Age. *PARIS-Jourdan Sciences Economiques, working paper N°* 29.

Clark, Andrew. E. (2007). Born to be mild? Cohort Effects don't (Fully) Explain why Well-Being is U-shaped in Age. *IZA Discussion Papers, No.* 3170.

Costa, Paul T., Jr., A. B. Zonderman, Robert R. McCrae, J. Cornoni-Huntley, B. Z. Locke, and H. E. Barbano. (1987). Longitudinal analyses of psychological well-being in a national sample: Stability of mean levels. *Journal of Gerontology,* 42, 50–55.

Cummins, Robert A., Ning Li, Mark Wooden and Mark Stokes. (2013). A Demonstration of Set-Points for Subjective Wellbeing. *Journal of Happiness Studies,* DOI 10.1007/s10902-013-9444-9

Diener, Edward, Eunkook M. Suh, Richard E. Lucas, and H. L. Smith. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125, 302–376.

Easterlin, Richard A. (2003). Explaining happiness. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 100(19), 11176–11183.

Fischer, Justina AV. (2009). Happiness and age cycles – return to start. . . *Munich Personal RePEc Archive (MPRA), Paper No.* 15249.

Frijters, Paul, and Tony Beatton. (2012). The Mystery of the U-Shaped Relationship between Happiness and Age. *Journal of Economic Behavior & Organization,* 82, 525–542.

Glenn, Norval. (2009). Is the Apparent U-Shape of Well-Being over the Life Course a Result of Inappropriate use of Control Variables? A Commentary on Blanchflower and Oswald (66: 8, 2008, 1733–1749). *Social Science & Medicine,* 69, 481–485.

Goldbeck, Lutz, Tim G. Schmitz, Tanja Besier, Peter Herschbach and Gerhard Henrich. (2007). Life Satisfaction Decreases during Adolescence. *Quality of Life Research,* 16(6), 969-979.

Gwozdz, Wencke, and Alfonso Sousa-Poza. (2010). Ageing, Health and Life Satisfaction of the Oldest Old: An Analysis for Germany. *Social Indicator Research*, 97, 397–417.

Kristoffersen, Ingebjørg. (2013). The Age-Happiness Puzzle: The Role of Economic Circumstances and Financial Satisfaction. *HILDA Survey Research Conference 2013 – Papers*
http://www.melbourneinstitute.com/downloads/conferences/HILDA_2013/HILDA_2013_papers/Kristoffersen,%20Inga_final%20paper.pdf   (as on 25 Feb 2014).

Kassenboehmera, Sonja, C. and John P. Haisken-DeNew. (2012). Heresy or Enlightenment? The Well-Being Age U-shape Effect is Flat. *Economics Letters,* 117, 235–238.

Kunzmann, Ute, Todd D. Little, and Jacqui Smith. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin ageing study. *Psychology and ageing*, 15(3), 511-526.

Lacey, Heather P., Todd A. Kierstead, and Diana Morey. (2012). De-Biasing the Age-Happiness Bias: Memory Search and Cultural Expectations in Happiness Judgments Across the Lifespan. *Journal of Happiness Studies*, 13, 647–658.

Lang, Frieder R., and Jutta Heckhausen. (2001). Perceived Control over Development and Subjective Well-Being: Differential Benefits across Adulthood. *Journal of Personality and Social Psychology*, 81, 509–523.

Mroczek, Daniel K., and Christian M. Kolarz. (1998). The Effect of Age on Positive and Negative Affect: A Developmental Perspective on Happiness. *Journal of Personality and Social Psychology,* 75 (5), 1333–1349.

Mroczek , Daniel K., and Avron Spiro III. (2005). Change in Life Satisfaction during Adulthood: Findings from the Veterans Affairs Normative Ageing Study. *Journal of Personality and Social Psychology*, 88 (1), 189–202.

Myers, David G. (2000). The funds, friends, and faith of happy people. *American Psychologist*, 55(1), 56–67.

Ree, Joppe de, and Rob Alessie. (2011). Life Satisfaction and Age: Dealing with Under-Identification in Age-Period-Cohort Models. *Social Science & Medicine*, 73, 177–182.

Summerfield, Michelle, S. Freidin, M. Hahn, P. Ittak, N. Li, N. Macalalad, N. Watson, R. Wilkins, and M. Wooden. (2013). *HILDA user manual - release 12*. Melbourne, Australia: Melbourne Institute of Applied Economic and Social Research.

