Economica (2001) **68**, 335–357

Identifying Welfare Effects from Subjective Questions

By Martin Ravallion and Michael Lokshin
World Bank

Final version received 30 August 2000.

We argue that the welfare inferences drawn from answers to subjective–qualitative survey questions are clouded by concerns over the structure of measurement errors and how latent psychological factors influence observed respondent characteristics. We propose a panel data model that allows more robust tests and we estimate the model on a high-quality survey for Russia. We find significant income effects on an individual's subjective economic welfare. Demographic effects are weak at given income per capita. Ill-health and becoming unemployed lower welfare at given current income, although the unemployment effect is not robust, and returning to work does not restore welfare without an income gain.

I. Introduction

A large literature in economics and psychology has sought to understand why some people purport to feel well off in interviews, while others do not. Answers have been sought in respondents' objective economic circumstances; and conclusions have been drawn about the welfare effects of changes in (for example) incomes, employment and household size.

The results have typically suggested that incomes do not have much power to explain perceived welfare. Substantial economic growth since the 1950s did not bring higher average self-rated happiness in either the United States (Easterlin 1974, 1995; Diener *et al.* 1999) or Japan (Veenhoven *et al.* 1993). Cross-sectional micro data often suggest a positive correlation with individual or household incomes, but it is generally not what one would call a strong correlation; DeNeve and Cooper (1999) quote a mean correlation coefficient between income and subjective wellbeing of 0.17 (over 85 independent samples).[1]

These studies have used broad welfare concepts ('happiness' or 'satisfaction with life', that undoubtedly embrace much more than command over market goods. One might conjecture that focusing on the narrower idea of 'economic welfare' or 'poverty' would reveal a far stronger relationship with income. We can offer two observations suggesting that it does not; one is from aggregate data and the other from micro data.

1. A well known example of a self-rated poverty measure is that used in the surveys conducted by the Social Weather Station (SWS) in the Philippines. Respondents in regular surveys are asked whether they are 'poor', 'borderline' or 'non-poor'. From the SWS data presented in Mangahas (1995), the proportion of the population saying they are 'poor' has not fallen with growth in GDP per capita in the Philippines.
2. In the main surveys for Russia that we use later, respondents rated their economic welfare on a nine-rung ladder from 'poor' to 'rich'. The correlation coefficients with household income per person are 0.11 for

1994 and 0.20 for 1996. Though both are highly significant by a conventional test,[2] they are not any higher than the correlations found with more holistic concepts of 'happiness' or 'life satisfaction'.

In attempting to explain observations such as these, it has been argued that it may in fact be *relative income*—relative to some reference group—that drives self-rated welfare, rather than absolute income. This is the now classic interpretation that Easterlin (1974) offered for the fact that aggregate happiness has responded so little to economic growth in the United States even though income and subjective wellbeing are correlated (albeit weakly) across people at one date.[3] Similarly, respondents to the SWS question on subjective poverty may well be strongly influenced by their perceived income relative to (say) the mean at each survey date. By this view, absolute income gains that are shared with all those in the relevant reference group do not raise perceived wellbeing.

Another respondent characteristic widely identified as important is unemployment. A number of papers have found that the unemployed have lower self-rated welfare with and without controls for income.[4] The adverse effect of unemployment at given income appears to contradict the prediction of the standard economic model of work–leisure choice that unemployment (and hence greater leisure) is desirable at given income. Implications have been drawn for policy discussions of the supposed disincentive effect of unemployment compensation (see, e.g. Blanchflower and Oswald 1997).

Evidence has also been found of strong demographic effects on subjective welfare (such as household size, stage of the life-cycle, marital status) and of effects of health status; for a recent survey of the socioeconomic correlates of subjective wellbeing, see Diener *et al.* (1999).

This paper identifies a number of problems that cloud the inferences that can be drawn from survey responses on subjective welfare. We draw on research in psychology suggesting that relatively stable personality traits influence how people respond to subjective welfare questions. Just as any welfare comparison requires a consistent preference ordering, identifying welfare effects in subjective data requires that we control for these latent psychological differences. We argue that failure to control for latent heterogeneity also biases welfare inferences at given tastes, since there is also evidence (mostly from psychology) that these same differences influence socioeconomic characteristics. We also point to a number of other potential biases in interpreting subjective data, related for example to the likely structure of measurement errors and to non-comparability of subjective welfare ladders.

We then propose and implement an approach that is likely to be more robust to the main problems identified. We use a comprehensive multi-purpose panel survey for Russia that also included the question on subjective economic welfare mentioned above. The income measure from these surveys is built up from a detailed questionnaire and so is undoubtedly more reliable than measures commonly found in the literature on subjective welfare, which are often based on a single question: 'What is your income?' The fact that the survey is longitudinal (re-surveying the same individuals over time) allows us to control for individual differences in personality that influence subjective welfare and are also likely to influence incomes, employment and other individual and household characteristics. In particular, we treat these effects as an idiosyncratic, time-invariant error component correlated with the socio-economic characteristics of respondents. Our model specification also incorporates other features of subjective welfare data suggested in the literature.

This is not the first use of panel data to study subjective welfare. The two closest antecedents are van de Stadt *et al.* (1985) and Winkelmann and Winkelmann (1998).[5] The former paper used panel data in modelling a money metric of subjective welfare in the Netherlands. In addition to the difference in the dependent variable (which we explain later), van de Stadt *et al.* do not allow for latent individual effects. They do, however, allow for dynamics, by including the lagged subjective welfare measure as a regressor (though they cannot reject the null that its coefficient is unity). The paper by Winkelmann and Winkelmann is our closest antecedent. Importantly, we believe, the authors allow for a latent individual effect when modelling satisfaction with life in Germany. In addition to exploring a similar question in different data, we depart from the methods of Winkelmann and Winkelmann in a number of ways. One difference is that we focus on subjective *economic* welfare, rather than the seemingly broader—but also more ambiguous—concept of 'satisfaction with life' that Winkelmann and Winkelmann use (in keeping with much of the literature). We focus on the more narrow concept in the expectation that it will offer sharper results on the welfare effects of economic variables: yes, 'money does not buy you happiness', but surely it makes you think you are less poor? We also use a better income measure, built up from a detailed survey. Other differences are that we allow for time-varying coefficients arising from the reference group effects discussed in the literature, and we allow for endogenous utility cut-offs in defining subjective ladders. By incorporating these features, we can offer a new estimation method for studying subjective data.

The following section surveys past work. Section III describes the setting of our data. Descriptive results can be found in Section IV, while our model and results on the determinants of subjective economic welfare can be found in Sections V and VI. Section VII offers some conclusions.

II. Sources of Bias in Identifying Welfare Effects from Subjective Data

Subjective welfare data are typically in the form of self-reported positions on a ladder that has a natural ordering, such as from 'poor' to 'rich', or from 'very bad' to 'very good'. We doubt that responses to this type of question are prone to serious measurement error; there is no obvious reason why respondents would not tell you how they feel at the time of interview. However, the literature in psychology has pointed to 'mood variability' as a factor in self-rated welfare (Diener *et al.* 1999). Two equally happy people may have very different variances in their happiness over time, and self-rated wellbeing can then vary greatly according to the time of interview. (An example of such a transient effect is how a recent experience ended; see Fredrickson and Kahneman 1993.) While mood variability reduces the statistical fit with regressors related to longer-term determinants of welfare, it is not an obvious source of bias.

However, there are reasons to be concerned about bias. Aggregation is one. Individual income in a cross-section may well be highly correlated with self-rated welfare across individuals, while at the same time aggregate economic growth results in little or no change in average self-rated welfare. This can happen if it is relative income that matters at the individual level—relative to (say) the mean income in the society as a whole—and if the economic growth is roughly distribution-neutral (incomes at all levels are growing at about the same rate). If inequality does not change much, or tends to increase with economic growth, then one can readily explain why economic growth does not translate into rising average happiness, without concluding that individual incomes are irrelevant to individual welfare.

