Is there a W-shape in Wellbeing across the Lifespan? A Misclassification Approach*

Ekaterina Oparina†
London School of Economics

March 25, 2021

Abstract

This paper re-examines the U-shaped relations between life satisfaction and age while controlling for the presence of measurement error in reported life satisfaction. My dataset comes from Wave 3 of UK Understanding Society that was surveyed between 2011 and 2013. I use the novel approach by Oparina and Srisuma (2021) and show that in this dataset latent, presumably error-free, life satisfaction was W-shaped. I estimate a pronounced adverse mid-life effect on LS between 50-55 years old, and a nadir at a younger age group, between 25 and 30 years old. I estimate the reporting pattern of the latter to explain the discrepancies between their reports and the latent state. Low levels of wellbeing in this group are likely to come from a combination of age effect, i.e. the quarter-life crisis, and cohort effects, these are the young professionals hit among the hardest by the global economic crisis. I conjecture that unwillingness of young adults to report low wellbeing can be linked to derogatory attitude to this age group as a *snowflake generation*.

JEL Classification Numbers: C14, C51, I31

Keywords: Measurement error, identification, subjective wellbeing, testing.

*I am most grateful to Sorawoot Srisuma for various discussions and suggestions. I thank Jo Blanden, Valentina Corradi, the seminar participants at IAAE Conference (2019), and the RES Junior Symposium (2019) for helpful comments and discussions.

†*E-mail address*: e.oparina@lse.ac.uk

1 Introduction

The U-shape of reported life satisfaction (LS) over the lifespan is largely a consensus in the economics of wellbeing literature. This empirical regularity is supported by a growing body of evidence coming from various countries, datasets, time periods and analytical approaches. Researchers typically observe a decline in reported wellbeing till the mid-age and an increase after. In practice, the presence of this phenomenon is usually tested by including both linear and quadratic terms for the respondent's age and then the conclusions are based on the signs of the coefficient. Most contemporary wellbeing models use the quadratic specification to control for age effect and, hence, rely on its validity for identifying the effects of other factors on individual wellbeing.

Asking respondents to report their state remains the most practical and authentic way of evaluating individual wellbeing. However, answering these questions is known to be a complex cognitive task that requires performing multiple comparisons and evaluations (see e.g., Diener et al. (2013)). To minimise the load, respondents might use mental shortcuts, e.g. some information that is easily accessible but might not be directly relevant to global LS. Those can include current mood, passing events or social norms (see, e.g., Schwarz and Clore (1983), Feddersen et al. (2016), Diener et al. (2013)). This behaviour is accurately summarised by the *cognitive miser* metaphor. The effect of these external factors can be seen as a measurement error, see e.g. Bertrand and Mullainathan (2001). Then, one can interpret *reported* LS as a possible mismeasurement of *latent* LS, where measurement error combines *irrelevant factors* that may affect the reporting of latent LS. Given that the LS responses are recorded on a discrete scale I refer to the measurement error as a misclassification. Oparina and Srisuma (2021) show that this type of misclassification can affect the conclusions of wellbeing analysis.

In this paper, I test if the U-shape, recorded for reported LS, holds for latent LS, which is presumably free of the effects of external factor that are not relevant to global wellbeing. I interpret latent LS as a hypothetical response that we would receive if the respondent spent all the time necessary and gave full mental effort to judge their life satisfaction. I follow the approach outlined in Oparina and Srisuma (2021) to identify the distribution of the latent LS and study the relationship between latent LS and age. I refer the reader to the aforementioned paper for a detailed account of identification and estimation strategies.

I use data from the Wave 3 of the UK Understanding Society survey, taken between January 2011 and June 2013. I do not find support for the U-shape of latent LS using quadratic specification for age. What is more, I find that in my dataset latent LS is W-shaped. To identify this shape I adopt a more flexible modelling framework. By splitting the respondents into 5-year age bands and including a dummy for each group I allow LS to depend on age arbitrarily. I find evidence supporting a pronounced adverse mid-life effect on LS for those between 50-55 years old. However, I find that this age is not the nadir of LS. I locate the nadir to be at a younger age group, between 25 and 30 years old, which corresponds to what is sometimes known as the *quarter-life crisis*. The low levels of wellbeing of this group could also be reinforced by the disproportional effect of the global economic crisis. The difference between the nadir points is explained by the fact that younger respondents tend to over-report their LS when they are in fact less satisfied.

My approach does not allow me to identify the reasons for this misreporting. Without further research, I can only offer potential explanations. One particular conjecture is based on the effect of stereotypes, in particular, that it is easier for respondents to report their latent state if it is socially acceptable. The mid-life crisis receives a lot of attention in academic and popular culture, which normalises reporting lower wellbeing states at this age. The respondents in their 20s, on the opposite, are sometimes referred to as the *snowflake generation* when showing signs of distress.[1] Hence, using the socially acceptable state as a *mental shortcut* can explain the misreporting pattern that I observe.

The rest of the paper is organised as follows. Section 2 provides background on wellbeing dynamic over the lifespan. The data is discussed in Section 3. Section 4 presents evidence of the presence of misclassification. It also presents the results of the study of the relationship between age and LS and discusses potential explanations of the findings. Section 5 concludes. Appendix provides robustness checks and supplementary results to support the empirical findings.

2 The U-shape of life satisfaction over age

The U-shape of reported LS over age is largely a consensus in the economics of wellbeing literature and it is supported by the rapidly growing body of evidence, see e.g. Blanchflower and Graham (2020). Despite some exceptions (Frijters and Beatton (2012)), researchers typically observe a decline in reported wellbeing from a young age, a low between the age of 32 and 50 depending on the dataset, and an increase after. This pattern in reported wellbeing was found by multiple authors in various datasets from around the globe, including North America, Western and Eastern Europe, Latin America, and Asia, see e.g. Blanchflower and Oswald (2008). This empirical phenomenon appears to hold for both cross-sectional and panel analysis (Winkelmann and Winkelmann (1998), Cheng et al. (2017), Clark (2019)). It also stays intact whether LS is treated as a cardinal variable, so the OLS is used (Di Tella et al. (2001)), or when it is treated as an ordinal variable by ordered choice models, usually logit or probit (Blanchflower and Oswald (2008)).

The lowest point corresponds to a period referred to as the *mid-life crisis*, a concept which originates from the psychology literature. Although there seems to be little consensus in the contemporary psychological literature, see e.g. Dear et al. (2002), it is widely exploited in popular outlets and popular culture. The term was coined in 1965 by Elliott Jaques and originally meant a decline in productivity around the age of 35 (Jaques (1965)). However, it is now more commonly used for

[1]The term *snowflake generation* was one of Collins English Dictionary's 2016 words of the year. Collins defines the term as "the young adults of the 2010s, viewed as being less resilient and more prone to taking offence than previous generations".

an identity and self-confidence crisis during the late 40s – early 50s. Some psychological research, however, suggests that this is not a phase that most middle-aged people experience, and questions the existence of this phenomenon overall, see e.g. Kruger (1994).