Van Landeghem, Bert G.M. (2008). Human Well-Being over the Life Cycle: Longitudinal Evidence from a 20-Year Panel. *LICOS Discussion Paper, No.* 213.

Wooden, Mark, and Ning Li. (2013). Panel Conditioning and Subjective Wellbeing. *Social Indicators Research*, DOI 10.1007/s11205-013-0348-1.

Wunder, Christoph, Andrea Wiencierz, Johannes Schwarze, And Helmut Küchenhoff. (2013). Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data. *The Review of Economics and Statistics,* 95(1), 154–167.

Supplementary Material

This material contains computer programs that were used to generate the results in the accompanying article. The program files are named as `AgeHappiness-stata.do` and `AgeHappiness-3D.R,` containing Stata and R codes respectively. These files can be stored in any sub-directory of your PC to execute the programs. But the Stata program should be run before the R program.

How to run the `AgeHappiness-stata.do`**?**

1. Before you can run the Stata codes, you need to have Stata 12 working in your PC.

2. You also need a license to access the data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, Release 11 or above. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). A license agreement can be obtained via the link http://www.melbourneinstitute.com/hilda/data/

3. The particular data sets used in the paper are `Combined_a110c.dta,` `Combined_b110c.dta, … …, Combined_j110c.dta,` and `Master_j110c.dta.` The letter `c` in the file names of the data sets stands for HILDA in-confidence data, the `110` stands for Release 11, the `_a, _b,…` and `_j` stand for Wave 1, 2, …, and 11 respectively.

4. This do file assumes that the above data sets are stored in the directory `"E:\HILDA 110c\STATA 110c"`. You will need to modify this path to suit your case. For example, if your data sets are stored in a directory `"C:\mydir\HILDA 110c\STATA 110c",` then please replace `"E:\HILDA 110c\STATA 110c"` with `"C:\mydir\HILDA 110c\STATA 110c"` in every instances in the program.

5. This do file creates a temporary directory "`C:\temp\Age-Happiness`" in your PC as a working directory to collect all the output, including data sets generated in the analysis, a log file, and a few graphs. The Stata command `mkdir "C:\temp\Age-Happiness"` in the first line of the program only needs to run once. After the first run, this line should be greyed out by placing a `*` in front of the line.

6. The headings in the do file are made consistent with those in the article.

7. To run the program, activate Stata, select File then select Do to open the do file, make relevant changes in Do-file Editor Window, save the changes. Then press control D while resting the mouse in the Do-file Editor Window. After the program has been executed, the program results can be read from the log file "`AgeHappiness.log`" in the working directory using WordPad or other text edit tools, and the graphs can be accessed by double clicking the file names in the working directory.

How to run the R program `AgeHappiness-3D.R`**?**

1. This program is for drawing the 3D plot in Figure 6 (Average Happiness as a Function of Cohort and Aging) in the article.

2. You need to have R Release 3.0.1 or above working in your PC. R is a free statistical software that can be obtained from the website http://www.r-project.org/

3. To run the program, activate R, go to File then select Open Script to open the file `AgeHappiness-3D.R.` Place your mouse in the R Editor Window, control A to select all the contents, and then control R to execute it. The resulting graph is saved in the working directory "`C:\temp\Age-Happiness`" and can be opened using MS Paint. The graph will also show in the R Graphics Window.