Distributional effects can also arise from the data aggregation process. Subjective welfare questions are typically asked of individuals, whereas income measures are usually for households or even for large aggregates, such as nations. Then inequality within households (or countries) can influence the relationship between subjective welfare and average income. And this can hold even if subjective welfare does not depend directly on relative position. In particular, if individual subjective welfare is a concave function of income, then higher income inequality will lower average subjective welfare, holding average income constant.[6] Economic growth with rising inequality may entail little or no gain in average subjective welfare even when there is a strong income effect at the micro level.

Another concern is measurement error in reported incomes. The subjective wellbeing surveys used in much of the literature appear to have obtained 'income' from just one or a small number of questions. For example, the German survey used by Winkelmann and Winkelmann (1998, p. 14) obtains income from answers to a single question: 'If everything is taken together, how high is the total monthly income of all household members at present?' The resulting estimates could deviate substantially from what one would get from a standard household income and expenditure survey, with individual incomes identified by source, with imputations for income-in-kind. One then expects the income effect to be underestimated in most past studies, owing to the usual attenuation bias. This expectation assumes a white-noise measurement error in incomes; we return to this point, and consider other possible biases arising from the structure of measurement errors.

Misspecification of the relevant income variable can be a concern even with more detailed surveys. There is scope for debate at virtually every step in obtaining a measure of 'income'. There are issues about whether it should be income or expenditure, what should be included, how one should adjust for differences in household size, how cost-of-living deflators should be constructed and so on. The methods used in practice—even with elaborate surveys—need not accord well with subjective assessments. There may be differences in the time period over which income is measured versus the time period on which self-perceptions of wellbeing are based. Past incomes can also matter, through savings. So too may expected future incomes (or determinants of these) matter, either when utility is not intertemporally separable, or when the time period over which subjective welfare is being assessed is longer than that over which income is measured. Defensible alternative methods of measurement may well yield a stronger correlation.[7]

There are also attributes of subjective data that need to be considered in assessing income and other effects. Precisely because it is 'subjective', different people can have different personal notions of what a 'high' or 'low' level of subjective welfare means.[8] Concerns about the interpersonal comparability of subjective ladders appear to have been an important obstacle to their wider application in economics and other social sciences. Yet standard practice has entailed using methods such as an ordered probit on ladder data that assume the utility cut-offs in defining ladder rungs to be constant across all respondents.

An important source of heterogeneity in perceived welfare is personality. Some people seem to have been born happy, or to have persistent personality traits that make them happy. There is evidence from psychological research that intrinsic, intertemporarily stable personality traits systematically influence reported wellbeing. In a meta-analysis of research in psychology, De Neve and Cooper (1999) identify 137 personality traits correlated with subjective wellbeing, grouped under five commonly used headings in psychology: 'extraversion', 'agreeableness', 'conscientiousness', 'neuroticism' or 'emotional stability', and 'openness to experience'. These psychological traits are not normally measured in standard socioeconomic surveys, and even if they were, including them as regressors would create concerns about their endogeneity. Identifying the causal effect of a personality trait on mean happiness (say) would seem problematic; it is hard to imagine a valid instrumental variable, that is, correlated with observed personality traits but not correlated with happiness given personality.

Are these psychological factors in perceptions of wellbeing of concern in using such data to assess welfare effects? Of the 137 personality traits identified by De Neve and Cooper, the strongest correlates with subjective wellbeing within the five categories mentioned above are:[9] *extraversion*: 'social competence'; *agreeableness*: 'collective self-esteem', 'fear of intimacy' (negative), 'interpersonal locus of control', 'social emotionality', 'social interest', 'social tempo', 'trust'; *conscientiousness*: 'desire for control', 'inhibition' (negative), 'plasticity'; *neuroticism*: 'distress' (negative), 'emotional stability', 'rebellious-distrustful' (negative), 'repressive defensiveness' (negative), 'social anxiety' (negative), 'tension' (negative); *openness to experience*: 'self-confidence', 'self respect'. These are differences in tastes which one would want to control for in making interpersonal comparisons of welfare for most purposes (such as for tax or welfare policy-making); the fact that a person is inhibited, rebellious or unconfident would not normally constitute a case for favourable tax treatment, for example. If these psychological factors happened to be uncorrelated with the other variables of interest, then we would not need to control for them when measuring the welfare effect of unemployment, say. Explanatory power will be lower, but the latent psychological factors will not bias the results.

However, it is plausible that a number of the personality traits that raise self-rated welfare are also positively correlated with income and negatively correlated with unemployment. The above list of personality traits thought to promote a feeling of wellbeing overlaps considerably with the desirable things human resource managers are told to look for when interviewing job candidates (Darity and Goldsmith 1996). This makes sense, since there is evidence that happy workers are more productive in various ways. (Frank 1985 reviews the evidence.) For example, there is a large literature in psychology suggesting that various personality traits influence worker absenteeism (examples include Judge *et al.* 1997; Kivimaki *et al.* 1997; Salgado 1997); some of the traits identified overlap noticeably with those thought to influence subjective wellbeing, such as extraversion, conscientiousness and emotional stability (De Neve and Cooper 1999). One can also conjecture that certain personality traits simultaneously promote happiness, but make survey respondents disinclined to say they are sick. Thus, it can be argued that the income and health effects on subjective welfare will be overestimated, as will the absolute effect of unemployment. (The actual effect will be less negative than the estimated effect.)

One can also expect the income measurement error to be correlated with other variables of interest. For example, it is often conjectured that the rich tend to understate their incomes when asked by a stranger in an interview for some survey. (This is not implausible in Russia in the mid-1990s.) They are also less likely to be unemployed. Then the negative correlation between unemployment and subjective welfare could be due entirely to this structure of measurement errors; unemployment will appear to lower subjective wellbeing even if it has no real welfare effect beyond the loss of income. Similarly, if the time period over which incomes are measured is too short—and it is a longer-term income concept that drives self-assessments of welfare—we can expect unemployment and possibly other characteristics to be correlated with the difference between the two income measures. Unemployment might have a significant negative effect on subjective welfare at given current income simply because respondents naturally worry about future income too. That does not of course mean that leisure is undesirable, or that there are no adverse incentive effects of unemployment compensation. The high estimate of the level of unemployment benefits needed to create unemployment implied by the results in the literature (see e.g. Winkelmann and Winkelmann 1998) could well reflect this structure of income measurement error, whereby the income effect is underestimated while the unemployment coefficient is overestimated.

The likely endogeneity of income to subjective welfare also clouds past efforts to test the claim that it is income relative to some reference group—not absolute income—that matters to wellbeing. Van de Stadt *et al.* (1985) and Clark and Oswald (1996) regress a subjective welfare indicator on both 'own income' and an estimate of 'comparison group income', namely the mean income of people with similar characteristics. The 'comparison group' income is found to have a significant negative coefficient with own income entering positively; the authors conclude that it is relative income that matters to welfare. However, the significant effect of predicted income could also reflect a misspecification. Suppose that earnings are influenced by latent personality traits in subjective welfare via the effects of higher job satisfaction on labour turnover and disputes (as discussed in, e.g., Frank 1985, and indeed Clark and Oswald 1996). Then the significance of predicted income could be due solely to a correlation between own income and unobserved determinants of subjective welfare.[10] Income endogeneity can generate spurious comparison group effects.