Several wellbeing studies support the U-shape in reported LS with alternative, presumably more objective, measures. Blanchflower and Oswald (2016) show that the probability of taking antidepressants follows an inverted U-shape curve that peaks in the late 40s. If taking antidepressants is negatively correlated with reported wellbeing, this finding supports the pattern. However, the authors also find that women in Western Europe are more likely to take antidepressants than men, though women systematically report being more satisfied with their lives. The correlation seems positive for women, which might e.g. mean that using medication improves subjective wellbeing. The direction of the correlation is not exactly clear. What is more, this analysis relies on a self-reported measure of the use of antidepressants, so it might as well be influenced by reporting behaviour artefacts. Weiss et al. (2012) show that a similar U-shape exists for 508 great apes (two samples of chimpanzees and one sample of orangutans). However, the authors use subjective reports of raters, who assess the wellbeing of apes. Hence, the reports of raters about the wellbeing of apes may be influenced by external factors in the same way as individual's reports about their own wellbeing. Then, individual stereotypes about life-long happiness dynamics might distort the reported scales.

There is ample evidence that reported wellbeing follows a U-shape throughout one's life. In the remainder of this paper, I investigate whether this pattern stays in place for latent LS, which is presumably free from the effect of mental shortcuts.

3 Empirical strategy

I use Wave 3 of the UK representative household study Understanding Society, which was collected between January 2011 and June 2013. I only use data on the respondents who reported their satisfaction with life overall. I drop the respondents above age 90 (75 observations), given that a low number of observations in this category does not allow me to perform constrained maximum likelihood to estimate the conditional distribution of latent LS. The final sample contains 40,284 respondents.

My empirical strategy follows the two-step procedure from Oparina and Srisuma (2021). I first identify the distribution of latent LS for different age groups (see Hu (2008)) and then use it to identify the parameters of the model with latent, X^*, instead of reported, X, LS as a dependent variable. This procedure requires carefully selecting two auxiliary variables, Y and Z. Selecting these variables is similar to selecting a good instrument. One the one hand, they should contain information on latent LS. On the other hand, they should be independent of the measurement error in reported LS that I control for, once I account for age. I use a derived measure of neuroticism, which is indicative of a respondent's emotional stability, as Y. And a measure of mental state that comes from the General Health Questionnaire (GHQ-12) as Z. I refer the reader to Oparina and Srisuma (2021) for the detailed discussion of the identification strategy and validity of the selected variables.

I reduce the support of the variables to guarantee that the algorithm that estimates the latent distributions is numerically stable. In particular, latent LS is recorded on the following 3-point scale:

$$X^* = \begin{cases} 1 & \text{dissatisfied} \\ 2 & \text{neither satisfied nor dissatisfied} \\ 3 & \text{satisfied} \end{cases}.$$

I group all the respondents between the age 20 and 70 in 5-year bands. Respondents under 20 and over 70 are grouped in the respective categories. I obtain 12 subgroups. As discussed further, different model specifications include a set of dummies for age groups or a vector of mid-values of the groups and their squares. In the following section, I report the estimates of the conditional distribution of the latent LS for respondents in each age band. I then use it to provide a comprehensive description of how latent LS changes over the lifespan, and how it differs from reported LS for people of different age.

4 Results

I begin this section by presenting the evidence of the presence of measurement errors in different age groups. In Section 4.2 I study how measurement error affects the relationship between age and LS. I start with modelling the quadratic effect of age on LS, as is standard in the literature. I then allow for a more flexible relationship by including age band dummies instead of a linear and a quadratic term for age.

4.1 Measurement error in reported LS

I test for the presence of measurement error for different age groups, following Wilhelm (2018) and Oparina and Srisuma (2021). Each group is used separately to construct test statistics and bootstrap critical values at different significance levels. I find very strong evidence for the presence of measurement error and reject the hypothesis of no measurement error at 1% significance level for all age groups (Table 1). However, the U-shape relation might hold even in the presence of measurement error if the misreporting patterns are similar across the age groups. I now investigate whether the analysis, that explicitly accounts for the presence of measurement error, supports the U-shape.

Age	TS	Critical values			N
		90%	95%	99%	
<20	0.097***	0.030	0.036	0.046	2,613
20-24	0.090***	0.028	0.033	0.041	2,690
25-29	0.075***	0.027	0.031	0.039	2,777
30-34	0.110***	0.026	0.030	0.035	3,293
35-39	0.114***	0.024	0.027	0.034	3,537
40-44	0.114***	0.021	0.024	0.029	4,087
45-49	0.112***	0.022	0.026	0.032	3,846
50-54	0.093***	0.023	0.026	0.032	3,611
55-59	0.129***	0.025	0.029	0.036	3,065
60-64	0.089***	0.026	0.030	0.037	3,190
65-69	0.104***	0.031	0.034	0.044	2,798
≥70	0.087***	0.025	0.029	0.038	4,777

* p<0.10, ** p<0.05, *** p<0.01

Table 1: Conditional test for the presence of measurement error for different age bands

4.2 Life satisfaction across lifespan

My main results come from the heteroskedastic ordered probit estimates of the models for reported and latent LS. I interpret these estimates as a component of the conditional median of the underlying continuous wellbeing variable.[2] I also report the estimates of homoskedastic probit, as well as linear projection estimates. The latter allows me to compare my results with the studies that find the U-shape while treating reported LS as cardinal (see, e.g., Frijters and Beatton (2012); Cheng et al. (2017)).

I proceed in the following steps. First, I show that the relationship between reported LS and age, the same as with the other covariates, in my dataset, is in line with the previous findings of the literature. Second, I show that reducing the support of LS, which guarantees the numerical stability of the estimator, does not change the results, and that the U-shape holds both unconditionally and conditionally on other covariates. Third, I estimate the model for latent LS with two different specifications for age and show that it follows the W-shape instead of the U-shape. I now discuss all these steps in details.

Tables 2 – 4 report the estimates of the model for reported and latent LS with a quadratic specification for age. The tables report the results of linear projection, homoskedastic and heteroskedastic ordered probit models respectively. The results are quantitatively similar for all three models. Column 1 contains the model with the full support of reported LS and a list of controls, that are widely

[2]This approach guarantees that my findings are free from the recent criticism of Bond and Lang (2019), who showed that the signs of the parameters of interests in ordered choice models that focus on average LS can be arbitrarily reverted unless knife-edge assumptions hold. Focusing on the median and explicitly modelling heteroskedasticity allows up to avoid this criticism. I refer the reader to Chen et al. (2019) for a detailed account of the approach.