Supplementary Material I: AgeHappiness_stata.do

```
*------------------------------ begin AgeHappiness_stata.do --------------------------*


******************************************************************************************
*                                                                                       *
*  How to use this do file?                                                             *
*                                                                                       *
*  1. Before you can run the codes, you need to have Stata 12 working in your PC.        *
*  2. You also need a license to access the data from the HILDA Survey, Release 11 or    *
*     above.                                                                             *
*  3. HILDA data sets come in three formats: SAS, Stata, and SPSS. This program uses     *
*     the data in Stata format.                                                         *
*  4. This do file assumes that input data sets are stored in the directory             *
*     "E:\HILDA 110c\STATA 110c". Please modify this path to suit your case.            *
*  5. This do file creates a temporary directory "C:\temp\Age-Happiness" as a working    *
*     directory to collect all the output. The command                                 *
*     -mkdir "C:\temp\Age-Happiness"- only needs to run once.                           *
*                                                                                       *
******************************************************************************************

mkdir "C:\temp\Age-Happiness"
cd "C:\temp\Age-Happiness"

cap log close
log using "AgeHappiness.log", replace

/*=====================================================================*

AgeHappiness.do                              last modified 25/2/2014

Ning Li, Melbourne University, ning.li@unimelb.edu.au

Illustrates different meanings of "age" in OLS from single wave data and
from the fixed-effects model regression. Settles the U-shape debate,
and raises awareness of multidimensionality of longitudinal data.

Inputs: E:\HILDA 110u\STATA 110u\Combined_a110u.dta, ...,
        E:\HILDA 110u\STATA 110u\Combined_k110u
         E:\HILDA 110u\STATA 110u\Master_k110u

Output: C:\temp\Age-Happiness\NGlenn0.dta
        C:\temp\Age-Happiness\NGlenn.dta
        C:\temp\Age-Happiness\HAP.dta
         C:\temp\Age-Happiness\NGlenn.gph    (Figure 1 of Glenn 2009)
         C:\temp\Age-Happiness\obsMeanLS_w1.gph
         C:\temp\Age-Happiness\ObsMeanLS_poly_w1.gph
         C:\temp\Age-Happiness\LSvscoh_w1w11.gph
         C:\temp\Age-Happiness\FE_coh6065_dummypolyn.gph
         C:\temp\Age-Happiness\FE_coh65.gph
         C:\temp\Age-Happiness\LSvsAge_OLS_FE.gph
         C:\temp\Oswald\Fig3c.gph   (Figure 3c of Cheng et al 2014)
         C:\temp\Oswald\Fig3c_16to85.gph (Quadratic fit in a longer age range)
         C:\temp\Age-Happiness\AgeHappiness.log

Article: Multidimensionality of Longitudinal Data: Unlocking the
        Age-Happiness Puzzle

Run on: Stata 12 (Windows 64bits)

*=====================================================================*/

*-------------------------------*
*                               *
*  Section:  Introduction       *
*                               *
*-------------------------------*
```

```
//  Replication of Glenn's(2009) Example

/*

Norval Glenn* (2009) -
Is the apparent U-shape of well-being over the life course a result of inappropriate
use of control variables? A commentary on Blanchflower and Oswald (66: 8, 2008, 1733-1749)

Extracted from the paper:

"Just how using marital status as a control variable in an age and happiness study can
distort estimates of the effects of the former on the latter is illustrated by a
simple exercise with hypothetical data based on some extreme assumptions.
Suppose that a sample of 100 young adults born in the same year is studied each year
as they age from 21 to 29 years old. Assume further that:
 a. growing older has no effect on happiness;
 b. no one's happiness changes for any reason during the period covered;
 c. when happiness is measured on a 10 point scale (0-9), 10 persons fall on each
    point on the scale;
 d. at age 21, only the 10 happiest persons are married;
 e. each subsequent year, the 10 happiest of the remaining unmarried persons marry,
    so that by age 29, only the 10 least happy persons have not married;
 f. no one divorces during the course of the study.

Taking the data set based on these assumptions and regressing happiness on age with
marital status controlled creates the strong negative relationship between age and
happiness shown by the lower line in Fig. 1. Even though age does not affect happiness,
controlling marital status creates the impression that it does."

*/

  // Data preperation for Glenn's example

cap clear
qui set obs 100
qui gen byte id=_n
label variable id "Person ID"

egen byte h21= fill(1/10 1/10)
label variable h21 "Happiness at age 21"
forvalue j=22/29 {
  qui gen byte h`j'= h21
  label variable h`j' "Happiness at age `j'"
}
//

label define ms 0 "[0] Unmarried" 1 "[1] married"
qui gen byte m21=0
qui replace m21=1 if h21==10
label variable m21 "Marital status age 21"
label value m21 ms

local k=9
forvalue j=22/29{
  local i=`j'-1
  qui gen byte m`j' = m`i'
    qui replace m`j'=1 if h`j'==`k'
  local k =`k'-1
  label variable m`j' "Marital status age `j'"
  label value m`j' ms
}
//
br
qui save NGlenn0.dta, replace

use NGlenn0.dta, clear
keep id m*
qui reshape long m, i(id) j(age)
qui gen wave= age-20
order id wave age m
sort id wave
qui save temp1.dta, replace

use NGlenn0.dta, clear
keep id h*
qui reshape long h, i(id) j(age)
```

```
qui gen wave= age-20
order id wave age h
sort id wave
qui merge 1:1 id wave using temp1.dta
drop _merge

