Journal of Economic Psychology 49 (2015) 74–83

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

A Samuelsonian validation test for happiness data ☆

Ricardo Perez-Truglia *

Microsoft Research, United States

ARTICLE INFO

Article history:
Received 16 November 2014
Received in revised form 19 March 2015
Accepted 6 May 2015
Available online 15 May 2015

JEL classification:
B41
D12
D60

PsycINFO classification:
2900
2910

Keywords:
Revealed-preference
Subjective well-being
Life satisfaction

ABSTRACT

There is growing interest in the use of subjective well-being data, such as survey questions about happiness and life satisfaction. The existing validation tests determine whether these subjective measures have a positive correlation with objective measures of well-being, such as suicide rates and frequency of smiling. We propose an alternative test consisting of three steps: using regression analysis to infer preferences from subjective well-being data; using those estimated parameters to predict how a rational utility-maximizer individual should have acted; and comparing predicted behavior with actual behavior. This validation test can be compelling for economists, because it compares decision utility (i.e., preferences inferred from behavior) to reported utility (i.e., preferences inferred from self-reported well-being). We provide an application of this test based on a model of food consumption, estimated with one of the most widely used measures of subjective well-being: life satisfaction. We find that, across individuals, a one percentage point increase in the actual expenditure share is associated with a 0.76 (SE 0.196) increase in the share predicted by life satisfaction. Additionally, life satisfaction performs significantly better than other objective and subjective measures of well-being (e.g., income, satisfaction with income). The evidence suggests that life satisfaction offers some useful information about experienced utility.

1. Introduction

There are two possible meanings of the term utility. On the one hand, Samuelson's decision utility refers to preferences as revealed by observed choices (Samuelson, 1948). On the other hand, Bentham's experienced utility refers to the actual feelings of pleasure and pain that an individual experiences in response to certain stimuli (Kahneman, Wakker, & Sarin, 1997). Samuelson's revealed preference principle is arguably the most important and influential concept in economics. However, this decision utility may sometimes be difficult to estimate or less reliable. For example, economic analysis often relies on the assumption that decision utility reflects experienced utility, but this assumption can be violated by bounded rationality, endogenous preferences, or problems of self-control, among others (Chetty, 2015; Kahneman et al., 1997).[1]

☆ Thanks to Rafael Di Tella and Sendhil Mullainathan for very valuable discussions. Thanks to two anonymous reviewers for their great feedback that substantially improved the paper. Thanks to Daniel Benjamin, Nicolas Bottan, Jan-Emmanuel De Neve, Martin Tetaz and seminar participants at the London School of Economics for useful comments. The author would like to acknowledge support from the Simon Kuznets Research Grant from the Department of Economics at Harvard University. The usual disclaimer applies.

* Address: Microsoft New England Research and Development (NERD) Lab, Office 12073, 1 Memorial Drive, Cambridge, MA 02142, United States.
E-mail address: rtruglia@microsoft.com

[1] In some other cases, decision utility may be useful in itself for predicting behavior. Additionally, for some other applications, even though the presence of issues like endogenous preferences and self-control may play a role, decision utility may be deemed a good enough approximation.

http://dx.doi.org/10.1016/j.joep.2015.05.002
0167-4870/

An alternative measure of experienced utility would be valuable for situations where decision utility is less reliable, too costly, or impossible to estimate. Economists and other social scientists have long sought a direct measurement of experienced utility. For example, Edgeworth (1881) fantasized about a hedonimeter, an instrument capable of measuring pleasure in real time. Neuroscience may someday provide this tool, but for now we rely on other sources of data. Among social scientists, there has been exponential growth in the use of subjective well-being data. For example, some studies use data on responses to questions like, "Are you very happy, pretty happy, or not too happy?" This literature, sometimes referred to as happiness economics, has grown rapidly over the last few decades (for a review, see Di Tella & MacCulloch (2006)). This growth was facilitated by the availability of subjective well-being data in countless surveys around the world and spanning several decades. These studies often compare the well-being of individuals or nations (e.g., Easterlin, 1974; Kahneman & Deaton, 2010; Ludwig et al., 2012; Stevenson & Wolfers, 2008, 2009). However, a significant number of these studies use subjective well-being data to test theories about human behavior (e.g., Easterlin, 1995; Gruber & Mullainathan, 2005; Luttmer, 2005; Oswald & Powdthavee, 2008; Perez-Truglia, 2013; Senik, 2004). In spite of the increasing use of subjective well-being data, a majority of economists remain skeptical, although this skepticism may be about subjective data in general and not just well-being data (Bertrand & Mullainathan, 2001).

Several studies provide validation tests for subjective well-being measures. These tests typically consist of gauging whether a subjective measure of well-being is positively correlated to objective measures of well-being, such as emotional expressions (e.g., Sandvik, Diener, & Seidlitz, 1993), neurological measures (e.g., Urry et al., 2004), third-party evaluations (e.g., Sandvik et al., 1993), aggregate suicide rates (e.g., Di Tella, MacCulloch, & Oswald, 2003), among others (for a recent review, see Diener, Inglehart, & Tay (2013)). Most economists seem unpersuaded by this evidence, because it does not discard the possibility that these subjective scores measure superficial aspects of life, like mood. For instance, suppose that reporting to like the Rolling Stones is positively correlated to the frequency of smiling. This correlation is not sufficient evidence to support a Rolling Stones Index to guide economic policy or test economic theories. Nevertheless, psychologists have devoted a large literature to investigating these subjective scores measures. They have identified several measurement challenges, such as sensitivity to contextual factors and selective memory, and developed tools to address them (e.g., Stone, Shiman, & DeVries, 1999).[2] However, methodological differences make such insights less compelling for the average economist. Thus, this paper proposes a framework for testing the validity of a subjective well-being measure that should be more compelling for economists.