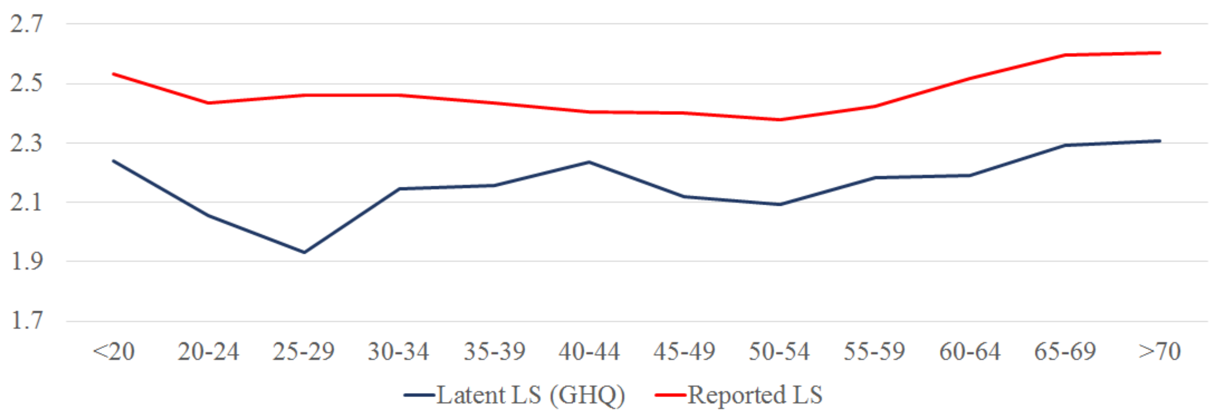

Figure 1: Plots of sample averages of reported and latent LS across age bands

considered to be the sources of LS. In addition to controlling for age (*age*) and age squared (*age2*), it includes university degree dummy (*degree*), gender (*fem*), long-standing illness or disability dummy (*illness*), marital status dummy (*mrd*), logarithm of gross personal income (*l_inc*), and unemployment dummy (*unempl*). Coefficients for linear and quadratic age terms support the U-shape. The effects of other controls are also in line with the literature. Women report being more satisfied with their lives than men. Long-standing illness and unemployment are associated with lower levels of LS, and married people are happier than singles. Reducing the support only influences the magnitude of the coefficients, while the U-shape relationship between age and happiness stay intact, the results are presented in Column 2.

I now eliminate the controls in the model and report the unconditional estimates of the age effects in Columns 3 and 4. The model in Column 3 uses the respondent's age, while the model in Column 4 uses the midpoint of the age group, where the respondent was allocated. Both approaches yield similar results. Unconditionally, the U-shape is less pronounced, which is a typical finding in the literature (see e.g. Frijters and Beatton (2012)), however, the coefficients for both terms are strongly statistically significant.

I estimate the model for latent LS and report the results in Column 5. The estimates do not support the U-shape: both coefficients keep the signs, however, point estimates are smaller in absolute value than for the reported counterpart, and neither of the coefficients is statistically significant. To further investigate the relations, I remove the parametric assumption on the dynamics of LS over the lifespan and allow LS to vary with age arbitrarily, by including a dummy for each age group.

I now present the results from the models that use age band dummies as covariates. Before reporting ordered probit estimates, I provide the estimates from a model where LS is treated as a cardinal variable that focuses on the mean rather than the median. Figure 1 plots the averages of reported (blue line) and latent (red line) LS across age bands. I find average latent LS to be lower than the reported one. A U-shaped pattern does not appear to be a natural fit. Although I see higher averages of latent LS in the youngest and oldest age groups, unlike the plot for reported LS,

	Reported LS	Reported LS	Reported LS	Reported LS	Latent LS
	Full support	Reduced support	Reduced support	Reduced support	Reduced support
age	-0.0540***	-0.0203***	-0.0125***		
	(0.0026)	(0.0011)	(0.0009)		
age2	0.000624***	0.000237***	0.000151***		
	(0.0000252)	(0.0000110)	(0.0000097)		
age_gr_avrg				-0.0126***	-0.0038
				(0.00092)	(0.00549)
age_gr_avrg2				0.000149***	0.000071
				(0.0000092)	(0.0000542)
degree	0.213***	0.105***			
	(0.0182)	(0.0080)			
fem	0.0317**	0.0166**			
	(0.0154)	(0.0067)			
illness	-0.489***	-0.195***			
	(0.0166)	(0.0073)			
mrd	0.290***	0.124***			
	(0.0166)	(0.0073)			
l_inc	0.0148**	0.00966***			
	(0.0068)	(0.0030)			
unempl	-0.548***	-0.215***			
	(0.0372)	(0.0163)			
cons	5.955***	2.738***	2.677***	2.683***	2.164***
	(0.0638)	(0.0279)	(0.0213)	(0.0211)	(0.1330)

Standard errors in parentheses: * p<0.1, ** p<0.05, *** p<0.01

Table 2: Linear projection estimates

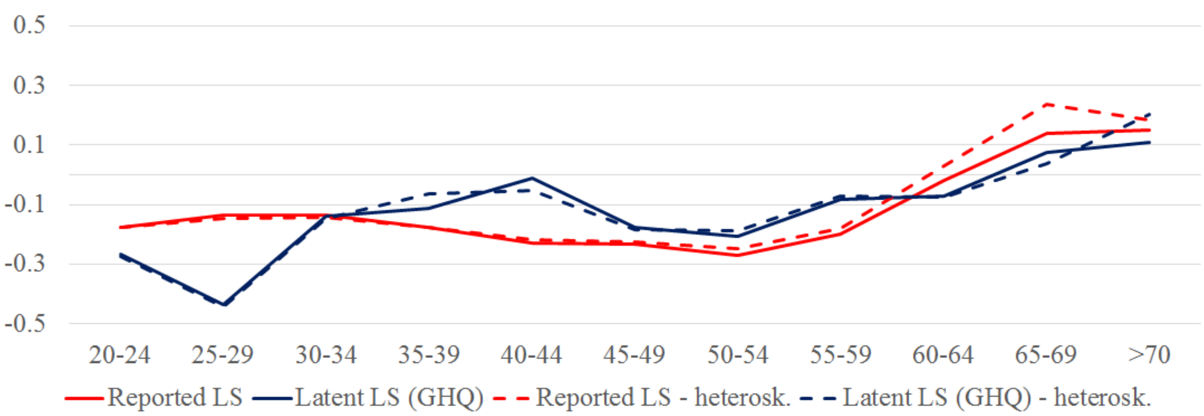

Figure 2: Plots of ordered probit estimates for reported and latent life satisfaction across age bands