label variable wave "Wave number"
label variable age "Age"
label variable h "Happiness score"
label variable m "Marital status"

xtset id wave
sort id wave
notes: Data as N.Glenn,Social Science&Medicine, Vol69(4), 2009, 481-485
qui save NGlenn.dta, replace

*------------ end data preperation for Glenn's example ----------*

use NGlenn.dta, clear
des
note

  // Conditional & marginal mean happiness - observed

tab wave m, s(h) nost


  // regress happiness on age without any control

reg h age
cap drop hhat1
qui predict hhat1

  // regression with marital status being controlled for

xi: reg h age i.m
cap drop hhat2
qui predict hhat2 if m==0

twoway (scatter hhat1 age if m==0, mc(lavender)) (scatter hhat2 age if m==0, mc(orange)) ///
  , ti("Average Happiness Over Age With & Without Controlling for Marital Status", ///
     size(small)) ///
     subti("N.Glenn, Social Science & Medicine, 69(2009) 481-485",size(small)) ///
  yti("Average Happiness",size(small)) ///
  xti("Age",size(small)) ///
  legend(label(1 "Without control") label(2 "With control")) ///
    saving("NGlenn.gph",replace)


  // The OLS without and with controlling for marital status are equivalent to
  // the observed unconditional and conditional mean happiness respectively:

  //  Observed mean(h) - without control

  tab age, s(h)

  // Observed mean(h) - with control

  tab age if m==0, s(h)

  //  My comments:
/*

Blanchflower & Oswald's (2008,2009) study is based on cross-sectional data,
where different age refers to different birth cohorts. In their words
(2008, p1736. line 5-7 left column)
"Our data do not follow the same person longitudinally. Instead, we use
statistically representative snapshots year after year."

Glenn's example is for a single cohort over time. In his data, different age
refers to different time period experienced by the same person.
Cohort and aging are distinct aspects. It is quite OK that a data set
is U-shaped in one aspect and downwards in another.

*/

*-------------- end Glenn's(2009) Example ------------------*
```

```
*----------------------------------------------------------------------*
*                                                                      *
* Section I: Age Difference in Cross-Sectional Data is Cohort Difference *
*                                                                      *
*----------------------------------------------------------------------*

// Data preperation - HILDA overal life satisfaction w1-w11

//  Get original data from Release 11

local j=1
qui foreach q in a b c d e f g h i j  {
    use xwaveid `q'hgsex `q'hgage `q'losat `q'ghmh `q'hgivw ///
    using "E:\HILDA 110u\STATA 110u\Combined_`q'110u", clear // <----- change path to suit you

    rename `q'* *
    gen wave=`j'
    recode losat ghmh (-10/-1 = .)
     qui save temp`j'.dta,replace
    local j=`j'+1
}
//
use xwaveid khgsex khgage klosat kghmh khgivw khhtup ///
    using "E:\HILDA 110u\STATA 110u\Combined_k110u", clear  // <----- change path to suit you
rename k* *
qui recode losat ghmh (-10/-1 = .)
qui gen wave=11
qui save temp11.dta,replace

append using temp10.dta
append using temp9.dta
append using temp8.dta
append using temp7.dta
append using temp6.dta
append using temp5.dta
append using temp4.dta
append using temp3.dta
append using temp2.dta
append using temp1.dta

qui keep if hgivw==1 | hgivw==2
qui destring xwaveid, gen(id)

rename hgage age
qui gen age2=age^2
label variable age2 "Age squared"
qui gen age3=age^3
label variable age3 "Age cubed"
qui gen ls=10*losat      // rescale to (0,100) range
label variable ls "happiness in 100-scale"
qui xtset id wave
sort id wave
notes: R11 respondents happiness data
qui save HAP.dta, replace

forvalue j=1/11 {
  erase temp`j'.dta
}
//

//  Get interview pattern

use xwaveid ivwptn using "E:\HILDA 110u\STATA 110u\Master_k110u", clear // <----- change path to suit
you
qui destring xwaveid, gen(id)
drop xwaveid
sort id
merge 1:m id using HAP.dta

qui keep if _merge==3
drop _merge
sort id wave
qui save HAP.dta, replace

//  Flag compact panel
```

```
egen flg=count(losat), by(id)
qui gen ls_comp = 0
qui replace ls_comp=1 if ivwptn=="XXXXXXXXXXX" & flg==11
drop flg
label variable ls_comp "Compact panel w1-11 for ls"

// Generate cohort indicator

qui gen coh = age - wave +1
label variable coh "Cohort as defined by age in wave 1"

keep id wave coh age age2 age3 ls ls_comp ivwptn
order id wave coh age age2 age3 ls ls_comp ivwptn
xtset id wave
sort id wave
qui save HAP.dta, replace

*------------------- end data preperation -------------------*


// A small proportion of data

    // Long form

use HAP.dta, clear
list id wave age ls in 1/33

    // Wide form

qui keep id wave coh ls
qui reshape wide ls, i(id) j(wave)
list id ls* in 1/5


// Distribution of cohorts - balanced panel

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"
su coh if wave==1, d
tab coh if wave==1


// Observed mean happiness Wave 1 at each age

tab coh if wave==1, s(ls)

egen mls = mean(ls), by(age wave)
label variable mls "Mean Happiness by age and wave"
tab coh if wave==1, s(mls)

twoway (scatter mls age if wave==1) ///
    , /*title(Average Happiness vs Age in Wave 1,size(mediam))*/ legend(off) ///
    xti("Age in Wave 1",size(small)) yti("Average Happiness",size(small)) ///
        saving(obsMeanLS_w1, replace)


// OLS regression of happiness on age dummies

xi: reg ls i.age if wave==1
cap drop y
qui predict y if e(sample)

// Based on the regression, the mean at age a is calculated as _b[_cons] + _b[_Iage_a]
// This can be compared to the direct results in the table

di _b[_cons] + _b[_Iage_19]


// Key points: age here refers to birth cohort


// OLS of happiness on age, age square and age cubic

reg ls age age2 age3 if wave==1
cap drop z
qui predict z if e(sample)
```