Consider a context in which decision utility is believed to provide an accurate approximation of experienced utility. If, in that context, reported utility is similar to decision utility, then that would provide a compelling argument for using reported utility in cases where decision utility is absent or less reliable. Based on this insight, we propose a test consisting of three steps: using regression analysis to infer preferences from subjective well-being data; using those estimated parameters to predict how a rational utility-maximizer individual should behave; and comparing the predicted behavior with the actual behavior.

In addition to formulating a general version of the test, we provide a simple application based on a model of household consumption. To illustrate the intuition behind this application, assume that we have an hedonimeter that is capable of measuring well-being, H_i. After experimenting with different combinations of consumptions of two goods, X_1 and X_2, we find the following relationship in the data: $H_i \approx \hat{\mu}_i^1 \cdot \ln(X_1) + \hat{\mu}_i^2 \cdot \ln(X_2)$. If this individual rationally maximized H_i, then she should choose an expenditure share equal to $\hat{s}_i^1 = \frac{\hat{\mu}_i^1}{\hat{\mu}_i^1 + \hat{\mu}_i^2}$. To test the validity of the hedonimeter, we should measure the expenditure share that this individual chooses in reality, and compare that share to the prediction of the hedonimeter.

We estimate a model of food consumption using panel data on subjective well-being, expenditures, and prices from a Russian household panel. The life satisfaction data is used to estimate consumption preferences in two categories of food: animal source and non-animal source. We assume that preferences are heterogeneous across individuals but stable over time. Thus, we can exploit the panel structure of the data to estimate individual-specific preference parameters. Ideally, we would test whether the actual and predicted choices are equal for every individual. However, because of data limitations, our application lacks enough power to perform an ideal version of the test. Instead, we perform a weaker version of the test by measuring the cross-individual association between actual and predicted expenditure choices.

We find that, across individuals, a one percentage point increase in the actual expenditure share is associated with an increase of 0.76 (SE 0.196) percentage points in the expenditure share predicted by life satisfaction. Indeed, we cannot reject the hypothesis of a regression coefficient of 1, which corresponds to the ideal value in the context of this test. However, given the imprecise estimation of this parameter, we cannot reject the hypothesis that the coefficient equals 0.5 either. Additionally, we show that life satisfaction performs significantly better in this test relative to other objective and subjective measures of well-being (e.g., income, satisfaction with income, perception of economic ranking, and health self-evaluation). The evidence suggests that life satisfaction offers useful information about experienced utility. However, it is unknown whether life satisfaction would perform well in the strong version of the test.

To the best of our knowledge, only a few papers provide validation tests that are similar to our test. First, Oswald and Wu (2010) use objective data to compare state-level averages of happiness and a quality-of-life measure inferred from residential choices. They report a statistically and economically significant correlation coefficient of 0.6 between the subjective and

[2] See Diener et al. (2013) for a recent review. Indeed, there are still important disagreements in some respects – for example, there is a debate as to whether happiness measures more frequency rather than average intensity of pleasurable moments (Diener, Sandvik, & Pavot, 1991).

objective measures of well-being. Although they do not compare predicted behavior to actual behavior, as we propose, their test is implicitly based on the comparison between reported and decision utility.

Second, Benjamin, Heffetz, Kimball, and Rees-Jones (2012) provide a test based on a cleverly designed survey. Subjects are shown pairs of hypothetical tradeoff scenarios (e.g., higher income but longer workdays). Subjects are then asked to report which scenario they would choose and which scenario would make them happier. The evidence suggests that, despite significant discrepancies, predicted choice is largely consistent with predicted well-being: 83% of the time, respondents chose the scenario with the highest anticipated well-being. In a follow-up paper, Benjamin, Heffetz, Kimball, and Rees-Jones (2014) present further evidence from a survey of students from U.S. medical schools shortly after submitting their choice rankings for the National Resident Matching Program. They analyze the students' choice rankings, anticipated subjective well-being rankings, and expectations about the quality of residencies (e.g., prestige). They estimate the marginal rates of substitution inferred from choice rankings and the anticipated subjective well-being rankings. They find that, even though life satisfaction does better than other subjective measures of well-being, the trade-offs inferred from life satisfaction are substantially different from the trade-offs inferred from choice rankings.[3]

Benjamin et al. (2012, 2014) study the validity of hypothetical well-being data (e.g., they ask whether the respondent anticipates more life satisfaction for particular hypothetical situations). However, the vast majority of the subjective well-being literature uses stated well-being data (e.g., whether the respondent feels a little or a lot satisfied with her life at the moment). Although the evidence in Benjamin et al. (2012, 2014) is suggestive, it does not directly address the validity of stated well-being data.[4] Our paper tries to fill this gap.

2. A general validation test for subjective well-being data

Let $c_{i,t}$ denote a vector with the choice variables, where the subscript i indexes individuals and subscript t indexes time. For instance, the elements in $c_{i,t}$ could denote consumption of different goods or work hours. Let $U(c_{i,t};\phi_i)$ denote the utility function, where ϕ_i is a vector of preference parameters. Let $H_{i,t}$ denote a measure of subjective well-being (i.e., reported utility). Consider the following regression equation:

$$H_{i,t} = U(c_{i,t};\phi_i) + \epsilon_{i,t}$$

Given a set of instrumental variables $Z_{i,t}$, i.e., $\mathrm{E}[\mathrm{Z}_{i,t} \cdot \epsilon_{i,t}] = 0$, we can consistently estimate the parameter values, $\hat{\phi}_i$. Let $F(c_{i,t};p_{i,t}) = 0$ denote the individual's feasible choice set. For instance, in a problem of optimal consumption the latter would be the budget constraint and $p_{i,t}$ would contain information about prices and income. Let $\hat{c}_{i,t}$ be the choices predicted by the subjective well-being data:

$$\hat{c}_{i,t} = \arg\max_{c} \mathrm{U}\left(\mathrm{c};\hat{\phi}_{\mathrm{i}}\right) \text{s.t.} \quad F(c;p_{i,t}) = 0$$

To test the consistency of reported utility with decision utility, we must measure how consistent the predicted choices, $\hat{c}_{i,t}$, are with the observed choices, $c_{i,t}$. [5] The attractiveness of this approach is that it does not require to specify or estimate a model for decision utility.[6] To perform this test, we must define a notion of consistency between predicted and observed choices. A strong version of consistency suggests testing the null hypothesis that $\hat{c}_{i,t} = c_{i,t}\ \forall i,t$ (in which case reported utility would be consistent with decision utility). If each $\hat{c}_{i,t}$ was precisely estimated, then the consistency between predicted and observed choices could be assessed as the R-squared in an scatterplot of $\{\hat{c}_{i,t}, c_{i,t}\}$. In other words, we could measure whether predicted choices explain 10%, 50% or 90% of the variation in actual choices.