	Reported LS	Reported LS	Reported LS	Reported LS	Latent LS
	Full support	Reduced support	Reduced support	Reduced support	Reduced support
age	-0.0693***	-0.0314***	-0.0200***		
	(0.0030)	(0.0017)	(0.0015)		
age2	0.000806***	0.000371***	0.000244***		
	(0.0000299)	(0.0000171)	(0.0000151)		
age_gr_avrg				-0.0202***	-0.0062
				(0.00143)	(0.00825)
age_gr_avrg2				0.000241***	0.000114
				(0.0000143)	(0.0000819)
degree	0.202***	0.163***			
	(0.0215)	(0.0124)			
fem	0.0542***	0.0270 ***			
	(0.0181)	(0.0103)			
illness	-0.587***	-0.300***			
	(0.0196)	(0.0111)			
mrd	0.354***	0.191***			
	(0.0194)	(0.0110)			
l_inc	0.004	0.015***			
	(0.0081)	(0.0046)			
unempl	-0.542***	-0.287***			
	(0.0431)	(0.0237)			

Standard errors in parentheses: * p<0.1, ** p<0.05, *** p<0.01

Table 3: Homoskedastic ordered probit estimates

	Reported LS	Reported LS	Reported LS	Reported LS	Latent LS
	Full support	Reduced support	Reduced support	Reduced support	Reduced support
age	-0.0787***	-0.0439***	-0.0278***		
	(0.0041)	(0.0028)	(0.0017)		
age2	0.000909***	0.000525***	0.000346***		
	(0.0000442)	(0.0000331)	(0.0000192)		
age_gr_avrg				-0.0255***	-0.0080
				(0.00157)	(0.00957)
age_gr_avrg2				0.000315***	0.000141
				(0.0000177)	(0.0000910)
degree	0.176***	0.143***			
	(0.0212)	(0.0166)			
fem	0.0683***	0.0490***			
	(0.0181)	(0.0120)			
illness	-0.586***	-0.359***			
	(0.0292)	(0.0258)			
mrd	0.373***	0.249***			
	(0.0233)	(0.0195)			
l_inc	0.017*	0.021***			
	(0.0089)	(0.0049)			
unempl	-0.513***	-0.266***			
	(0.0497)	(0.0274)			

Standard errors in parentheses: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 4: Heteroskedastic ordered probit estimates

	Homoskedastic ordered probit estimates		Heteroskedastic ordered probit estimates	
	Reported LS	Latent LS	Reported LS	Latent LS
20-24	-0.150***	-0.275	-0.175***	-0.274
	(0.0287)	(0.1953)	(0.0318)	(0.1974)
25-29	-0.114***	-0.450	-0.147***	-0.444
	(0.0291)	(0.2452)	(0.0338)	(0.3151)
30-34	-0.114***	-0.143	-0.142***	-0.146
	(0.0237)	(0.1793)	(0.0312)	(0.1816)
35-39	-0.150***	-0.117	-0.177***	-0.065
	(0.0270)	(0.1730)	(0.0323)	(0.1762)
40-44	-0.194***	-0.012	-0.219***	-0.055
	(0.0269)	(0.1375)	(0.0318)	(0.1425)
45-49	-0.199***	-0.182	-0.227***	-0.184
	(0.0264)	(0.1380)	(0.0322)	(0.1358)
50-54	-0.231***	-0.215	-0.247***	-0.188
	(0.0278)	(0.1564)	(0.0331)	(0.1597)
55-59	-0.169***	-0.085	-0.182***	-0.073
	(0.0309)	(0.1402)	(0.0341)	(0.1517)
60-64	-0.018	-0.074	0.031	-0.076
	(0.0299)	(0.1393)	(0.0365)	(0.1751)
65-69	0.118***	0.076	0.237***	0.038
	(0.0285)	(0.1608)	(0.0383)	(0.1972)
$\geq$70	0.126***	0.114	0.183***	0.201
	(0.0246)	(0.1505)	(0.0321)	(0.1613)
const	1.228***	0.853***	1.238***	0.848***
	(0.0285)	(0.1608)	(0.0267)	(0.1305)

Table 5: Ordered probit estimates of the coefficients for the age band dummies

the 50-55 age band is not the nadir. The lowest levels of latent LS is found for the 25-30 age band. Latent LS follows a W-shape.

Figure 2 contains four plots of estimates of the coefficients from homoskedastic ordered probit models (solid lines) and heteroskedastic ordered probit models (dashed lines) of latent and reported LS. The first thing to note is that the results from homoskedastic and heteroskedastic models are qualitatively similar, whether I look at reported or latent LS. What is more, the probit model results for latent LS are qualitatively similar to simple averages and are different from the reported LS estimates.

Respective probit estimates are presented in Table 5. As expected, standard errors for estimates from the heteroskedastic models are larger than the homoskedastic ones as I have to estimate the skedastic function (nonparametrically). The standard errors for estimating the latent model of LS are even higher because I have to first estimate the nonparametric distribution of all the variables. This, in fact, leaves all age effects statistically insignificant in the heteroskedastic model with latent

LS. Nevertheless, for descriptive analysis, the differences between reported and latent LS appear to be systematic. This can be seen in the striking similarities between the shape of Figure 1 and Figure 2. Latent LS does not follow a U-shape, which explains why the coefficients in the model with a linear and a square terms, i.e. Column 5 of Tables 2 – 4 are not statistically significant.

4.3 Analysis of reporting behaviour

In order to better understand the differences between reported and latent LS, I focus on the distributions of reported and latent LS, as well as on the (mis-)reporting probabilities. Table 6 contains the estimates of the distribution of reported and latent LS for the 25-30 age group. This group exhibits the largest difference between reported and latent LS and is responsible for the deviation from the U-shape.