The existence of social effects also raises concerns about assuming time-invariant utility parameters. Suppose that the self-rated welfare of person i at date t, u_{it}, is a linear function of income relative to the mean y_{it}/m_i, where the income of person i at date $t = 1, 2$ is y_{it} with mean m_t formed over all i at t. We can write this relationship as a linear utility function with time-varying parameters $u_{it} = a + by_{it}/m_t = a + \beta_t y_{it}$ where $\beta_t = b/m_t$. (If the model is $u_{it} = a + b\log(y_{it}/m_t)$, then the time-varying parameter is the intercept.) This offers an alternative approach to allowing for reference group effects to those based on comparison group means, as found in the literature. It is however an inconclusive test, since accepting the null that the model's parameters do not vary over time could mean either that relativities do not matter (i.e. that only own income matters) or that mean income does not change over time.

An important strand of the economics literature on subjective welfare has instead tested for effects on self-reported *money metrics* of welfare. Respondents are asked what income they need to secure stipulated welfare levels. An example is the 'income evaluation question' (IEQ): 'What after-tax income do you consider very bad, bad, sufficient, good, very good?'[11] A special case is the minimum income question: 'What income do you need to make ends meet?' The answers are then regressed on actual income and other variables, such as reference group income as in van de Stadt *et al.* (1985).

The money metric approach can offer important insights (such as in setting a social subjective poverty line, as in Kapteyn *et al.* 1988; but also see Pradhan and Ravallion 2000). However, the estimated regressions do not in general provide unbiased estimates of the effects of socioeconomic characteristics on subjective welfare. To see why, let u denote subjective welfare, which is a function of income y and characteristics x, as $u = u(y, x)$, with $u_y > 0$. The welfare level to make ends meet is u^*. The income level needed to reach u^* when characteristics are x is denoted y^* and is defined implicitly by $u^* = u(y^*, x)$. Now assume that u^* is a rising function of actual welfare, $u^* = g(u)$. Then y^* is a function of y and x (the function is given implicitly by equating $u(y^*, x) = g[u(y, x)]$) with derivatives:

$$(1) \qquad \frac{\partial y^*}{\partial y} = \frac{g_u(u)u_y(y, x)}{u_y(y^*, x)}; \qquad \frac{\partial y^*}{\partial x} = \frac{g_u(u)u_x(y, x) - u_x(y^*, x)}{u_y(y^*, x)}.$$

It is evident from (1) that the derivatives of y^* with respect to y and x (as estimated by regressing y^* on y and x) do not identify the corresponding marginal utilities. The marginal rate of substitution (MRS) between y and x (u_x/u_y) is identified in the special case in which the MRS does not vary with income (precisely, $u_x(y, x)/u_y(y, x) = u_x(y^*, x)/u_y(y^*, x)$); then

$$(2) \qquad \frac{u_x(y, x)}{u_y(y, x)} = \frac{\partial y^*/\partial x}{(\partial y^*/\partial y) - 1}.$$

In summary, both the cross-sectional micro evidence and the aggregate time-series evidence available could well have a hard time revealing the true welfare effects of changes in socioeconomic characteristics. The direction of bias is unclear on *a priori* grounds.

III. Setting and Data

The sharp drop in GNP in Russia in the 1990s was accompanied by a sharp rise in income poverty (Lokshin and Popkin 1999). What happened to subjective wellbeing? Table 1 combines answers to a standard question on 'satisfaction with life' from two different nationally representative surveys. The questions are not identical, but they are similar. (We do not use the first survey in the main analysis.) While only 13% of respondents to the 1991 survey said they were unsatisfied (including 'not at all satisfied' or 'completely unsatisfied'), this rose to 72% immediately after the economic reforms. Possibly the wording 'less than satisfied' (in the 1992 and subsequent surveys) is somewhat less strong than 'unsatisfied' (in the 1991 survey); this might account for some of the difference. However, looking at the distribution across the categories of potential responses, it appears from Table 1 that there was a sharp deterioration in subjective welfare.[12]

Of course (following the observations made in the last section), these aggregate results do not imply that self-rated welfare is responsive to absolute income changes at the micro level. Possibly self-rated welfare is driven entirely by relative income, and it is the rise in inequality in Russia that is driving these results. We next turn to micro panel data.

Since the pre- and post-reform data used in creating Table 1 are from different surveys, and interviewed different people, we cannot span the reforms. The post-reform data, however, are longitudinal, though there are two distinct panels, 1992–93, and 1994–96. Between the latter two years there was a marked increase in the proportion of the sample in the least-satisfied category, and this is the period we will focus on.

We use the Russian Longitudinal Monitoring Survey (RLMS) for 1994 and 1996.[13] RLMS is based on the first nationally representative sample of several

Table 1

Satisfaction with Life in Russia, 1991–96 (%)

	Pre-reform	Post-reform				
	1991	1992	1993	1994	1995	1996
Fully satisfied ('Completely satisfied' in the 1991 survey)	5	2	2	3	3	2
Rather satisfied ('Satisfied' in 1991)	39	9	10	11	10	9
Both yes and no (same)	43	18	20	20	20	20
Less than satisfied ('Unsatisfied' in 1991)	11	40	36	41	37	37
Not at all satisfied ('Completely unsatisfied' in 1991)	2	32	32	24	30	31
Sample size		10,692	10,894	8082	7682	7584

Sources: The 1991 survey is the General Social Survey of the European USSR, April–May 1991. The data and documentation are found on web site of the Inter-university Consortium for Political and Social Research: http://www.iopsr.umich.edu, ICPSR number: 6500. The other surveys are the Russian Longitudinal Monitoring Survey (RLMS), which is described in Section II of this paper; the sample for the latter changed in 1994.

thousand households across the Russian Federation.[14] The RLMS was designed as a panel, and we can track 5588 adults over the rounds for 1994 and 1996, slightly over 5000 with complete data.

The survey included the following question: 'Please imagine a 9-step ladder where on the bottom, the first step, stand the poorest people, and on the highest step, the ninth, stand the rich. On which step are you today?' We will call this the Economic Ladder Question (ELQ). The question does not presume that 'income' is the relevant variable for defining who is 'poor' and who is not, but leaves that up to the respondent. At the same time, by using the words 'poor' and 'rich', the question focuses on a more narrow concept of welfare than the questions often used in surveys, which refer to broader welfare concepts such as 'happiness' or 'satisfaction with life'. It does not appear plausible to us that discrepancies between answers to the ELQ, as posed above, and an objective measure of real income reflect the fact that they are aiming to measure different things. The real income measure is, after all, calibrated (in theory at least) to a utility function that can be more or less broad. It is not unreasonable to assume that both are aiming to measure the same thing, which we will call 'economic welfare'.

All adults in the sampled households were asked the ELQ. We decided to condense the highest 7th, 8th and 9th rungs into one, because of the small number of respondents who assigned themselves to rungs 8 and 9. (Only 28 of the 7405 respondents put themselves in rung 8 and only 3 put themselves on rung 9.) So we treat the data as a seven-rung ladder.

The income variable we use is total real monthly disposable household income (in June 1992 prices); this includes wages and salaries, social security, private transfers, and income in-kind and from home production. To convert to real values, we use well established region-specific poverty lines as deflators (Popkin *et al.* 1995).

As always, there are various sources of error in measured real incomes, in both the levels and changes over time. There are concerns about whether the variables in the deflators have been measured well, and weighted correctly. There is likely to be deliberate under-reporting of certain components of income, which are not legal or for which tax was evaded. This will be less of a problem for the data from those households that depend more heavily on legal income sources, notably wages from a regular job and governmental transfers. As one check, we will redo our regressions on a subsample restricted to the 2000 adults living in households for whom reported income is at least 90% from wages, pensions, unemployment benefits, child benefits, other governmental transfers and/or stipends. Our expectation is that these income components will be measured more accurately, though there is undoubtedly some measurement error here too, if only because we are basing the calculation on reported incomes. This restricted sample cannot be considered representative.