29

```
// Average happiness of W1 vs Age in W1 (Figure 1)

twoway (scatter y age if wave==1) ///
       (line z age if wave==1,sort lc(orange)) ///
     , /*title(Average Happiness vs Age in Wave 1,size(mediam)) */ legend(off) ///
     xti("Age in Wave 1",size(small)) yti("Average Happiness",size(small)) ///
     saving("ObsMeanLS_poly_w1.gph", replace)




// Pooled data regression gives similar results as single wave data regression
// when the between-person variation is larger than within-person variation

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"

reg ls age age2 age3 if wave==1
cap drop z
qui predict z if e(sample)

reg ls age age2 age3 if wave==11
cap drop zz
qui predict zz if e(sample)

reg ls age age2 age3
cap drop zzz
qui predict zzz if e(sample)

twoway (line z age if wave==1,sort lc(orange))(line zz age if wave==11,sort lc(blue)) ///
       (line zzz age,sort lc(green)) ///
     ,xti("Age",size(small)) yti("Fitted values",size(small)) ///
     legend(label(1 "W1 data") label(2 "W11 data") label(3 "Pooled data")) ///
     ti("btw SD dominates, pooled data give similar fits to single wave data",size(median))


*-------- Understanding btw & wth SD in longitudinal data ---------*

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"

xtsum ls

// Between SD is the standard deviation of all persons' mean scores.
// It reflects the difference in happiness btw different people.
// There are 79488 scores from 7229 persons in the balanced panel.
// The between SD for happiness is calculated as follows.

cap drop xi
qui egen xi= mean(ls), by(id)
label variable xi "Person time-average"
qui keep id  xi
qui duplicates drop
su xi



// The within SD is the standard deviation of all person's deviates.
// It is the variability of variabilities.
// The within SD is calculated over 79488 person-years, as follows.

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"
qui egen xi= mean(ls), by(id)
cap drop xbar
egen xbar=mean(ls)
label variable xbar "Grand mean happiness"
qui gen y=ls -xi+xbar
label variable y "Deviation of person score from grand and person means"
su y

// N is total nbr of non-missing x
codebook ls if ls<.

// n is total nbr of distinct id with non-missing x
codebook id if ls<.
```