In practice, the data may not be good enough and then each individual $\hat{c}_{i,t}$ may be very imprecisely estimated. In that case, the strong version of the test would be uninformative. Indeed, this turns out to be the case in the application to food consumption presented in the following sections. In that case, we can use a weak version of the test, based on an alternative definition of consistency between predicted and actual choices. The null hypothesis of the week test is that the coefficient from a regression of $\hat{c}_{i,t}$ on $c_{i,t}$ should be equal to one. Note that it is perfectly plausible for reported utility to satisfy the weak test while at the same time violating the strong test.[7]

[3] Note that, however, they still find a significant prediction rate for subjective well-being, although it is somewhat smaller than that of the Benjamin et al. (2012) study.

[4] Indeed, Benjamin, Heffetz, Kimball, and Rees-Jones (2014) discusses some biases in survey responses (e.g., halo effect, cognitive dissonance) that may suggest that the validity of hypothetical well-being data may serve as an upper bound to the validity of stated well-being data.

[5] For simplicitly, we take observed choices, $c_{i,t}$ and the extra choice parameters, $p_{i,t}$, as given. In practice, due to issues such as measurement error, we may only be able to observe an estimate of such variables (e.g., average consumption shares over a period of time).

[6] Alternatively, we could estimate the preference parameters with a traditional choice model, $\hat{\phi}_i^{choice}$, and compare those estimated parameters directly with the corresponding parameters estimated with life satisfaction data, $\hat{\phi}_i$. However, depending on how complex the choice model is, that could make the test less straightforward. For example, Benjamin, Heffetz, Kimball, and Rees-Jones (2014) compare directly marginal rates of substitution inferred from subjective well-being and choice, which adds an additional layer of complexity.

[7] Assume that $\hat{c}_{i,t}$ is very precisely estimated. In the scatterplot of $\{\hat{c}_{i,t}, c_{i,t}\}$, it is possible that the regression coefficient is exactly one (satisfying the weak test) and still the R-squared can be arbitrarily small (violating the strong test).

We should also note that, even if the data is good enough to pass the strongest version of the test, the results would not imply that reported utility is valid for every purpose. Intuitively, our test assesses the validity of subjective data for estimating marginal rates of substitution. And since marginal rates of substitution are estimated using within-individual variation in reported utility, the fact that the marginal rates of substitution are valid does not imply that the cross-individual variation in reported utility is also valid (e.g., reported utility may still over-estimate the levels of experienced utility of Spanish speakers relative to English speakers).

Last, it is important to note that it would be very unrealistic to expect that a single question about subjective well being could accurately predict choices. There are multiple questions about subjective well-being that measure different combinations of emotional states, and some of these measures of reported well-being may provide a better representation of some marginal rates of substitution (e.g., consumption of different food items) but worse representation of other rates of substitution (e.g., consumption vs. leisure). Indeed, it is straightforward to extend the basic framework to allow for a utility function that depends on a combination of emotional states (Becker & Rayo, 2008). Assume that the experienced utility can be approximated by a linear combination of emotional states:

$$H^*_{i,t} = \sum_{k=1}^{K} \gamma_k \cdot H^k_{i,t}, \sum_{\mathrm{k}=1}^{\mathrm{K}} \gamma_{\mathrm{k}} = 1$$

where $\left\{H^k_{i,t}\right\}_{k=1}^{K}$ is the set of K measures of subjective well-being. We could use this index $H^*_{i,t}$ to estimate jointly preferences and weights (i.e., $\hat{\phi}$ and γ) to best predict choices.[8]

3. An application to consumption choices

We illustrate the approach with a very simple application: the choice of allocate a fixed budget across J goods.[9] Let subscripts i and t denote individuals and time, and superscript $j = 1, \ldots, J$ denote consumption goods. Let $U_{i,t}$ be the Cobb–Douglas utility function over consumption goods.[10] Let c^j_i denote consumption on good j, e^j_i denote expenditure in j and p^j_i denote j's price. We drop the superscript j to denote the vector for all J goods. Using the identity $e^j_i = c^j_i \cdot p^j_i$, the utility function can be expressed as follows:

$$U_{i,t}(e_{i,t}, p_{i,t}; \alpha_i) = \sum_{j=1}^{J} \alpha^j_i \ln\left(e^j_{i,t}\right) - \sum_{j=1}^{J} \alpha^j_i \ln\left(p^j_{i,t}\right)$$

Note that we are assuming that preferences are heterogeneous but stable over time. In that case, we can exploit the panel structure of the data to obtain consistent estimates of the individual-specific parameters, α_i (Swamy, 1970). To do so, we must regress subjective well being on the interaction between the expenditure variables, $\left\{\ln\left(e^j_{i,t}\right)\right\}_{j=1}^{J}$, and the individual dummies. This specification compares – for each individual – changes in expenditures of different goods to the change in well-being. If i's happiness reacts more favorably to the consumption of good j, that will result in a higher $\hat{\alpha}^j_i$.