We will use expenditures as well as incomes. The expenditure measure is comprehensive, including imputations for consumption in kind (such as from family farms or enterprises) as well as cash expenditures. Including expenditures can help compensate for certain types of income measurement error; for example, a household with illegal income is unlikely to report that income accurately, but it could well be better reflected in expenditures. It can also be argued that subjective welfare will depend more on long-run ('permanent'), income, which will be better reflected in current expenditures than current incomes.

IV. Some Descriptive Statistics

Table 2 summarizes responses to the ELQ. The row and column totals give the number of respondents for each ladder rung. By exploiting the panel nature of the survey, we also give the numbers of respondents with each possible combination of responses. Thus, one can use the table to see how much movement up and down the ladder there was. Comparing the column and row totals, we find that there was an increase in the proportion of adults reporting that they are in the poorest few rungs, though there is not first-order dominance over the distribution. Taking the poorest two rungs to be the subjectively poor, the subjective poverty rate rose from 28.7% to 31.6%.

Table 2
Movements Up and Down the Subjective Economic Welfare Ladder, Russia, 1994–96[a]

1994	1996							Total (%) [cum. %]
	1	2	3	4	5	6	7+	
1	253	181	140	69	59	14	6	722 (12.92) [12.92]
2	189	222	212	132	107	9	9	880 (15.75) [28.67]
3	148	245	388	305	226	38	26	1376 (24.62) [53.29]
4	111	153	271	288	299	61	22	1205 (21.56) [74.85]
5	95	114	214	254	326	71	24	1098 (19.65) [94.50]
6	24	19	29	50	71	16	9	218 (3.90) [98.46]
7+	3	6	16	30	17	6	11	89 (1.59) [100.00]
Total	823	940	1270	1128	1105	215	107	5588
(%)	(14.73)	(16.82)	(22.73)	(20.19)	(19.77)	(3.85)	(1.91)	(100.00)
[cum. %]	[14.73]	[31.55]	[54.28]	[74.47]	[94.24]	[98.09]	[100.00]	n.a.

[a] The number within each cell is the number of respondents from the 1994–96 panel with each combination of answers to what their economic welfare is on a nine-rung ladder. (Rungs 7–9 were aggregated because of small number of responses.)

TABLE 3
CHANGES IN SUBJECTIVE ECONOMIC WELFARE VERSUS REAL INCOME

Change in subjective economic welfare 1994–96	Number of responses (%)	Mean growth rate of income relative to the poverty line (%)	Falling income relative to the poverty line (%)					Rising income relative to the poverty line (%)					
			<−80+	−80–60	−60–40	−40–20	−20–0	0–20	20–40	40–60	60–80	80–100	100+
Fall by 3 or more rungs	381 (7.86)	−21.91	87	56	60	27	52	19	14	17	9	6	34
Fall by 2 rungs	488 (10.86)	−2.91	91	79	71	37	66	29	26	19	12	9	49
Fall by 1 rung	884 (18.23)	−1.91	147	129	148	66	121	61	50	38	27	10	87
No change	1307 (26.95)	15.46	167	169	212	92	176	125	79	64	42	21	160
Increase by 1 rung	953 (19.65)	22.95	104	115	107	85	134	98	74	40	32	34	130
Increase by 2 rungs	517 (10.66)	49.52	57	52	64	34	57	57	37	25	25	16	93
Increase by 3 of more rungs	320 (6.60)	73.42	42	31	39	15	32	37	23	10	13	13	65

These data suggest considerable transient subjective poverty. Of the 1602 adults who said in 1994 that they were on the poorest two rungs, 757 put themselves on the third or higher rung in 1996; by this definition, 47% of the subjectively poor escaped poverty within two years. However, they were all replaced—of the 1763 respondents on the poorest two rungs in 1996, 918 had been on the third or a higher rung two years earlier.

The panel nature of the RLMS allows us to see how answers to the ELQ changed over time for the same people, and to compare this with to the growth rates of income relative to the poverty line. As is evident in Table 2, there is considerable transient subjective poverty. There is a similar degree of transient poverty when assessed by incomes relative to the poverty line: 43% of responding adults who lived in households with an income below the poverty line in 1994 had escaped poverty in 1996, while 70% of the poor in 1996 had not been poor two years earlier.[15]

Are the changes in ladder positions correlated with changes in household incomes? Table 3 looks at the relationship between changes in ladder positions and income growth. We compare the answers to the ELQ for 1996 with those given by the same individuals in 1994. We find that the average growth rates of real income tend to rise as the gain in ladder rungs rises. However, there is also a high variance in growth rates within each category. There are people reporting a substantial improvement in their subjective welfare among those experiencing the largest income drops, and similarly, there are people reporting a large drop in their subjective welfare among those with large measured gains in income.

Table 3 suggests an association between changes in ladder positions over time and the growth rates in incomes relative to the poverty line. When we construct the contingency table, the Cramer V statistic is 0.0778. ($\chi^2 = 189$, which is significant at the 0.001 level.)[17] While average growth rates are appreciably higher for those who report that their economic welfare has improved, there is clearly a large dispersion in growth rates among those in any given category according to their self-reported ladder positions.

V. A Multivariate Model of Subjective Welfare with Latent Heterogeneity

The discussion in Section II suggests four attributes of subjective welfare data that we want to build into our econometric specification. First, the data come in the form of ordered qualitative variables; we cannot assume that the difference between rungs 1 and 2 of the ladder (say) means the same in terms of welfare as the difference between rungs 2 and 3. Second, there is no good reason for assuming that the perceived minimum utility to reach a given ladder rung is the same across people; different people have different ladders. Third, we can expect that time-invariant but unobserved personality differences jointly influence observed socioeconomic characteristics and reported well-being. And fourth, social reference-group effects can entail that the level of a person's subjective welfare at given personal and household characteristics will vary over time.

Our econometric model aims to incorporate these features. We allow the unobserved continuous utility function to contain an additive individual effect.

We also allow the minimum perceived gains in utility to move up (or down) the ladder to vary across respondents. By taking first differences over time in the utility function and netting out idiosyncratic differences between people in their perceived utility gaps relative to the ladder rungs, we are able to identify the key utility parameters from an ordered probit for the changes in ladder rungs.[17] We can avoid the usual incidental parameters problem in estimating a fixed effects probit (Chamberlain 1980) since the latent continuous variable in our case is the net utility gain. Thus, we can difference out the fixed effect in utility levels and still generate a one-to-one mapping from the latent continuous variable to the observed changes in ladder positions.

To outline the method more formally, let the utility of person i at date t, u_{it}, be a function of a vector of exogenous characteristics x_{it}; this can be interpreted as an indirect utility function. We make the standard assumption that the function is static, in that u_{it-1} does not influence u_{it} given x_{it}. However, we allow the utility function to vary between people at given x_{it}. In particular, we allow for a latent time-invariant individual effect in u_{it}; this is interpretable as an idiosyncratic taste shifter in the utility function arising from psychological differences.

We assume that the utility function is linear in parameters and that it contains an additive error term. This has two components. The first is a time-invariant individual effect, η_i, representing personality traits and any other sources of latent heterogeneity, and time-invariant measurement errors; η_i is allowed to be correlated with x_{it}. The second error component is a normal i.i.d. innovation error, ε_{it}, interpretable as the mood variability effects discussed in the psychology literature, under the assumption that these are orthogonal to x_{it}. The latter assumption is crucial; while we have little practical choice, violations of this assumption arising from correlations between mood effects and observed circumstances cannot be ruled out theoretically, and will bias our estimates.