30

```
// T-bar is the average number of years a person was observed in the data
di 79488/7229

xtsum ls

di 100*(r(sd_b)-r(sd_w))/r(sd_w)

* Interpretation: If two persons were selected at random from the data, the difference in H btw
*               them is expected to be 11.7% larger than the difference in H btw two randomly
*               selected years for a single person.


*------------- end understanding btw & wth SD -------------------*




*------------------------------------------------------------------------*
*                                                                        *
*  Section II:  Age Difference in FE Regressions is Time of Years        *
*               Experienced by Fixed Cohorts                             *
*                                                                        *
*------------------------------------------------------------------------*

// A. Longitudinal Comparison at Wave Aggregate Level

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"

  // OLS of happiness on age polynomial - Wave 1

reg ls age age2 age3 if wave==1
cap drop y1 y1_se a1 b1
qui predict y1 if e(sample)
label variable y1 "Fitted value w1"
qui predict y1_se if e(sample), stdp
qui gen a1 = y1 - 1.96*y1_se if e(sample)
qui gen b1 = y1 + 1.96*y1_se if e(sample)

  // OLS of happiness on age polynomial - Wave 11

reg ls age age2 age3 if wave==11
cap drop y11 y11_se a11 b11
qui predict y11 if e(sample)
label variable y1 "Fitted value w11"
qui predict y11_se if e(sample), stdp
qui gen a11 = y11 - 1.96*y11_se if e(sample)
qui gen b11 = y11 + 1.96*y11_se if e(sample)


  // w1 curve & w11 curve aligned with cohort, pointwise 95% CI limites (Figure 2)

twoway (line y1 coh if wave==1, sort lp("_##") lc("orange") lw("medthick") ) ///
       (line y11 coh if wave==11, sort lp("solid") lc("lavender")  lw("medthick") ) ///
      (line a1 coh if wave==1,sort lp("-#") lc("orange") lw("vvthin") ) ///
      (line b1 coh if wave==1,sort lp("-#") lc("orange") lw("vvthin")) ///
      (line a11 coh if wave==11,sort lp("-#") lc("lavender")lw("vvthin") ) ///
      (line b11 coh if wave==11,sort lp("-#") lc("lavender") lw("vvthin")) ///
      ,/*ti("Fitted regression curves W1 & W11",size(small))*/ legend(off) ///
      yti("Average Happiness",size(small)) xti("Age in Wave 1",size(small))  ///
      saving("LSvscoh_w1w11.gph", replace)



// B. Longitudinal Comparison at the Person Level - via FE regression


// The FE regression of y on x is just a shorter name for the OLS regression of
// the deviates of y on the deviates of x. (NB: SE are different due to df)

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"

egen temp=mean(ls),by(id)
egen temp2=mean(ls)
gen dls=ls-temp + temp2
drop temp temp2
```

```
egen temp=mean(age),by(id)
egen temp2=mean(age)
gen dage=age-temp+temp2

drop temp temp2
egen temp=mean(age2),by(id)
egen temp2=mean(age2)
gen dage2=age2-temp+temp2

drop temp temp2
egen temp=mean(age3),by(id)
egen temp2=mean(age3)
gen dage3=age3-temp+temp2


    // OLS dls on dage dage2 dage3

reg dls dage dage2 dage3

  // FE reg ls on age age2 age3

xtreg ls age age2 age3, fe

    // OLS using dummy for each person return right df (se)

areg ls age age2 age3, absorb(id)


// What exactly the FE regression does?

  // Single Cohort Case

    // FE regression on age dummies

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"
keep if ls_comp==1
codebook coh if wave==1
qui keep if coh==65

xi: xtreg ls i.age, fe   // (Table 1)
cap drop y
qui predict y if e(sample)


    // Observed mean happiness in age grps:

cap drop mls_a
egen mls_a=mean(ls), by(age)
label variable mls_a "Mean happiness by age (W1-W11)"  // NB: this is for a single cohort!

tab age, s(mls_a)

    // The fitted values from FE on age dummies = observed means

tab age, s(y)

    // FE regression on age polynomials

xtreg ls age age2 age3, fe
cap drop y2
qui predict y2 if e(sample)

   // ( Figure 3, upper graph )
   // Fitted Values from the Fixed-Effects Regression of Age Polynomials
   // Approximate Average Happiness in the Observed Data

twoway (scatter mls_a age) (line y2 age, lc(orange)  sort) ///
       , /*tit("Fixed-effects regression of happiness (Cohort 65)")*/ ///
        xti("Ageing",size(small)) yti("Average Happiness",size(small)) ///
         legend(off) ///
         saving("FE_coh65.gph", replace)

* Summary: estimated difference in mean happiness btw consecutive ages = obs differece in
*          mean happiness btw consecutive ages when age dummies are used.