I start by examining the matrix of (mis-)reporting probabilities, $\mathbf{M}_{X|X^*}$. The diagonal elements of this matrix hold the probability of reporting the latent state, off-diagonal elements present the probability of reporting other states, different from the latent. Figure 3 illustrates how the distributions of reported and latent LS are linked through the misreporting probabilities. Dark blue vertical columns in the lower row represent the distribution of latent LS, $\mathbf{M}_{X^*}$. The cumulative bar chart shows the reporting behaviour of the respondents in each latent state, $\mathbf{M}_{X|X^*}$, where the lower part of each bar represents the share of reporters of the high state and the upper part of the bar – the lower state. The reports are accumulated across the respondents of different latent states in the distribution of reported LS, $\mathbf{M}_X$, represented by the horizontal bars of matching colours. The respondents between age 25 and 30 are more likely to report the latent state if it is high. Respondents who are latently in the low or medium state prefer to report being more satisfied than they really are. While I find the same pattern across age groups, it is particularly pronounced in the 25-30 group. As a result, I observe the nadir in latent LS at this age. As a robustness check, I use different constructs of auxiliary variables, particularly of the GHQ, to identify the misclassification model and re-perform the above analysis. The results support the nadir at the late 20s, see Appendix A.

$$\mathbf{M}_X = \begin{array}{c} \begin{bmatrix} 0.0857 & 0.3687 & 0.5456 \\ (0.0052) & (0.0072) & (0.0089) \end{bmatrix} \\ \begin{matrix} f_X(1) \quad & \quad f_X(2) \quad & \quad f_X(3) \end{matrix} \end{array}$$

$$\mathbf{M}_{X|X^*} = \begin{array}{cc} \begin{bmatrix} 0.1702 & 0.0296 & 0.0561 \\ (0.0407) & (0.0788) & (0.0173) \\ 0.6070 & 0.3200 & 0.1389 \\ (0.0474) & (0.0906) & (0.0410) \\ 0.2228 & 0.6504 & 0.8049 \\ (0.0550) & (0.0996) & (0.0304) \end{bmatrix} & \begin{matrix} f_{X|X^*}(1|\cdot) \\ \\ f_{X|X^*}(2|\cdot) \\ \\ f_{X|X^*}(3|\cdot) \\ \\ \end{matrix} \\ \begin{matrix} f_{X|X^*}(\cdot|1) & f_{X|X^*}(\cdot|2) & f_{X|X^*}(\cdot|3) \end{matrix} & \end{array}$$

$$\mathbf{M}_{X^*} = \begin{array}{c} \begin{bmatrix} 0.3462 & 0.3742 & 0.2796 \\ (0.0598) & (0.0863) & (0.1039) \end{bmatrix} \\ \begin{matrix} f_{X^*}(1) \quad & \quad f_{X^*}(2) \quad & \quad f_{X^*}(3) \end{matrix} \end{array}$$

Table 6: Distribution of reported and latent LS for 25-30 age band

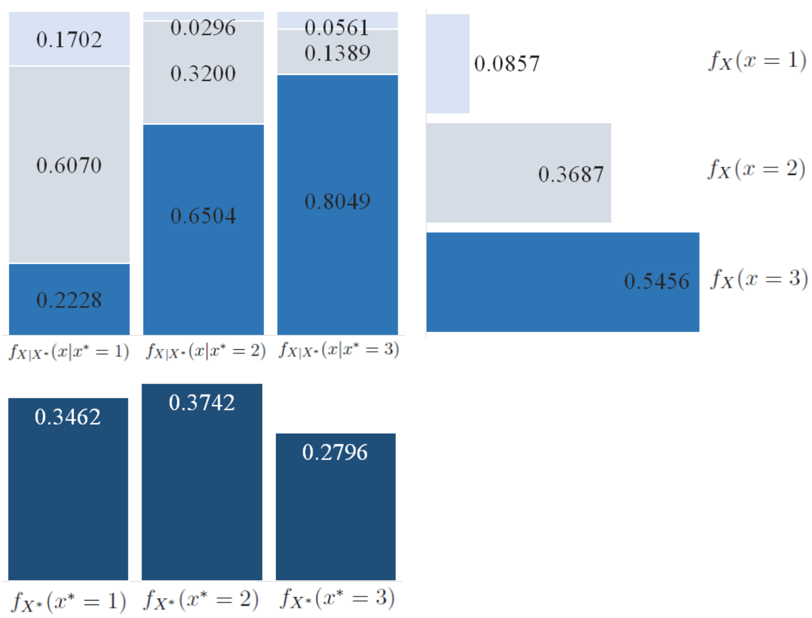

Figure 3: Distribution of reported and latent LS for 25-30 age band

My analysis is purely statistical and does not offer behavioural reasons for the nadir in the late 20s. However, I can hypothesise about its causes based on the insights from the psychology literature. The low level of LS for respondents in late 20s is known in the literature and sometimes labelled *quarter-life crisis* (see, e.g., Robbins and Wilner (2001), Robinson et al. (2013), Byock (2015) and references therein). It is thought to occur due to a mix of factors, including insecurity or disappointment about one's current and future life that occurs after graduation or leaving the family home (Byock (2015)), or as a result of the commitments of working or personal life that has been made before but is no longer desired (Robinson et al. (2013)). The nadir I observe in the late 20s can also be a result of a cohort effect. I use the data collected in 2011-2013, hence the respondents at the age of 25-30 in my dataset are those who graduated and started jobs at the time or after the global financial crisis. As a result, those respondents can be the most heavily affected by the adverse economic environment.

Future research is needed to distinguish between age and cohort effect in my findings, which can be done by exploring the panel dimension of the dataset at hand. An obvious obstacle is that although Understanding Society is collected as a panel, neuroticism levels, that are essential for identification, are not collected in other waves. A potential solution is to proceed under an assumption that individual neuroticism level stays stable over time. This assumption has a degree of support in the psychology literature (see Caspi et al. (2005) for review). Using neuroticism levels measured in the Wave 3 one can proceed in two ways. The first approach includes extending the identification strategy by Oparina and Srisuma (2021), including the application of the misclassification model by Hu (2008), to the panel context, which in itself is an interesting avenue for future research. Alternatively, one can take a more qualitative approach, that does not require additional methodological extensions. Namely, by applying the existing cross-sectional approach to the later waves one can explore if/how the age of nadir changes over time. Whether the age of nadir stays stable in later waves or moves later in life would allow distinguishing between the effect of age and the cohort effect.

5 Conclusion

There is a consensus in the economics of wellbeing literature that reported LS is U-shaped throughout one's life. It is now standard in parametric models to include a linear and a quadratic term for age to account for its effect. However, it is also known that answering wellbeing questions requires significant cognitive effort, that respondents may be willing to minimise. In this paper, I use novel nonparametric techniques to investigate the relationship between age and latent LS, i.e. the state a respondent would have reported if they did not use mental shortcuts. I find that latent LS follows a W-shape: there is some evidence of a dip in LS in the mid-age, but it does not represent the nadir as the literature often suggests. I find the nadir in the late 20s.