The utility function is then

$$u_{it} = \beta_t x_{it} + \eta_i + \varepsilon_{it}, \qquad t = 1, 2;\ i = 1, n. \tag{3}$$

Notice that the parameter vector β_t varies over time, reflecting the changes in reference group income, as discussed above. Following standard practice in panel data econometrics, the bias in an OLS estimate of (3) resulting from non-zero correlations between x_{it} and η_i can be eliminated by taking differences over time (or by taking deviations from time means, which is equivalent with two observations in time):

$$\Delta u_{it} = \beta_t \Delta x_{it} + \Delta\beta_t x_{it-1} + \Delta\varepsilon_{it}, \tag{4}$$

where $\Delta u_{it} = u_{i2} - u_{i1}$ is the change in subjective welfare for person i, $\Delta x_{it} = x_{i2} - x_{i1}$ is the vector of changes over time in the explanatory variables, and $\Delta\beta_t = \beta_2 - \beta_1$ is the corresponding vector of changes over time in the parameters. Notice that a variable (such as gender) that does not change over time still has a non-zero parameter in equation (4) if the parameters of the utility function in levels (equation (3)) change over time. Of course, having eliminated η_i, we are unable to identify permanent effects on subjective welfare of time-invariant variables, such as gender.

We cannot estimate (4) directly since we do not observe Δu_{it}. Rather, we know the respondent's positions at each of two dates on a ladder with R rungs. There is not a one-to-one mapping from changes in utility to changes in the ladder rung, since individuals on the same rung can have different initial utility levels. Let $g_{it}(k)$ denote the perceived increment to utility needed for person i to advance k rungs; we call this the 'utility gap function'. For example, if between dates $t-1$ and t utility falls by more than $g_{it}(-2)$, then one will be at least two rungs lower at t than at $t-1$. If Δu_{it}, is found within the interval $[g_{it-1}(-1), g_{it-1}(1)]$, then one will be on the same rung at time t as $t-1$.

Our strategy for identifying the β_1 parameters in (3) is to postulate a specific form for the utility-gap functions. It is unlikely that the utility gaps are the same across individuals. The key assumption is that the utility gaps contain a common additive component across all individuals plus an idiosyncratic component that depends on current characteristics (similarly to the utility function). The specific parametric form we assume is

$$(5) \qquad g_{it}(k) = c(k) + \gamma x_{it} + \upsilon_{it},$$

where $c(k)$ and γ are parameters and υ_{it} is a zero-mean normal error term.[18] While this allows for heterogeneity in utility gaps, it does so in a rather special way. It does not allow for a latent individual effect in utility gaps, just as (3) does not allow for an individual effect in the changes in utility; people with the same observed characteristics agree on the utility gaps, on average. And it assumes that people agree on $g_{it}(k) - g_{it}(r)$ for any $r \neq k$.

These assumptions permit us to estimate an ordered probit on the observed changes in ladder positions where the latent continuous variable is the change in utility net of the idiosyncratic component in the lagged utility gap:

$$(6) \qquad y_{it} \equiv \Delta u_{it} - g_{it-1}(k) + c(k) = \beta_t \Delta x_{it} + (\Delta\beta_t - \gamma)x_{it-1} + \mu_{it},$$

in which the composite error term, $\mu_{it} \equiv \Delta\varepsilon_{it} - \upsilon_{it-1}$, is normally distributed under our assumptions. The reported ladder position will be unchanged if y_{it} is within the interval $[c(-1), c(1)]$, which is equivalent to requiring that Δu_{it} is found in $[g_{it-1}(-1), g_{it-1}(1)]$, given the definition of y_{it} in (6) and the assumed form of the gap function in (5). In a ladder of R rungs, there are $2R-1$ possible rung changes and $2(R-1)$ values of $c(k)$, giving the appropriate (common) cut-offs in the y_{it} dimension. At the extremes, if y_{it} is less than $c(1-R)$ (greater than $c(R-1)$), then person i will have fallen (risen) by the maximum number of rungs $(R-1)$. While $c(k)$ must be strictly increasing in k, we do not require that $c(k)$ be positive (negative) when k is positive (negative).

Thus, an ordered probit for the changes in ladder rungs allows us to estimate the β_t parameters in the utility function and the cut-offs $c(k)$ in utility gaps, up to a scalar (the inverse of the standard deviation of the error term in equation (6)). The estimates will no longer be biased as a result of the presence of the latent individual effects in utility levels.

Nonetheless, there are a number of limitations of our method; three stand out.

1. We deal only with one source of bias, namely the dependence of respondent characteristics on a latent time-invariant psychological effect in subjective

welfare. We do not deal with endogeneity with respect to mood effects or other time-invariant unobservables.
2. As is clear from equation (6), the utility parameters for $t-1$ cannot be distinguished from the parameters of the utility gap functions. This would require the stronger assumptions of either time-invariant utility parameters or fixed (expected) utility gaps.
3. The method is feasible only with two observations in time, since only then is there a natural ordering of the changes in ladder rungs. Since one wants to allow for changing parameters over time, this is not unduly restrictive.

VI. Estimation Results

If we use the log difference of income relative to the poverty line as the sole explanatory variable, then we obtain an ordered probit regression coefficient of 0.0245 with a standard error of 0.004 (a t-ratio of 6.15). This is significant at the 0.0005 level. However, two observations are notable. First, zero growth in income relative to the poverty line was still associated with lower subjective welfare (Table 3). Second, the bulk of the log-likelihood of the qualitative perceptions of welfare is left unexplained by income growth rates; the pseudo R^2 for the OP regression on the growth rate is only 0.0095.[19] There is clearly a lot more to changing perceptions of economic welfare than measured income growth rates.

We consider two sets of additional explanatory variables. The first includes those one would expect to be included in a measure of real income per equivalent single adult. Here we include a comprehensive set of demographic variables, as might be used to form an equivalence scale. The second set of variables is not normally included in a measure of real income, but might well be considered to influence economic welfare independently of income. Here we include a variety of individual characteristics, including age and marital status, and measures of health, education, employment and consumption. We also include geographic dummy variables; these eliminate variance resulting from unexplained locational effects, such as local public goods and reference groups effects. (Recall that variables that do not vary over time naturally appear in the regression since their coefficients may change over time.)

Table 4 gives our estimates for the full sample and the sample restricted to those for whom wages and government transfers account for at least 90% of reported income.[20] (As usual, one cannot directly compare the coefficient estimates for the two columns, given that variance of the intertemporal difference in innovation errors need not be the same.)

The changes in log household income and expenditure (both per capita) are both highly significant, with roughly equal proportionate effects in the full sample, though income has a higher weight in the restricted sample. The expenditure variable could well be picking up income measurement error in the whole sample.

Beyond the per capita normalizations for income and expenditure, there is only weak evidence of demographic effects. People living in households with a higher proportion of female adults tended to say that their ladder rung had improved over time. Women tended to say they are worse off, and especially