The advantage of estimating individual-specific preferences is that we can predict expenditures shares for each individual. As a result, we can test not only whether the life satisfaction data predicts the average expenditure choices, but also if it predicts cross-individual differences in expenditure choices. In the case of two goods, this regression yields a pair of estimated marginal utilities for each individual, $\{\hat{\alpha}^1_i, \hat{\alpha}^2_i\}$. If both parameters are positive, then the predicted expenditures share in the first good is $\frac{\hat{\alpha}^1_i}{\hat{\alpha}^1_i + \hat{\alpha}^2_i}$, which is the interior solution to the optimal consumption problem.[11] If one of the marginal utilities is positive and the other is negative, then the prediction is a corner solution: all of the expenditures should go to the consumption category with the positive coefficient. If both marginal utilities are negative, then the prediction is another corner solution: the individual should spend zero in both consumption goods, which would leave the expenditures share indeterminate.

This model makes a number of assumptions that are worth discussing. One important assumption is the exclusion restriction. Indeed, we can think of a number of potential sources of omitted-variable-biases. For instance, not controlling for prices or consumption of non-food items could in principle introduce systematic biases. In the results section, we offer robustness checks aimed at testing some of these potential sources of biases. In any case, we must keep in mind that the predicted

[8] See Benjamin, Heffetz, Kimball, and Rees-Jones (2014) and Benjamin, Heffetz, Kimball, and Szembrot (2014) for a related discussion on the optimal combination of multiple subjective scores. In particular, the results in Benjamin, Heffetz, Kimball, and Szembrot (2014) suggest that measures of well-being that use multiple questions do not seem to explain marginal rates of substitution significantly better than single-question measures.

[9] This is a partial optimization problem: i.e., we are taking as given that the individual wants to spend a budget M over those N goods. That problem is of course part of a more general optimization problem where the individual must also solve trade-offs between other goods outside those N goods, including expenditures at different points in time.

[10] Even though the Cobb–Douglas assumption may seem strong, this functional form happens to fit the data very well.

[11] In multiple-member households, the prediction for the household-level expenditures should depend on the combination of the individual preferences of all household members, weighted by their corresponding bargaining power. Since we do not observe their bargaining powers, this is an additional source of model misspecification.

choice is the ratio between the marginal utility over the two goods. As a result, these omitted-variable-biases will not affect predicted choices if they bias the estimated marginal utilities for the two goods to a similar extent. For example, we are assuming that preferences vary across individuals but are held constant for a given individual over time. If this assumption was violated, that may inflate the estimates of the marginal utilities for both goods in a similar manner, thereby not affecting the predicted choice.[12] Last, an implicit but important identification assumption is that the individuals do not always choose the optimal consumption bundle: otherwise, given the assumption on stable preferences, there would be perfect multi-collinearity between the expenditure variables from the right hand side of the regression equation. These deviations from the optimal bundle can be due to various plausible factors, such as experimentation and informational/choice frictions.

4. Data

We use panel data from rounds 5 to 20 (corresponding to years 1994–2011) of the Russian Longitudinal Monitoring Survey[13] (RLMS).[14] We chose this survey for several reasons. First, it contains one of the most widely used measures of subjective well-being, life satisfaction, as well as some additional measures of subjective well-being.[15] Second, the data contain detailed measures of household expenditures, prices, and income. Finally, the panel structure of this dataset links the same individuals over several years, which is a critical requirement for the estimation of individual-specific preference parameters.

Let J be the number of individuals and let R be the number of variables with individual-specific parameters (including the intercept). In the heterogeneous-preference specification, the number of individual-specific parameters is equal to $R \cdot J$. To limit the number of parameters, we fix $R = 3$ by focusing on just two consumption categories: animal source food (e.g., meats, dairy products, fish, eggs, fat) and non-animal source food (e.g., potatoes, breads, sugar, fruits, vegetables).[16] These measures of expenditures are constructed using responses to a battery of questions about expenditures on dozens of specific items (e.g., milk, potatoes). Additionally, the RLMS contains a community-level survey on prices of specific food items (e.g., milk, potatoes), which we used to construct price indexes for each of the two food categories.[17]

The key measure of subjective well-being is the standard *life satisfaction* question: "To what extent are you satisfied with your life in general at the present time?" The possible answers range from "not at all satisfied" (1) to "fully satisfied" (5). We employ the Probit-OLS (POLS) regression model for this dependent variable (van Praag and Ferrer-i-Carbonell, 2007). This model consists of an OLS regression where the dependent variable does not take the values from 1 to 5 but, instead, some other values that take into account the ordinal nature of the life satisfaction question.[18] We find very similar results when using OLS with the raw scores of 1–5. This evidence is consistent with Ferrer-i-Carbonell and Frijters (2004), who show that using ordinal or cardinal models for the subjective well-being outcomes makes little difference in practice if individual-specific intercepts are included.

The data also include other measures of subjective well-being, which we use to provide a benchmark for life satisfaction. Economic satisfaction is based on the question: "Tell me, please, how satisfied are you with your economic conditions at the present time?" The scale of possible answers is the same as for the life satisfaction question, ranging from "not at all satisfied" (1) to "fully satisfied" (5). Unfortunately, this question was not included in all rounds of the survey, resulting in a slightly smaller sample size. Another measure of subjective well-being, economic rank, is based on the question: "And now another nine-step ladder where on the lowest step stand people who are absolutely not respected, and on the highest step stand those who are very respected. On which of the nine steps are you personally standing today?" The possible

[12] This is the same bias that would arise from ignoring the stable part of the preference heterogeneity. Intuitively, if individuals who have higher marginal utility from a certain consumption good are on average more likely to consume more of that good, then ignoring preference heterogeneity will lead to an over-estimation of the average marginal utility for that good.

[13] Source: Russia Longitudinal Monitoring survey, RLMS-HSE, conducted by Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS. RLMS-HSE sites: http://www.cpc.unc.edu/projects/rlms-hse, http://www.hse.ru/org/hse/rlms.

[14] This database is representative of the Russian population. For previous economic applications of this dataset on life satisfaction, see for example Senik (2004) and Perez-Truglia (2013).

[15] Other studies commonly find significant differences when using happiness instead of life satisfaction (Kahneman & Deaton, 2010). Indeed, Benjamin et al. (2012, 2014) report that evaluative subjective well-being measures, such as life satisfaction, are better at predicting choices than affective happiness. Thus, our results with life satisfaction data should not be directly extrapolated to other subjective measures such as happiness scores.