This finding raises two further questions: (i) why the respondents in their late 20s are unsatisfied with their lives and (ii) why they do not report so. The answer to the first question is likely to be a combination of age and cohort effects. Given the timing of the survey, respondents in their late 20s are those who graduated and started their career during and after the global financial crisis. The young professionals were among those who were hit the hardest, both in terms of uncertainly and in terms of unemployment. These adverse economic circumstances might have enhanced the dip in LS which is sometimes referred to as the *quarter-life crisis*. Further analysis is needed to disentangle the time and cohort effects in my findings, which can be done, under mild assumptions, by exploring the panel nature of the data. This potential extension is a natural avenue for further research that would allow me to further understand the relations between age and LS.

Understanding the reasons for misreporting is even less trivial. A potential explanation is an effect of stereotypes and social norms, that respondents use as a *mental shortcut* (see e.g. Holtgraves (2004)). Being less satisfied with one's life at mid-age is largely normalised in the Western culture, due to the wide use of the *mid-life crisis* concept in popular culture. Showing signs of distress while being in the late 20s, on the contrary, is still considered less acceptable, and is sometimes characterised as a *snowflake* behaviour.

My results have important implications, both in terms of econometric analysis of LS data and from the wellbeing perspective. From the econometric analysis perspective, I show that respondents of different age misreport in systematically different ways, i.e. misclassification in reported LS is correlated with a regressor. As a result, the error significantly distorts the relative levels. If we think of happiness measures as a mean of policy evaluation, such influence of measurement error makes it difficult to compare effects across groups or to access any aggregate effects if the groups include people of different ages.

From the wellbeing perspective, I find that under my choice of instruments there is a cohort of respondents, those who were in their late 20s in 2011-2013, who experience low wellbeing and who are not willing to report it. According to the reported consumption of antidepressants (Blanchflower and Oswald (2016)) this group of respondents is also less likely to seek medical help or to report receiving it. Given poor mental health can have long-lasting effects on an individuals life, including earnings, educational success, employment and physical health, more research is needed to rationalise these findings and to mitigate potential adverse consequences.

References

Bertrand, M. and Mullainathan, S. (2001). Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review*, 91(2):67–72.

Blanchflower, D. and Graham, C. (2020). The Mid-Life Dip in Well-Being: Economists (Who Find It) Versus Psychologists (Who Don't)! Technical Report w26888, National Bureau of Economic Research, Cambridge, MA.

Blanchflower, D. G. and Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science & Medicine*, 66(8):1733–1749.

Blanchflower, D. G. and Oswald, A. J. (2016). Antidepressants and age: A new form of evidence for U-shaped well-being through life. *Journal of Economic Behavior & Organization*, 127:46–58.

Bond, T. N. and Lang, K. (2019). The Sad Truth about Happiness Scales. *Journal of Political Economy*, 127(4):1629–1640. Publisher: The University of Chicago Press.

Byock, S. D. (2015). The Inner World of the First Half of Life: Analytical Psychology's Forgotten Developmental Stage. *Psychological Perspectives*, 58(4):399–415.

Caspi, A., Roberts, B. W., and Shiner, R. L. (2005). Personality Development: Stability and Change. *Annual Review of Psychology*, 56(1):453–484.

Chen, L. Y., Oparina, E., Powdthavee, N., and Srisuma, S. (2019). Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales. *Working paper.*

Cheng, T. C., Powdthavee, N., and Oswald, A. J. (2017). Longitudinal Evidence for a Midlife Nadir in Human Well-being: Results from Four Data Sets. *The Economic Journal*, 127(599):126–142.

Clark, A. E. (2019). Born to Be Mild? Cohort Effects Don't (Fully) Explain Why Well-Being Is U-Shaped in Age. In Rojas, M., editor, *The Economics of Happiness*, pages 387–408. Springer International Publishing, Cham.

Dear, K., Henderson, S., and Korten, A. (2002). Well-being in Australia. *Social Psychiatry and Psychiatric Epidemiology*, 37(11):503–509.

Di Tella, R., MacCulloch, R. J., and Oswald, A. J. (2001). Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness. *American Economic Review*, 91(1):335–341.

Diener, E., Inglehart, R., and Tay, L. (2013). Theory and Validity of Life Satisfaction Scales. *Social Indicators Research*, 112(3):497–527.

Feddersen, J., Metcalfe, R., and Wooden, M. (2016). Subjective wellbeing: why weather matters. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(1):203–228.

Frijters, P. and Beatton, T. (2012). The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behavior & Organization*, 82(2-3):525–542.

Holtgraves, T. (2004). Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding. *Personality and Social Psychology Bulletin*, 30(2):161–172.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1):27–61.

Jaques, E. (1965). Death and the mid-life crisis. *The International Journal of Psycho-Analysis*, 46(4):502–514.

Kruger, A. (1994). The midlife transition: Crisis or chimera? *Psychological Reports*, 75(3, Pt 1):1299–1305. Place: US Publisher: Psychological Reports.

Oparina, E. and Srisuma, S. (2021). Analyzing Subjective Well-Being Data with Misclassification. *Journal of Business & Economic Statistics*. (forthcoming).

Robbins, A. and Wilner, A. (2001). *Quarterlife Crisis: How to Get Your Head Round Life in Your Twenties*. Bloomsbury Publishing plc, London.

Robinson, O. C., Wright, G. R. T., and Smith, J. A. (2013). The Holistic Phase Model of Early Adult Crisis. *Journal of Adult Development*, 20:27–37.

Schwarz, N. and Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45(3):513–523.

Weiss, A., King, J. E., Inoue-Murayama, M., Matsuzawa, T., and Oswald, A. J. (2012). Evidence for a midlife crisis in great apes consistent with the U-shape in human well-being. *Proceedings of the National Academy of Sciences*, 109(49):19949–19952.

Wilhelm, D. (2018). Testing for the Presence of Measurement Error. *CeMMAP Working Paper CWP45/18*.

Winkelmann, L. and Winkelmann, R. (1998). Why are the Unemployed So Unhappy? Evidence from Panel Data. *Economica*, 65(257):1–15.

A Robustness checks

I divide GHQ-12 into two subgroups, one of those includes questions which are relatively less subjective (GHQ1), another one contains questions which are relatively more subjective and closer to the LS question (GHQ2). Despite some differences in the results, the main findings remain unchanged. I further describe the variables that I use:

1. GHQ1
    - Have you recently been able to concentrate on whatever you're doing?
    - Have you recently lost much sleep over worry?
    - Have you recently felt capable of making decisions about things?
    - Have you recently felt you couldn't overcome your difficulties?
    - Have you recently been able to face up to problems?
    - Have you recently been losing confidence in yourself?
2. GHQ2
    - Have you recently felt that you were playing a useful part in things?
    - Have you recently felt constantly under strain?
    - Have you recently been able to enjoy your normal day-to-day activities?
    - Have you recently been feeling unhappy or depressed?
    - Have you recently been thinking of yourself as a worthless person?
    - Have you recently been feeling reasonably happy, all things considered?