TABLE 4

ORDERED PROBITS FOR THE CHANGES IN SUBJECTIVE ECONOMIC WELFARE

	Whole sample		Restricted sample	
	Coefficient	s.e.	Coefficient	s.e.
Δ Log household income per capita	0.102***	0.017	0.122***	0.036
Log of total household income per capita, 1994	−0.021	0.027	−0.045	0.049
Δ Log household expenditure per capita	0.109***	0.023	0.088**	0.037
Log household expenditure per capita, 1994	0.026	0.028	0.008	0.045
Δ individual income (×10,000)	0.064	0.043	0.258**	0.088
Individual income, 1994 (×10,000)	0.017	0.064	0.305**	0.111
Household composition variables				
Δ Log household size	−0.033	0.066	−0.093	0.111
Log of household size, 1994	−0.046	0.051	−0.060	0.080
Δ Proportion of small children	−0.459	0.580	1.098	0.844
Proportion of small children, 1994	0.237	0.178	0.584	0.291
Δ Proportion of big children, 1994	−0.437	0.554	1.031	0.795
Proportion of big children, 1994	0.282**	0.119	0.300**	0.189
Δ Proportion of adult men	−0.707	0.564	0.612	0.816
Proportion of adult men, 1994	−0.088	0.099	0.022	0.151
Δ Proportion of adult women	0.765	0.568	0.841	0.830
Proportion of adult women, 1994	0.233**	0.094	0.328**	0.147
Δ Proportion of pensioners	−0.767	0.0564	0.363	0.823
Proportion of pensioners, 1994		*Reference*		
Geographic dummies				
Territory 1		*Reference*		
Territory 2	−0.250**	0.118	−0.573***	0.180
Territory 3	−0.189	0.126	−0.354**	0.158
Territory 4	−0.131	0.086	−0.400***	0.120
Territory 5	−0.165**	0.078	−0.217**	0.110
Territory 6	−0.134	0.084	−0.236**	0.118
Territory 7	−0.187**	0.081	−0.377**	0.127
Territory 8	0.036	0.079	−0.150	0.114
Territory 9	0.057	0.092	−0.252	0.175
Territory 10	−0.103	0.085	−0.199*	0.107
Territory 11	−0.082	0.091	−0.014	0.138
Territory 12	0.031	0.091	−0.253*	0.150
Territory 13	0.042	0.083	0.063	0.110
Territory 14	0.051	0.096	−0.127	0.154
Individual characteristics				
Age (×100)	0.033	0.573	−0.686	0.911
Age-squared (x1000)	0.005	0.062	0.046	0.099
Female		*Reference*		
Male	0.083**	0.035	0.064	0.056
Single		*Reference*		
Married	−0.072	0.059	0.020	0.089
Divorced	−0.137*	0.078	−0.099	0.120
Widowed	−0.242***	0.077	−0.109	0.117
Has job	−0.131	0.093	−0.088	0.171
Change in employment status				
No change in employment status		*Reference*		
Unemployed both rounds	−0.074	0.147	−0.230	0.267
Unemployed to employed	−0.008	0.084	0.055	0.134
Employed to unemployed	−0.292***	0.074	−0.152	0.123

(*continued*)

TABLE 4 *Continued*

	Whole sample		Restricted sample	
	Coefficient	s.e.	Coefficient	s.e.
Self-assessment of health, 1994				
Very good		Reference		
Good	−0.043	0.129	−0.195	0.200
Normal	0.008	0.128	−0.017	0.198
Bad	0.035	0.133	0.025	0.206
Very bad	0.080	0.206	0.005	0.306
Δ health: to very good	0.211	0.148	0.475**	0.225
Δ health: to very bad	−0.227**	0.109	−0.547**	0.183
Δ health: no change	−0.056**	0.032	−0.038	0.050
Δ health: from very bad	0.106	0.191	0.153	0.281
Education, 1994				
High school	0.067	0.051	0.035	0.079
Technical/vocational	0.061	0.051	0.112	0.078
University		Reference		
Change job over the period	−0.021	0.053	−0.084	0.079
Occupation, 1994				
Officials managers	0.066	0.173	0.126	0.240
Professionals	0.022	0.066	−0.114	0.104
Technicians and assistant profession	−0.006	0.066	−0.070	0.104
Clerks	0.062	0.082	−0.072	0.128
Service, shop, market worker	−0.127	0.081	−0.274**	0.130
Skilled agricultural and fishery	0.377**	0.203	0.773**	0.373
Craft and related work	−0.098	0.063	−0.335***	0.099
Plant machinery operation assembly	−0.052	0.063	−0.123	0.100
Manual labour	0.113*	0.069	0.136	0.112
Armed forces	−0.416	0.228	−0.078	0.327
Common cut-offs in utility gaps				
$c(-6)$	−3.499	0.360	−3.969	0.575
$c(-5)$	−2.835	0.329	−3.421	0.550
$c(-4)$	−2.180	0.324	−2.741	0.543
$c(-3)$	−1.679	0.323	−2.308	0.541
$c(-2)$	−1.160	0.322	−1.786	0.541
$c(-1)$	−0.575	0.322	−1.199	0.540
$c(1)$	0.139	0.322	−0.439	0.540
$c(2)$	0.757	0.322	0.207	0.540
$c(3)$	1.342	0.323	0.798	0.540
$c(4)$	1.852	0.324	1.377	0.542
$c(5)$	2.415	0.329	1.872	0.549
$c(6)$	2.965	0.349	2.443	0.580
No. of observations	5003		2012	
Pseudo-R^2	0.152		0.286	

*significant at 10% level; **significant at 5% level; ***significant at 1% level.

widows. However, neither effect is evident in the restricted sample, so these effects may stem from income measurement errors.

The fact that (log) household size is insignificant, controlling for income and expenditure per capita, suggests that there is little or no economy of size in subjective welfare. This appears to be in marked contrast to cross-sectional results on subjective welfare (see the survey by Diener *et al.* 1999). Using the cross-sectional data for 1996, Ravallion and Lokshin (2000) find that self-rated welfare in Russia rises with family size (though income relative to the poverty line falls).

Why do we see such different effects of household size on subjective welfare between the cross-sectional data and our panel data model? One possibility is that there is too little variation in household size over a two-year period to identify the true effect. Another possibility is that intrinsically happier people tend to have larger families. Then the cross-sectional results are biased because of this omitted personality trait; once we switch to the panel data model, the effect vanishes.

Those who became unemployed tended to say that they were worse off, even controlling for the loss of income. Our results from the full sample imply large monetary values of the subjective welfare loss from unemployment. From Table 4 (full sample) we can readily calculate that log household income and expenditure would have to increase by a factor of $0.292/(0.102+0.109) = 1.4$ to compensate for unemployment. (We assume that the savings rate is constant, so both income and expenditure increase by the same proportion. We hold individual income constant, but this variable has such a small coefficient that its effect on the calculation is negligible.)[21] This implies that a large unemployment benefit would be needed to attract a worker out of work. Consider again a worker choosing between staying employed (which is the only source of income for the household) and being unemployed and receiving unemployment benefits. Then the unemployment benefit level would have to be four ($= e^{1.4}$) times higher than the wage to attract the worker out of work.

While our qualitative welfare effect of becoming unemployed is consistent with past results, three caveats emerge. First, our estimate of the welfare cost of unemployment is far lower than the most comparable estimate in the literature, namely Winkelmann and Winkelmann (1998). Their coefficient on a dummy variable for unemployment is seven times their coefficient on log household income. Consider again a worker choosing between staying employed (the only source of income) and being unemployed and receiving unemployment benefits. Then the Winkelmann and Winkelmann regressions imply that the unemployment benefit level would have to be 1096 ($= e^{7}$) times higher than the wage to attract the worker out of work. Our lower estimate may reflect the fact that we have used a narrower definition of welfare. But even so, we find a large welfare loss from unemployment.

Second, our estimation method allows us to test whether there is symmetry in the effect of employment, by separating the welfare impact of going from employed to unemployed from that of changing in the opposite direction. Strikingly, we find that going from unemployed to employed does not raise subjective economic welfare controlling for income (Table 4); if you lose your job, then getting it back does not restore your subjective welfare. This casts doubt on incentive interpretations of the employment effect on subjective welfare—notably its supposed implications for setting unemployment benefits. Unless there is sufficient income gain, the mere availability of a job will not attract the unemployed back to work.

Third, when we restrict the analysis to the subsample for which incomes are more reliably estimated. the unemployment coefficient is halved in size, and its 95% confidence interval now includes zero (Table 4).[22] It might be conjectured that this is because fewer workers in this subsample become unemployed during the period. However, that does not seem to be the reason; the proportion of sampled adults who become unemployed is 3.7% in the restricted sample versus 4.1% in the full sample. A more plausible explanation is that the unemployment variable is picking up time-varying income measurement error, as we discussed in Section II. Notice that, while unemployment drops out in the restricted model, individual income becomes significant. This is consistent with a nontrivial measurement error in individual incomes in the full sample, assuming that the income measurement error is negatively correlated with unemployment.