[16] The RLMS does include data on expenditures in non-food items. However, we focused on expenditures of food items because expenditures in non-food items were measured with much less detail. For example, expenditures in clothing were measured with a single survey question, in contrast to the dozens of questions used to measure food expenditures. Using non-food items would also introduce other complications, such as having to deal with zero or missing values in expenditures (e.g., over 35% of households report zero or missing clothing expenditures) and having to deal with consumption of non-durable goods.

[17] Expenditures can still be systematically different from consumption due to a variety of reasons such as measurement error, stockpiling and household production. However, if the degree of measurement error is similar across the two consumption items, then the attenuation bias will be roughly proportional to the coefficients and it will not affect the resulting marginal rate of substitution. And given the logarithmic specification, stockpiling and household production will not be a problem as long as expenditures are roughly proportional to consumption.

[18] That is, the underlying difference in satisfaction from answering "not at all satisfied" (1) and "less than satisfied" (2) need not be equal to the corresponding difference between "rather satisfied" (4) and "fully satisfied" (5). The POLS method consists of assigning values to the categories by fitting an ordered probit to the raw sample fractions. For example, if a fraction q reports the lowest category (not at all satisfied), that means that the highest satisfaction among the lowest category must be $\Phi^{-1}(q)$, where Φ is the cumulative distribution of a standard normal. Thus, the POLS method assigns the lowest category an score of $E[z|z < q]$, where z is distributed standard normal.

answers range from 1 to 9. The last measure of subjective well-being considered is the health self-evaluation, which measures how the individual perceives her own health, on a scale from "Very Bad" (1) to "Very Good" (5).

The regressions always include individual-specific fixed effects, year dummies, and a set of additional control variables: the number of household members, a dummy for household head, age, age squared, a set of four dummies for marital status, eight region dummies and their interactions with the time effects. We had to exclude some households from the sample. First, we dropped households who reported missing or zero food expenditures or income (less than 1% of the sample).[19] To limit the influence of outliers, we excluded households in the top 1% of either the distribution of income or the distribution of food expenditures. Last, since the model has to be estimated with up to four variables with individual-specific coefficients, we restricted the final sample to individuals who respond in more than four years. The final sample includes 70,317 observations on 7,104 individuals, implying that individuals in the sample respond for an average of ten years. See Table 1 for data definitions and Table 2 for descriptive statistics.

Fig. 1 shows some basic facts about the distribution of the expenditure shares in animal source food. First, Fig. 1a shows a histogram of the expenditure shares across individuals. This histogram suggests that there is substantial cross-individual variation in expenditure choices.[20] This heterogeneity is crucial for the empirical analysis, because the main test consists of predicting cross-individual differences in expenditure choices. Second, Fig. 1b shows a binned scatter plot of the relationship between the share of expenditures in animal source food and the total food expenditures. For each 1% increase in total food expenditures, the share of food expenditures from animal source increases by just 0.037 percentage points ($p\text{-value}<0.01$). Even though statistically significant, the magnitude of this association is very small.[21] This finding suggests that the vast majority of the cross-individual differences in expenditure choices cannot be explained mechanically by differences in income. In turn, this finding suggests that life satisfaction cannot predict choices due to a mechanical correlation between life satisfaction and income. We will provide more direct evidence on this respect in the following section, where we use income instead of subjective well-being as a benchmark for predicting expenditure choices.

5. Results

Table 3 presents the final results. All coefficients are shown with bootstrapped standard errors, clustered at the individual level. Column (1) reports the baseline specification, which uses life satisfaction as the dependent variable. Columns (2) though (7) report specifications with additional control variables or with alternative dependent variables.

Table 3 reports the averages of the predicted and actual shares of expenditures in animal source food. Both of these averages are computed for the subset of individuals without indeterminate predictions. Table 3 also reports the proportions of individuals with an indeterminate prediction and with an extreme prediction. The proportions of individuals with indeterminate and extreme predictions are substantial: 14.4% and 60%, respectively, according to column (1). This result is partially due to the low statistical precision of the individual-specific parameters, which depend largely on the number of time periods that the individuals are observed (on average, only ten periods). However, a significant proportion of indeterminate and extreme predictions may also arise because reported utility is a bad approximation for decision utility. In any case, we must note that the proportions of indeterminate and extreme predictions are similar across all specifications reported in Table 3, suggesting that indeterminate and extreme predictions cannot explain the differences in findings across different specifications.

Column (1) of Table 3 shows that the difference between the average predicted share (51.2%) and the average actual share (57.1%) is statistically significant but economically small. Although this finding suggests that life satisfaction can accurately predict average expenditure choices, similar (or even better) accuracy is attained by other objective and subjective measures of well-being, such as income in column (4) and economic rank in column (6). The high prediction accuracy is most likely due to the fact that the correct prediction rate is close to 50%. As a result, using pure noise as a dependent variable would similarly result in an average predicted share close to 50%.

The bigger challenge for the life satisfaction data is predicting cross-individual differences in expenditure choices. However, this test would not be informative, because individual predicted shares are imprecisely estimated. The lack of precision results from the fact that each pair of marginal utilities is estimated from a regression with ten observations (i.e., the average number of years that the individuals remain in the sample) and three independent variables (i.e., the two expenditure variables plus the individual fixed effect). Nevertheless, the estimates are precise enough to study the weak version of the validity test introduced in Section 2 (i.e., the cross-individual association between actual and predicted choices). The first row of Table 3 presents this cross-individual association between predicted and actual expenditure shares. To account for the fact that predicted shares cannot be lower than zero or higher than one, the association coefficient was using a Tobit regression.

---

[19] For observations with missing values in the additional control variables (e.g., number of household members, household head dummy), we imputed those values by using the average values in the entire sample. In any case, the differences are negligible if we instead drop those observations.

[20] It is important to note that the vast majority of responses to this survey are collected every year during the same calendar month, so that this cross-individual variation is not a mechanical product of seasonality. Additionally, note that some of this cross-individual variation in choices can reflect other things besides preference heterogeneity, such as measurement error, experimentation, etc.