*          As a bonus, in single cohort case, the estimated
*          constant = the mean happiness in the youngest age, age 15 here.
```

```
*           When age polynomials are used, the estimated difference in mean happiness btw
*           consecutive ages approximately equal the obs differences.


// Multiple Cohort Case
// Relationship btw FE estimates & observed data

//  Two cohorts 60 & 65 (Table 2)

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"
qui keep if ls_comp==1
qui keep if coh==60 | coh==65
tab wave coh

//  FE regression on age dummies

xi: areg ls i.age, absorb(id)   // (Table 2 - column 2)
cap drop y1
qui predict y1 if e(sample)

// observed mean happiness by age & cohort grps:

egen mls_ac=mean(ls), by(age coh)

tab age coh, s(mls_ac) nost nofreq  // (Table 2 - column 5 & 6)


// FE regression on age polynomials

areg ls age age2 age3, absorb(id)
cap drop y2
qui predict y2 if e(sample)

//  ( Figure 3, lower graph )
twoway (scatter y1 age) (line y2 age, lc(orange) sort)  ///
      ,/*tit("Fitted values from age polynomials approximate that from age dummies",size(median))*/
///
     xti("Aging",size(small)) yti("Average Happiness",size(small)) ///
         legend(off) ///
         /* legend(label(1 "FE using age dummies") label(2 "FE using age polynomials")) */ ///
         saving("FE_coh6065.gph", replace)

* Summary: Differences in fitted mean happiness btw consecutive ages = weighted differences in obs
*          means btw consecutive ages. The weights are proportional to relative cohort sizes.
*          The fitted values when ploted againt age tell how mean happiness change from
*          one year to the next within cohorts. The mean is averaged over all cohorts.
*          When age polynomials replace age dummies, the differences in fitted values
*          btw consecutive ages approximately equal the observed change in mean happiness
*          btw concecutive years, within cohorts.


// Multiple graphs in the same plot

graph combine FE_coh65.gph FE_coh6065.gph, iscale(1) col(1) xcommon ycommon ///
     , saving("FE_combined.gph",replace)

* Notes: The dots in upper graph represent observed average happiness at each age for cohort 65.
*        The dots in lower graph represent average happiness at each age for both cohorts.
*        The line in upper graph is estimated average happiness at each age from FE regression
*        of happiness on age age2, age3 for cohort 65.
*        The line in lower graph is estimated average happiness at each age from FE regression
*        of happiness on age age2, age3 for cohorts 60 & 65.




*-----------------------------------------------------------------------*
*                                                                       *
* Section III: The Root of the Mystery: A Graph with an Ambiguous Axis  *
*                                                                       *
*-----------------------------------------------------------------------*

// OLS of happiness on age polynomials without controlling for individual FE (W1)

use HAP.dta, clear
qui keep if ivwptn=="XXXXXXXXXXX"
```

```
qui keep if ls_comp==1

reg ls age age2 age3 if wave==1
cap drop y1
qui predict y1 if e(sample)
label variable y1 "Fitted values from OLS on W1 data"

//  OLS of happiness on age polynomials controlling for individual fixed-effects

xtreg ls age age2 age3, fe
cap drop y2
qui predict y2 if e(sample)
label variable y2 "Fitted values from FE regression"


// Regressions with & without controlling for individual fixed-effects (Figure 4)

twoway (line y1 age if wave==1,lc(orange) sort) (line y2 age, lc(lavender) lp(longdash)sort) ///
       ,xtit("Age",size(small)) ytit("Average Happiness",size(small)) ///
         /*title("A Graph with an Ambiguous Axis") */ ///
      legend(label(1 "Without control") label(2 "With control for indiv.fixed effects")) ///
         saving("LSvsAge_OLS_FE.gph", replace)



// A better graph by separating different meanings of age

*  see R script for (Figure 5)



*-------------------------------------------------------------------------*
*                                                                         *
* Section IV: Conclusion and Discussion                                   *
*                                                                         *
*-------------------------------------------------------------------------*


*********************** (end do) *****************************

log close

*---------------------------- end AgeHappiness_stata.do ----------------------------*
```