Turning to the health variables in our model, we find that worse health lowers subjective economic welfare. In the full sample, 'health becoming very bad' has almost the same effect as becoming unemployed. Income and expenditure would have to increase by $0.227/(0.102 + 0.109) = 1.1$ to compensate for health becoming 'very bad'. The effect of a perceived health improvement is not however significant. The health effect is equally significant in the restricted sample, and health improvements also emerge as significant to perceived economic welfare.

We find some significant base-year effects. Income in 1994 does not matter to the change in ladder position from 1994 to 1996 given other variables. None the less, we can reject the null hypothesis that the initial values are jointly zero. (The χ^2 test is significant at the 4% level, in both the full sample and the restricted sample.) However, recall that the interpretation of the parameters on the initial year's characteristics is unclear; significant effects could arise either from changes in utility parameters owing to comparison group effects or from effects on the perceived utility gaps needed to change ladder positions.

VII. Conclusions

A number of potential biases cloud the inferences that can be drawn from standard methods of modelling the socioeconomic determinants of self-rated welfare. There is likely to be an attenuation bias on the estimated income effect owing to the poor income measures typically used in such studies. And there is an endogeneity bias, resulting from a dependence of relevant individual characteristics on the latent personality traits that are known to influence self-rated welfare. The overall direction of bias in (say) the income effect on welfare is unclear on *a priori* grounds. With a high-quality longitudinal survey one can go some way toward dealing with these concerns, though even with very good income measurements (by industry standards) there remain concerns about time-varying measurement errors.

We have proposed an econometric model for subjective–qualitative welfare data that tries to take account of the main properties of such data. By treating the mood effects as normal and i.i.d., and the personality effects as additive and time-invariant, we can use an ordered probit regression to retrieve the mean welfare effects of changes in observed characteristics from reported changes over two survey dates in self-reported positions on a welfare ladder of any length. This allows for endogenous (individual-specific) 'utility gaps' that must be spanned to change one's ladder rung. The latent continuous variable in the ordered probit is the change in utility net of the idiosyncratic component of the perceived utility gap.

On applying this method to panel data on subjective economic welfare for Russia in 1994–96, we find that some of the results widely reported in past studies appear to be robust, but others do not. Household income is a highly significant predictor of self-rated economic welfare. Individual income is a far weaker predictor. Health shocks lower subjective economic welfare at given values of other variables in the model, including incomes. The demographic effects found in cross-sectional studies (notably of household size, at given income per capita) are not robust. The extent of the economy of household size in individual subjective welfare suggested by past work may reflect latent personality effects on the demographic characteristics of the respondent's household.

In keeping with past work, we find evidence of an income-compensated cost to subjective economic welfare of unemployment. In the full sample, it would take a large gain in current income to compensate for becoming unemployed. However, our results point to a number of qualifications on past findings about the welfare effects of unemployment. While becoming unemployed entails a large welfare loss, that loss is not restored when an unemployed person gets a job, except via the income gain. This implies a permanent welfare loss from even transient unemployment at given income. It also suggests that high unemployment benefits do not attract people out of work, but they may well discourage a return to work. Also, the unemployment effect is not robust to restricting the sample to those for whom incomes are almost certainly better measured. There could well be a bias in past estimates of the (income-compensated) welfare effect of unemployment, arising from the structure of income measurement errors.

It is plausible that becoming unemployed or sick generates a welfare loss, even if there is full replacement of the income loss. People naturally become less happy with their lives when such a shock occurs. But we doubt that this is being captured in answers to the economic ladder question we have studied. People are telling us that they feel significantly poorer now, given their current income, when they suffer a shock. The more plausible interpretation is that the current welfare loss arises from some combination of lower expected future incomes and more uncertain incomes, the latter arising from greater exposure to uninsured risk outside the formal employment sector.

ACKNOWLEDGMENTS

The financial support of the World Bank's Research Committee (under RPO 681–39) is gratefully acknowledged. The findings, interpretations and conclusions of this paper are those of the authors, and should not be attributed to the World Bank, its executive directors or the countries they represent. For their helpful comments, the authors are grateful to Angus Deaton, Jyotsna Jalan, Dominique van de Walle, the Journal's anonymous referees and seminar participants at the World Bank and participants at a conference in honour of Bernard van Praag at the Tinbergen Institute, Amsterdam.

NOTES

1. Surveys of the literature can be found in Argyle (1987), Diener (1984, 1994), Furnham and Argyle (1998, ch. 11) and Diener *et al.* (1999).
2. The *t*-ratios are 7.70 and 14.95, respectively, with sample sizes of about 5400, although such tests ignore the discrete ordinal nature of the subjective welfare data.
3. Reference group effects have been discussed by (*inter alia*) Runciman (1966), Easterlin (1974, 1995), Frank (1985), van de Stadt *et al.* (1985), Clark and Oswald (1996) and Clark (1999a).

4. Examples include Clark and Oswald (1994), Blanchflower (1996), Blanchflower and Oswald (1997), Theodossiou (1998), Winkelmann and Winkelmann (1998) and Clark (1999a).
5. In an earlier paper, we also used panel data to study the determinants of subjective welfare in Russia, but we did not allow for individual effects arising from latent psychological factors (Ravallion and Lokshin, 2000). Other papers using panel data in studying subjective welfare include Clark (1999a, b), though he does not use the panel nature of the data to eliminate latent individual effects.
6. This follows straightforwardly from Jensen's inequality, and is well known from the literature on measuring inequality (following Atkinson 1970).
7. Indeed, it has been argued that subjective welfare data can provide the extra information needed to calibrate the equivalence scales used in real income comparisons (van Praag 1991; Kapteyn 1994).
8. Some researchers have strongly defended the consistency and stability of the scales used across respondents (Diener *et al.* 1999). However, there does not appear to be agreement on this point. For example, Veenhoven (1996, p. 1) writes that 'the prefix "subjective" means that criteria for judgement may vary from person to person'. It has also been argued that the scale of reference in subjective assessments of health varies systematically with the health status and socioeconomic characteristics of respondents (Groot 2000).
9. We chose personality traits with a weighted mean correlation coefficient (across samples) of 0.30 or higher; the correlation is positive unless noted otherwise.
10. Though this point applies to van de Stadt *et al.* as well, it is particularly clear in the case of the Clark–Oswald test, since their 'comparison group' income is the predicted income of a worker with the same characteristics from a first stage regression. Then its coefficient is just -1 times the coefficient on predicted residuals from the first-stage regression in a Wu–Hausman specification test (Hausman 1978). Clark and Oswald note this possibility and try to address it using income estimated from a different data-set as the comparison group income; this also has a negative and significant coefficient. (The same method of identifying comparison group effects is used by van de Stadt *et al.* 1985.) However, this does not avoid the problem of income endogeneity; indeed, it may be even more affected by the problem, if this alternative test is based on better instrumental variables.
11. The answers are often fitted to the normal distribution function, following van Praag (1968). Seidl (1994) questions the theoretical basis for doing so, and van Praag and Kapteyn (1994) defend it.
12. Using different data, Rose and McAllister (1996) also report subjective assessments of psychological wellbeing that suggest rising dissatisfaction in the 1990s.
13. The sample was interviewed in 1995, though our main analysis confines attention to the 1994 and 1996 rounds.
14. A range of issues related to the sample design and collection of these data are explained in the documents found in the home page of the RLMS, where the data-sets can also be obtained free; see http://www.cpc.unc.edu/projects/rlms/rlms_home.html.
15. These calculations are based on simple headcounts using the panel of adults: 3027 were not poor in either year; 1244 were not poor in 1994 but were in 1996; 411 were not poor in 1996 but were in 1994, and 537 were poor both years. If we use the number of people, we find that 4380 were not poor in both 1994 and 1996; 2086 were not poor in 1994 but were poor in 1996; 674 were not poor in 1996 but were in 1994, and 1010 were poor in both years. The extent of transient poverty is similar.
16. Cramer's V statistic lies between zero and one and is a measure of association between any two categorical variables; for further discussion see Agresti (1984).
17. We know of no previous example of this method. Ordered probits have been used on utility levels but without an allowance for latent heterogeneity (Clark and Oswald 1994). Winkelmann and Winkelmann (1998) use the Chamberlain fixed-effects logit model on subjective welfare data. This requires that the 10-rung ladder (in their case) be collapsed into a two-rung ladder. It also treats the ladder rungs as fixed. Our method does not require these assumptions.
18. Normality implies that it is possible for the utility gap to be negative. One could assume instead a one-sided error term, such that the gaps are strictly positive, though the ordered probit we use below would no longer be appropriate; nor does there appear to be an existing estimator for such a model.
19. We use the normalized Aldrich–Nelson (1985) pseudo-R^2 rather than that of McFadden (1974), which is known to have a sizeable bias downwards for ordered probits with more than three categories (Veall and Zimmerman 1996).
20. The standard errors are corrected for clustering (given that there is typically more than one respondent per household), but this made negligible difference.
21. Individual income is not logged (since it is not always positive). At mean individual income for 1994, the coefficient for the full sample implies an elasticity of 0.01; for the restricted sample, it is 0.05.