[21] Note that each of the dots from Fig. 1b correspond to a decile of the distribution of total food expenditures. Even the extreme move from the lowest to the highest decile has a small effect in the expenditure share.

Table 1
Data definitions.

Variable name	Definition
Life satisfaction	Individual response to the question: "To what extent are you satisfied with your life in general at the present time? [Not at all satisfied 1] [Less than satisfied 2] [Both yes and no 3] [Rather satisfied 4] [Fully satisfied 5]"
Economic satisfaction	Individual response to the question: "Tell me, please, how satisfied are you with your economic conditions at the present time? [Not at all satisfied 1] [Less than satisfied 2] [Both yes and no 3] [Rather satisfied 4] [Fully satisfied 5]"
Economic rank	Individual response to the question: "And now another nine-step ladder where on the lowest step stand people who are absolutely not respected, and on the highest step stand those who are very respected. On which of the nine steps are you personally standing today? [Lowest Step 1] [2] ... [Highest Step 9]"
Health self-evaluation	Individual response to the question: "Tell me, please: How would you evaluate your health? [Very bad 1] [Bad 2] [Average, not good, but not bad 3] [Good 4] [Very good 5]"
Expenditures in animal source food	Household monthly expenditures (in thousands of rubles deflated to year 2000) on meats, dairy products, fish, eggs and fat, constructed using data on several questions on expenditures on individual items
Expenditures in non-animal source food	Household monthly expenditures (in thousands of rubles deflated to year 2000) on potato, bread, sugar, fruits, vegetables and other food items, constructed using data on several questions on expenditures on individual items
Income	Household monthly income (in thousands of rubles deflated to year 2000). This variable was constructed by the RLMS using data on several sources of income, such as cash and non cash salaries, other paid work, unemployment benefits and pensions, state transfers, private transfers, among others

Table 2
Descriptive statistics.

	Observations	Mean	Sd	Min	Max
Life satisfaction	70,317	2.80	1.11	1.00	5.00
Economic satisfaction	58,950	2.28	1.07	1.00	5.00
Economic rank	69,494	3.81	1.40	1.00	9.00
Health self-evaluation	70,096	3.11	0.70	1.00	5.00
Expenditures in animal source food	70,317	1922.86	1392.92	19.65	8821.92
Expenditures in non-animal source food	70,317	1397.24	1102.75	15.49	7316.46
Income	70,317	5881.37	4250.78	1.15	54138.96

Notes: Data from the RLMS over the period 1994–2011. The sample excludes individuals with zero or missing information for expenditures/income, observations in the top-1% of the distribution of expenditures/income and individuals who were respondents in fewer than five rounds. See Table 1 for data definitions.

The results from the baseline specification shown in column (1) suggest that the correlation between predicted and actual expenditure shares is economically large and statistically significant at the 1% level. The regression coefficient of 0.76 implies that, for each percentage point increase in the actual expenditure share of an individual, the individual's predicted expenditure share increases by 0.76 percentage points. Indeed, we cannot reject the null hypothesis that the regression coefficient is equal to 1, which would be the ideal value (for the weak test). However, given how imprecisely estimated this regression coefficient is, we cannot reject the null hypothesis of a coefficient of 0.5 either.

Columns (2) and (3) reproduce the results, with some changes in the specification for robustness checks. A first concern is that the baseline regression does not control for non-food expenditures. If expenditures in one of the food categories were more strongly correlated to non-food expenditures, then that could introduce systematic biases. Given the previous evidence that the composition of food expenditures does not change with income (Fig. 1b), this possibility seems unlikely. To provide more direct evidence, the specification in column (2) is identical to that in column (1), except that the regression includes the logarithm of income as an additional control variable (with individual-specific coefficients). This variable is intended to control for differences in non-food expenditures. As expected, the results in column (2) are similar to the baseline results in column (1). If anything, the association between predicted and actual shares increases slightly from 0.76 in column (1) to 0.79 in column (2).

Failing to control for relative prices can also introduce omitted-variable biases because of the mechanical correlation between prices and expenditures. Indeed, some of the control variables included in the baseline specification are meant to control for prices indirectly (e.g., the interaction between region and time effects). To address this concern more directly, we construct a price index for each of the two consumption categories using data on prices from the community-level survey of the RLMS. We correct the expenditure variables by dividing them by their respective price indexes. The results, presented in column (3), are similar to the results from the baseline specification in column (1). If anything, controlling for prices increases the association between predicted and actual expenditure shares from 0.76 in column (1) to 0.89 in column (3).

Another concern is that even an objective measure of well-being, such as income, may perform as well as life satisfaction in this weak version of the validity test.[22] Again, the finding that the composition of food expenditures does not change with income (Fig. 1b) suggests otherwise. To address this concern more directly, column (4) presents the results for the baseline

[22] Note that, at least in theory, an outcome variable could perform worse in the weak version of the test and yet perform better in the strong version of the test.