The estimates for the effect of age on latent LS, estimated with different instruments, are presented in Table 7 and Table 8.

	Reported LS	Latent LS GHQ	Latent LS GHQ1	Latent LS GHQ2
<20	2.531	2.240	2.121	2.315
	(0.0132)	(0.0853)	(0.0879)	(0.0950)
20-25	2.435	2.054	1.812	2.199
	(0.0123)	(0.0805)	(0.1032)	(0.1469)
25-30	2.460	1.933	1.781	1.985
	(0.0127)	(0.1460)	(0.1741)	(0.1531)
30-35	2.460	2.144	1.954	2.014
	(0.0118)	(0.0852)	(0.1545)	(0.0759)
35-40	2.435	2.159	2.103	2.086
	(0.0115)	(0.0710)	(0.1692)	(0.1230)
40-45	2.405	2.236	2.115	2.227
	(0.0102)	(0.0463)	(0.0586)	(0.0901)
45-50	2.402	2.118	2.096	2.125
	(0.0095)	(0.0540)	(0.0851)	(0.1063)
50-55	2.379	2.093	2.102	2.061
	(0.0113)	(0.0498)	(0.0538)	(0.0962)
55-60	2.422	2.182	2.147	2.141
	(0.0133)	(0.0650)	(0.0821)	(0.1130)
60-65	2.518	2.191	2.109	2.113
	(0.0127)	(0.0396)	(0.0683)	(0.0985)
65-70	2.596	2.292	2.060	2.108
	(0.0118)	(0.0557)	(0.1256)	(0.0749)
>70	2.604	2.308	2.202	2.315
	(0.0076)	(0.0507)	(0.0752)	(0.0739)

Table 7: Estimates of average LS for different age bands

	Homoskedastic ordered probit				Heteroskedastic ordered probit			
	Reported LS	Latent LS GHQ	Latent LS GHQ1	Latent LS GHQ2	Reported LS	Latent LS GHQ	Latent LS GHQ1	Latent LS GHQ2
20-25	-0.150	-0.275	-0.391	-0.164	-0.175	-0.274	-0.392	-0.139
	(0.0287)	(0.1953)	(0.1022)	(0.2473)	(0.0318)	(0.1974)	(0.0872)	(0.2323)
25-30	-0.114	-0.450	-0.434	-0.458	-0.147	-0.444	-0.448	-0.414
	(0.0291)	(0.2452)	(0.2775)	(0.2583)	(0.0338)	(0.3151)	(0.3058)	(0.3500)
30-35	-0.114	-0.143	-0.212	-0.416	-0.142	-0.146	-0.225	-0.352
	(0.0237)	(0.1793)	(0.2300)	(0.1835)	(0.0312)	(0.1816)	(0.2218)	(0.1936)
35-40	-0.150	-0.117	-0.023	-0.313	-0.177	-0.065	-0.033	-0.200
	(0.0270)	(0.1730)	(0.2462)	(0.2318)	(0.0323)	(0.1762)	(0.3198)	(0.3511)
40-45	-0.194	-0.012	-0.011	-0.126	-0.219	-0.055	-0.040	-0.108
	(0.0269)	(0.1375)	(0.1281)	(0.1842)	(0.0318)	(0.1425)	(0.1367)	(0.1754)
45-50	-0.199	-0.182	-0.034	-0.266	-0.227	-0.184	-0.054	-0.228
	(0.0264)	(0.1380)	(0.1807)	(0.1990)	(0.0322)	(0.1358)	(0.1989)	(0.2037)
50-55	-0.231	-0.215	-0.023	-0.351	-0.247	-0.188	-0.014	-0.291
	(0.0278)	(0.1564)	(0.1347)	(0.1896)	(0.0331)	(0.1597)	(0.1594)	(0.2062)
55-60	-0.169	-0.085	0.035	-0.243	-0.182	-0.073	0.039	-0.209
	(0.0309)	(0.1402)	(0.0950)	(0.2251)	(0.0341)	(0.1517)	(0.1099)	(0.2139)
60-65	-0.018	-0.074	-0.017	-0.281	0.031	-0.076	-0.038	-0.242
	(0.0299)	(0.1393)	(0.1492)	(0.1693)	(0.0365)	(0.1751)	(0.1686)	(0.1993)
65-70	0.118	0.076	-0.079	-0.292	0.237	0.038	-0.099	-0.264
	(0.0285)	(0.1608)	(0.2340)	(0.1712)	(0.0383)	(0.1972)	(0.3434)	(0.1657)
>70	0.126	0.114	0.106	0.005	0.183	0.201	0.114	0.034
	(0.0246)	(0.1505)	(0.1681)	(0.1808)	(0.0321)	(0.1613)	(0.1889)	(0.1717)
const	1.228	0.853	0.654	0.936	1.238	0.848	0.669	0.885
	(0.0285)	(0.1608)	(0.1341)	(0.1375)	(0.0267)	(0.1305)	(0.1286)	(0.1369)

Table 8: Ordered probit estimates of the coefficients for the age band dummies

B Distribution of reported and latent LS for age bands

<20

$$\mathbf{M}_X = \begin{bmatrix} 0.0769 & 0.3150 & 0.6081 \\ (0.0049) & (0.0093) & (0.0104) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1363 & 0.0564 & 0.0667 \\ (0.0266) & (0.0112) & (0.0094) \\ 0.6692 & 0.3790 & 0.1013 \\ (0.0662) & (0.0411) & (0.0196) \\ 0.1945 & 0.5646 & 0.8320 \\ (0.0841) & (0.0448) & (0.0203) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2004 & 0.3595 & 0.4401 \\ (0.0563) & (0.0714) & (0.0549) \end{bmatrix}$$

20-25

$$\mathbf{M}_X = \begin{bmatrix} 0.0955 & 0.3736 & 0.5309 \\ (0.0060) & (0.0085) & (0.0087) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1679 & 0.0603 & 0.0781 \\ (0.0259) & (0.0153) & (0.0107) \\ 0.6354 & 0.3920 & 0.1324 \\ (0.0318) & (0.0340) & (0.0248) \\ 0.1967 & 0.5477 & 0.7895 \\ (0.0477) & (0.0384) & (0.0246) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2735 & 0.3991 & 0.3274 \\ (0.0478) & (0.0604) & (0.0527) \end{bmatrix}$$