Supplementary Material II: AgeHappiness_3R.R

```
#------------------------------ begin AgeHappiness_3R.R -------------------------#

#========================================================================#
#                                                                        #
#   3D plot - Mean life satisfaction as a function of cohort and age     #
#                                                                        #
#   File:  threeD.R                         last modified  10Mar2014     #
#                                                                        #
#   By:    Ning Li, Melbourne University, Ning.Li@unimelb.edu.au         #
#                                                                        #
#   Input: Stata results from OLS regression of happiness on age,        #
#          age sq, age cubed W1                                          #
#          Stata result from the FE regression of happiness on age,      #
#          age sq, age cubed W1-W11                                      #
#                                                                        #
#   Output: 3D graph "C:/temp/Age-Happiness/CAH.wmf"                     #
#                                                                        #
#   Run on: R 3.0.1 Windows                                              #
#                                                                        #
#========================================================================#


setwd("C:/temp/Age-Happiness")

x=15:89; y=15:101

a=x; b=y;
plot(a, 93.09159 -1.02549*a +  .0196367*a^2 -.0000932*a^3) #  min=(age=35,h=77.3),
max=(age=89,h=91.7)

a=35
93.09159  -1.02549*a +  .0196367*a^2 -.0000932*a^3 # 77.26
a=89
93.09159  -1.02549*a +  .0196367*a^2 -.0000932*a^3  # 91.66

b=15
96.08303 -1.062614*b +  .0219588*b^2  -.0001494*b^3 # 84.58
b=100
96.08303 -1.062614*b +  .0219588*b^2  -.0001494*b^3 # 60.01

(84.58-60.01)/(100-15)  # = 0.289

 # btw age 15-89, h decreased about 0.3 points per yr
 # For coh35, w1 mean h is min=77, by age 89 (ie,54 yrs later),
 # mean h decreases to 77-0.3*54=60.8. This is the lowest mean h.


coh=function(a){
 res=93.09159 -1.02549*a + .0196367*a^2 -.0000932*a^3
 return(res)
}

coh(70)
coh(50)
(coh(70)-coh(50))/coh(50) # In W1,av. H is 7.9% higher in coh(70) than coh(50)
(coh(75)-coh(50))/coh(50) # In w1, a 75 yo is 10% happier than a 50 yo.



aging=function(b){
 res=96.08303 -1.062614*b +  .0219588*b^2  -.0001494*b^3
 return(res)
}
aging(70)
aging(50)
(aging(70)-aging(50))/aging(50) # A person is 1.4% less happy in 70 yo than in 50 yo.
(aging(75)-aging(50))/aging(50) # a person is 2.9% less happy in 75 yo than in 50 yo.
```

```
#--------- Results from AgeHappiness.do Section III, (1) (2) ----------#
#   w1 OLS (1): ls=93.09159 -1.02549 Age + .0196367 Age^2 -.0000932 Age^3
# w1-11 FE (2): ls=96.08303 -1.062614 Age +.0219588 Age^2 -.0001494 Age^3

# z= (w1 OLS)*(w1-11 FE):

x=15:89; y=15:101
z=outer(x,y,function(a,b)
  (93.09159  -1.02549*a +  .0196367*a^2 -.0000932*a^3)*
  ( 96.08303 -1.062614*b +  .0219588*b^2  -.0001494*b^3) )


#--- save plots in wmf and pdf files using col & border (Figure 6) ------#

win.metafile("CAH.wmf")
persp(z,phi=19,theta=40,xlab="Cohort",ylab="Aging",zlab="Average Happiness",col=3,border=4)
dev.off()

persp(z,phi=19,theta=40,xlab="Cohort",ylab="Aging",zlab="Average Happiness",col=3,border=4)

#=================== end 3D w1 OLS & w1-11 FE plots ===============#



#------------------------------- end AgeHappiness_3R.R ---------------------------#
```