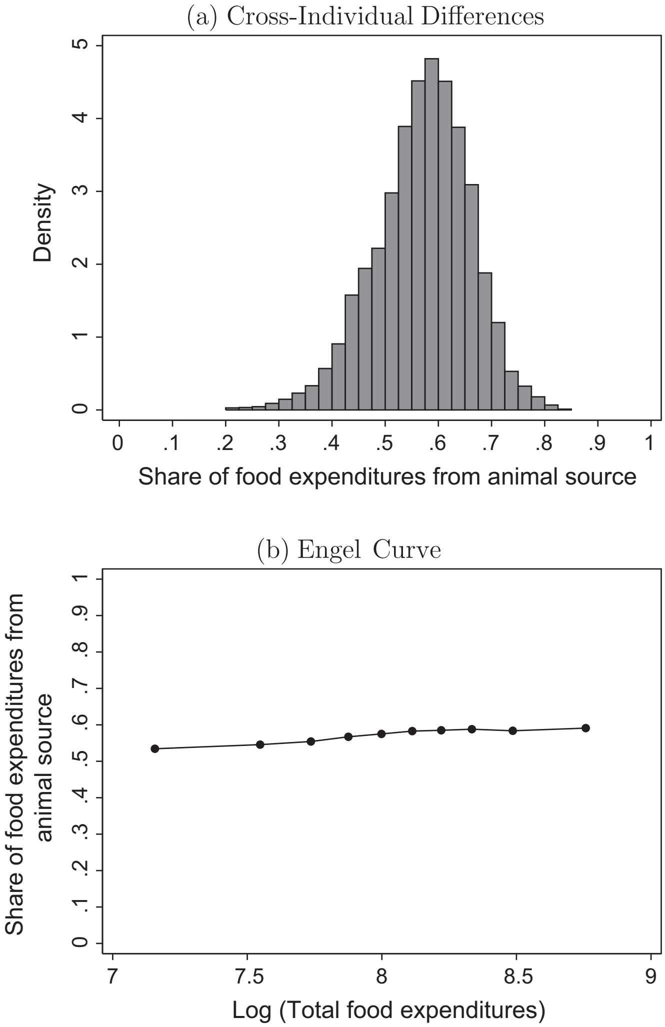

Fig. 1. Share of food expenditures in animal source food. *Notes*: Panel a. is the histogram of the average share in animal source food (averaged over all the years that each household is observed). Panel b. is a binned scatter plot of the relationship between the share in animal source food and total food expenditures. Data from the RLMS over the period 1994–2011. The sample only includes observations for the 7,104 individuals used to estimate the model in column (1) from Table 3. See Table 1 for data definitions, Table 2 for descriptive statistics and its note for more details about the sample.

specification, using the logarithm of income instead of life satisfaction as the dependent variable. As expected, there is no significant association between the actual shares and the shares "inferred" from income data: the regression coefficient is close to zero (0.140) and statistically insignificant.

Finally, we can assess whether other measures of subjective well-being besides satisfaction can also satisfy this same validity test. Indeed, Benjamin et al. (2012, 2014) find marked differences between evaluative measures, such as life satisfaction, and other measures of subjective well-being. The last three columns reproduce the baseline specification, using other measures of well-being instead of life satisfaction as the dependent variable: specifically, economic satisfaction in column (5), perceived economic rank in column (6), and health self-evaluation in column (7). Column (5) shows that economic satisfaction does not pass the same validity test: the regression coefficient between predicted and actual choices is close to zero (-0.003) and statistically insignificant. It is remarkable that the results are so different, given that life satisfaction and economic satisfaction are strongly correlated (correlation coefficient of 0.54) and would be more highly correlated if not for the degree of measurement error commonly found in this type of subjective measure.[23]

Column (6) shows that, when economic rank is the outcome variable, the regression coefficient between predicted and actual choices is 0.39 (statistically significant at the 10% level). Still, this regression coefficient of 0.39 is half the magnitude of the corresponding coefficient of 0.76 for life satisfaction (and their difference is statistically significant at the 10% level).

[23] For instance, the test–retest correlation over four weeks for a five-item life satisfaction measure is about 0.77 (Lucas, Diener, & Suh, 1996), which is significantly below the corresponding correlation of about 0.9 for objective outcomes such as education.

Table 3
Regression results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Life sat.	Life sat.	Life sat.	Log(income)	Econ. sat.	Econ. rank	Health eval.
Regression coefficient	0.763	0.792	0.899	0.140	−0.003	0.391	−0.013
Predicted-actual behavior	(0.196)	(0.240)	(0.270)	(0.158)	(0.271)	(0.222)	(0.258)
Expenditures share in animal source food:							
Mean, actual	0.571	0.572	0.572	0.571	0.571	0.572	0.572
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean, predicted	0.512	0.504	0.510	0.533	0.403	0.492	0.519
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.007)
Proportion indeterminate	0.144	0.167	0.168	0.079	0.143	0.136	0.181
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.005)	(0.006)
Proportion extreme	0.601	0.603	0.604	0.568	0.626	0.612	0.618
	(0.007)	(0.007)	(0.007)	(0.008)	(0.015)	(0.008)	(0.008)
Controlling for income	No	Yes	Yes	No	No	No	No
Correcting for prices	No	No	Yes	No	No	No	No
Observations	70,317	70,317	70,317	70,317	58,950	69,494	70,096
Individuals	7,104	7,104	7,104	7,104	7,104	7,103	7,104

Notes: Each column corresponds to a separate regression of the corresponding dependent variable on two main independent variables: the logarithms of expenditures in animal and non-animal source food. Their coefficients are allowed to be individual-specific. Those coefficients were used to evaluate the formulas described in Section 5. All standard errors are bootstrapped and clustered at the individual level. The *Regression Coefficient Predicted-Actual Behavior* corresponds to the coefficient from a Tobit regression of predicted expenditure shares on actual expenditure shares from animal food. The *Proportion Indeterminate* corresponds to the proportion of individuals for which the estimated marginal utilities are negative for both food categories. The *Proportion Extreme* corresponds to the proportion of individuals for which exactly one of the two estimated marginal utilities are negative, which results in the extreme prediction that all expenditures should go to one of the two food categories. The regressions always include individual-specific fixed effects, year dummies, household size, a dummy for household head, age, age squared, a set of four dummies for marital status, eight region dummies and their interactions with the time effects. Columns (2) and (3) include the logarithm of income as additional control variable, allowing for individual-specific coefficients. In column (3) the expenditures variables are divided by their corresponding price indexes. All dependent variables with a subjective scale (life satisfaction, economic satisfaction and health self-evaluation) are transformed using the Probit-OLS method. Data from the RLMS over the period 1994–2011. The sample excludes individuals with zero or missing information for expenditures/income, observations in the top-1% of the distribution of expenditures/income and individuals who were respondents in fewer than five rounds. See Table 1 for data definitions and Table 2 for descriptive statistics.