25-30

$$\mathbf{M}_X = \begin{bmatrix} 0.0857 & 0.3687 & 0.5456 \\ (0.0052) & (0.0072) & (0.0089) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1702 & 0.0296 & 0.0561 \\ (0.0407) & (0.0788) & (0.0173) \\ 0.6070 & 0.3200 & 0.1389 \\ (0.0474) & (0.0906) & (0.0410) \\ 0.2228 & 0.6504 & 0.8049 \\ (0.0550) & (0.0996) & (0.0304) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.3462 & 0.3742 & 0.2796 \\ (0.0598) & (0.0863) & (0.1039) \end{bmatrix}$$

30-35

$$\mathbf{M}_X = \begin{bmatrix} 0.0875 & 0.3650 & 0.5475 \\ (0.0046) & (0.0078) & (0.0089) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1700 & 0.0508 & 0.0735 \\ (0.0295) & (0.0125) & (0.0102) \\ 0.6741 & 0.4120 & 0.1256 \\ (0.0322) & (0.0458) & (0.0233) \\ 0.1559 & 0.5371 & 0.8009 \\ (0.0535) & (0.0509) & (0.0213) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2352 & 0.3855 & 0.3793 \\ (0.0497) & (0.0576) & (0.0531) \end{bmatrix}$$

35-40

$$\mathbf{M}_X = \begin{bmatrix} 0.0947 & 0.3752 & 0.5301 \\ (0.0052) & (0.0076) & (0.0083) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1542 & 0.0618 & 0.0790 \\ (0.0190) & (0.0108) & (0.0085) \\ 0.6596 & 0.4150 & 0.1681 \\ (0.0275) & (0.0455) & (0.0141) \\ 0.1861 & 0.5233 & 0.7530 \\ (0.0379) & (0.0477) & (0.0146) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2757 & 0.2900 & 0.4343 \\ (0.0393) & (0.0467) & (0.0455) \end{bmatrix}$$

40-45

$$\mathbf{M}_X = \begin{bmatrix} 0.1040 & 0.3866 & 0.5094 \\ (0.0048) & (0.0077) & (0.0077) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.2589 & 0.0758 & 0.0719 \\ (0.0320) & (0.0093) & (0.0061) \\ 0.7141 & 0.4715 & 0.1597 \\ (0.0289) & (0.0269) & (0.0132) \\ 0.0270 & 0.4528 & 0.7683 \\ (0.0473) & (0.0288) & (0.0139) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.1625 & 0.4388 & 0.3987 \\ (0.0352) & (0.0402) & (0.0253) \end{bmatrix}$$

45-50

$$\mathbf{M}_X = \begin{bmatrix} 0.1032 & 0.3911 & 0.5057 \\ (0.0043) & (0.0071) & (0.0072) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.2034 & 0.0629 & 0.0790 \\ (0.0190) & (0.0078) & (0.0104) \\ 0.6676 & 0.4296 & 0.1628 \\ (0.0211) & (0.0301) & (0.0186) \\ 0.1290 & 0.5075 & 0.7583 \\ (0.0305) & (0.0274) & (0.0190) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2455 & 0.3911 & 0.3635 \\ (0.0248) & (0.0377) & (0.0394) \end{bmatrix}$$

50-55

$$\mathbf{M}_X = \begin{bmatrix} 0.1166 & 0.3877 & 0.4957 \\ (0.0057) & (0.0078) & (0.0078) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.2071 & 0.0836 & 0.0732 \\ (0.0209) & (0.0105) & (0.0092) \\ 0.6385 & 0.4665 & 0.1353 \\ (0.0221) & (0.0325) & (0.0157) \\ 0.1545 & 0.4499 & 0.7916 \\ (0.0315) & (0.0359) & (0.0184) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.3005 & 0.3056 & 0.3939 \\ (0.0357) & (0.0450) & (0.0313) \end{bmatrix}$$

55-60

$$\mathbf{M}_X = \begin{bmatrix} 0.1047 & 0.3687 & 0.5266 \\ (0.0062) & (0.0086) & (0.0093) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.2006 & 0.0690 & 0.0806 \\ (0.0223) & (0.0153) & (0.0108) \\ 0.6730 & 0.4596 & 0.1212 \\ (0.0275) & (0.0405) & (0.0194) \\ 0.1264 & 0.4714 & 0.7982 \\ (0.0371) & (0.0474) & (0.0188) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2347 & 0.3486 & 0.4167 \\ (0.0374) & (0.0482) & (0.0433) \end{bmatrix}$$

60-65

$$\mathbf{M}_X = \begin{bmatrix} 0.0928 & 0.2966 & 0.6107 \\ (0.0050) & (0.0083) & (0.0094) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1562 & 0.0772 & 0.0732 \\ (0.0199) & (0.0177) & (0.0084) \\ 0.6178 & 0.3365 & 0.0888 \\ (0.0771) & (0.0840) & (0.0115) \\ 0.2260 & 0.5863 & 0.8380 \\ (0.0787) & (0.0881) & (0.0123) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2182 & 0.3728 & 0.4090 \\ (0.0288) & (0.0555) & (0.0387) \end{bmatrix}$$

65-70

$$\mathbf{M}_X = \begin{bmatrix} 0.0790 & 0.2455 & 0.6755 \\ (0.0051) & (0.0074) & (0.0084) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1295 & 0.0705 & 0.0689 \\ (0.0230) & (0.0100) & (0.0081) \\ 0.6450 & 0.2710 & 0.0842 \\ (0.0532) & (0.0359) & (0.0124) \\ 0.2255 & 0.6585 & 0.8469 \\ (0.0575) & (0.0327) & (0.0151) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.1556 & 0.3965 & 0.4479 \\ (0.0264) & (0.0488) & (0.0452) \end{bmatrix}$$

>70

$$\mathbf{M}_X = \begin{bmatrix} 0.0676 & 0.2613 & 0.6711 \\ (0.0033) & (0.0054) & (0.0057) \end{bmatrix}$$

$$\mathbf{M}_{X|X^*} = \begin{bmatrix} 0.1093 & 0.0531 & 0.0587 \\ (0.0169) & (0.0079) & (0.0062) \\ 0.5852 & 0.3235 & 0.0974 \\ (0.0461) & (0.0309) & (0.0096) \\ 0.3056 & 0.6235 & 0.8438 \\ (0.0562) & (0.0318) & (0.0100) \end{bmatrix}$$

$$\mathbf{M}_{X^*} = \begin{bmatrix} 0.2072 & 0.2775 & 0.5153 \\ (0.0313) & (0.0434) & (0.0354) \end{bmatrix}$$

Table 9: Distribution of reported and latent LS for different age bands