Last, column (7) shows the results for a subjective measure of well-being that should not accurately predict expenditure choices: health self-evaluation. As expected, the association between predicted and actual expenditure shares reported in column (7) is close to zero (-0.01) and statistically insignificant.

6. Conclusions

We proposed a general validity test for subjective well-being data. The test consists of three steps: inferring preferences from subjective well-being data through regression analysis; using those estimates to predict how a rational utility-maximizer individual should act; and comparing predicted behavior with actual behavior. We provided an application of this test based on a model of food consumption. In this application, the estimates of the predicted choices are not nearly precise enough to compare actual and predicted choices for each individual. However, we conducted a weak version of the test by measuring the cross-individual association between actual and predicted expenditure shares. We found that, for each percentage point increase in the actual expenditure share of an individual, the individual's predicted expenditure share (inferred from life satisfaction data) increases by 0.76 (SE 0.196) percentage points. Other objective and subjective measures of well-being (e.g., income, satisfaction with economic conditions) performed significantly worse. Overall, the evidence suggests that life satisfaction contains some useful information about experienced utility.

Our empirical application has obvious limitations regarding the data and the model. However, if anything, this means that there is a positive association between predicted and actual choices in spite of these limitations. Future research can apply our validity tests in other contexts. For example, one interesting application would be to test whether preferences inferred from life satisfaction data can predict work-related choices, such as occupation and work hours.[24] Furthermore, there is no reason to believe that any single measure of well-being should be consistent with all possible marginal rates of substitution (Becker & Rayo, 2008; Benjamin, Heffetz, Kimball, & Szembrot, 2014). It is possible that life satisfaction provides accurate representations of some marginal rates of substitution (e.g., between two categories of food) but inaccurate representations of other marginal rates of substitution (e.g., between consumption and leisure). Thus, future research should aim at identifying the combination of subjective measures of well-being that is most consistent with a range of behaviors.

[24] The challenge is to exploit credible identification strategies. For example, in the case of work-related choices one could exploit changes in labor laws or income taxes to obtain exogenous variation from which to infer preferences.

References

Becker, G. S., & Rayo, L. (2008). Comment on 'Economic growth and subjective well-being: Reassessing the Easterlin Paradox' by B. Stevenson and J. Wolfers. Brookings papers on economic activity, Spring 2008 (pp. 88–95)..

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2012). What do you think would make you happier? What do you think you would choose?. *American Economic Review, 102*(5), 2083–2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2014). Can marginal rates of substitution be inferred from happiness data? Evidence from residency choices. *American Economic Review, 104*(11), 3498–3528.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Szembrot, N. (2014). Beyond happiness and satisfaction: Toward well-being indices based on stated preference. *American Economic Review, 104*(9), 2698–2735.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review Papers and Proceedings, 91*(2), 67–72.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review, 105*(5), 1–33.

Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research, 112*(3), 497–527.

Diener, E., Sandvik, E., & Pavot, W. (1991). Happiness is the frequency, not the intensity, of positive versus negative affect. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being*. New York: Pergamon Press.

Di Tella, R., & MacCulloch, R. (2006). Some uses of happiness data in economics. *The Journal of Economic Perspectives, 20*(1), 25–46.

Di Tella, R., MacCulloch, R., & Oswald, A. (2003). The macroeconomics of happiness. *Review of Economics and Statistics, 85*(4), 809–827.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. *Nations and Households in Economic Growth, 89*, 89–125.

Easterlin, R. A. (1995). Will raising the incomes of all increase the happiness of all?. *Journal of Economic Behavior and Organization, 27*, 35–47.

Edgeworth, F. Y. (1881). Mathematical psychics: An essay on the application of mathematics to the moral sciences. Reprinted (1967). New York: M. Kelly..

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness?. *The Economic Journal, 114*, 641–659.

Gruber, J., & Mullainathan, S. (2005). Do cigarette taxes make smokers happier? Advances in economic analysis and policy (Vol. 5 (1)), Article 4..

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

Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics, 112*(2), 375–405.

Lucas, R., Diener, E., & Suh, E. M. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71*(3), 616–628.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R., Kling, J. R., & Sanbomatsu, L. (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science, 337*, 1505–1510.

Luttmer, E. F. P. (2005). Neighbors as negatives: Relative earnings and well-being. *Quarterly Journal of Economics, 120*(3), 963–1002.

Oswald, A. J., & Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics, 92*(5–6), 1061–1077.

Oswald, A. J., & Wu, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence of the U.S.A. *Science, 327*, 576–579.

Perez-Truglia, R. (2013). A test of the conspicuous-consumption model using subjective well-being data. *Journal of Socio-Economics, 45*, 146–154.

Samuelson, P. (1948). Consumption theory in terms of revealed preference. *Econometrica, 15*, 243–253.

Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective well-being: The convergence and stability of self-report and non-self-report measures. *Journal of Personality, 61*(3), 317–342.

Senik, C. (2004). When information dominates comparison: Learning from Russian subjective panel data. *Journal of Public Economics, 88*(9–10), 2099–2123.

Stevenson, B., & Wolfers, J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin Paradox. NBER Working Paper No. 14282..

Stevenson, B., & Wolfers, J. (2009). The paradox of declining female happiness. *American Economic Journal: Economic Policy, 1*(2), 190–225.

Stone, Arthur A., Shiman, Saul S., & DeVries, Marten W. (1999). Ecological momentary assessment. In Daniel Kahneman, Edward Diener, & Norbert Schwarz (Eds.), *Well-being: The foundations of hedonic psychology*. New York: Russell-Sage.

Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica, 38*(2), 311–323.

Urry, H. L., Nitschke, J. B., Dolski, I., Jackson, D. C., Dalton, K. M., Mueller, C. J., Rosenkranz, M. A., Ryff, C. D., Singer, B. H., & Davidson, R. J. (2004). Making a life worth living: Neural correlates of well being. *Psychological Science, 15*(6), 367–372.

Van Praag, B., & Ferrer-i Carbonell, A. (2007). *Happiness quantied: A satisfaction calculus approach*. Oxford: Oxford University Press.