22. If we repeat the calculation of the (household and individual) income and expenditure gains needed to compensate for unemployment using the estimates from the restricted sample (at mean individual income), we find that the unemployment benefit level would need to be only 80% higher than the wage to attract the worker out of work.

REFERENCES

AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data*. New York: John Wiley.

ALDRICH, J. and NELSON, F. (1985). *Linear Probability, Logit, and Probit Models*. Beverly Hills, Calif.: Sage.

ARGYLE, M. (1987). *The Psychology of Happiness*. London: Methuen.

ATKINSON, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, **2**, 244–63.

BLANCHFLOWER, D. G. (1996). Youth labor markets in 23 countries: a comparison using micro data. In D. Stern (ed.), *School to Work Policies and Practices in Thirteen Countries*. Cresskill, NJ: Hampton Press.

—— and OSWALD, A. (1997). A study of labor markets and youth unemployment in Eastern Europe. Warwick Economic Research Paper no. 499, Department of Economics, Warwick University.

CHAMBERLAIN, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, **47**, 225–38.

CLARK, A. E. (1999a). Unemployment as a social norm: psychological evidence from panel data. Mimeo, CNRS and LEO-CRESEP, University of Orleans.

—— (1999b). Are wages habit-forming? evidence from micro-data. *Journal of Economic Behavior and Organization*, **39**, 179–200.

—— and OSWALD, A. J. (1994). Unhappiness and unemployment. *Economic Journal*, **104**, 648–59.

—— and —— (1996). Satisfaction and comparison income. *Journal of Public Economics*, **61**, 359–81.

DARITY, W. Jr. and GOLDSMITH, A. H. (1996). Social psychology, unemployment and macroeconomics. *Journal of Economic Perspectives*, **10**, 121–40.

DE NEVE, K. M. and COOPER, H. (1999). The happy personality: a meta analysis of 137 personality traits of subjective wellbeing. *Psychological Bulletin*, **125**, 197–229.

DIENER, E. (1984). Subjective wellbeing. *Psychological Bulletin*, **93**, 542–75.

—— (1994). Assessing subjective wellbeing: progress and opportunities. *Social Indicators Research*, **31**, 103–57.

—— and SUH, E. (1997). Measuring quality of life: economic, social and subjective indicators. *Social Indicators Research*, **40**, 189–216.

——, ——, LUCAS, E. and SMITH, H. L. (1999). Subjective wellbeing: three decades of progress. *Psychological Bulletin*, **125**, 276–302.

EASTERLIN, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David and M. W. Rider (eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*. New York: Academic Press.

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

FRANK, R. H. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.

FREDRICKSON, B. L. and KAHNEMAN, D. (1993). Duration neglect in retrospective evaluation of affective episodes. *Journal of Personality and Social Psychology*, **65**, 45–55.

FURNHAM, A. and ARGYLE, M. (1998). *The Psychology of Money*. London: Routledge.

GROOT, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*, **19**, 403–20.

HAUSMAN, J. (1978). Specification tests in econometrics. *Econometrica*, **46**, 1251–71.

JUDGE, T., MARTOCCHIO, J. J. and THORESEN, C. (1997). Five-factor model of personality and employee absence. *Journal of Applied Psychology*, **82**, 745–55.

KAHNEMAN, D. and VAREY, C. (1991). Notes on the psychology of utility. In J. Elster and J. E. Roemer (eds.), *Interpersonal Comparisons of Wellbeing*. Cambridge: Cambridge University Press.

KAPTEYN, A. (1994). The measurement of household cost functions: revealed preference versus subjective measures. *Journal of Population Economics*, **7**, 333–50

——, Kooreman, P. and Willemse, R. (1988). Some methodological issues in the implementation of subjective poverty definitions. *Journal of Human Resources*, **23**, 222–42.

Kivimaki, M., Vahtera, J., Thomson, L., Griffiths, A., Cox, T. and Pentti, J. (1997). Psychosocial factors predicting employee sickness absence during economic decline. *Journal of Applied Psychology*, **82**, 858–72.

Lokshin, M. and Popkin, B. M. (1999). The emerging underclass in the Russian Federation. *Economic Development and Cultural Change*, **47**, 804–29.

Mangahas, M. (1995). Self-rated poverty in the Philippines, 1981–1992. *International Journal of Public Opinion Research*, **7**, 40–55.

McCrae, R. R. and Costa, P. T. (1991). Adding liebe and arbeit: the full five factor model and wellbeing. *Personality and Social Psychology Bulletin*, **17**, 227–32.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (ed.), *Frontiers in Econometrics*. New York: Academic Press.

Oswald, A. J. (1997). Happiness and economic performance. *Economic Journal*, **107**, 1815–31.

Plug, E. (1997). *Leyden Welfare and Beyond*, Tinbergen Institute Research Series. Amsterdam: University of Amsterdam.

Popkin, B., Baturin, A., Mozhina, M. and Mroz, T. with A. Safronova, I. Dmintrichev, E. Glinskaya and M. Lokshin (1995). The Russian Federation subsistence level: the development of regional food basket and other methodological improvements. Chapel Hill, NC: Carolina Population Center.

Pradhan, M. and Ravallion, M. (2000), Measuring poverty using qualitative perceptions of consumption adequacy. *Review of Economics and Statistics*, **82**, 462–71.

Ravallion, M. and Lokshin, M. (2000). Self-rated economic welfare in Russia. Policy Research Working Paper, World Bank, Washington DC.

Rose, R. and McAllister, I. (1996). Is money the measure of welfare in Russia? *Review of Income and Wealth*, **42**, 76–90.

Runciman, W. G. (1966). *Relative Deprivation and Social Justice*. London: Routledge & Kegan Paul.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, **82**, 30–43.

Seidl, C. (1994). How sensible is the Leyden individual welfare function of income? *European Economic Review*, **38**, 1633–59.

Theodossiou, I. (1998). The effects of low pay and unemployment on psychological well-being: a logistic regression approach. *Journal of Health Economics*, **17**, 85–104.

van de Stadt, H., Kapteyn, A. and van de Geer, S. (1985). The relativity of utility: evidence from panel data. *Review of Economics and Statistics*, **67**, 179–87.

van Praag, B. M. S. (1968). *Individual Welfare Functions and Consumer Behavior*. Amsterdam: North-Holland.

—— (1991). Ordinal and cardinal utility: an integration of the two dimensions of the welfare concept. *Journal of Econometrics*, **50**, 69–89.

—— and Kapteyn, A. (1994). How sensible is the Leyden individual welfare function of income? A reply. *European Economic Review*, **38**, 1817–25.

Veall, M. and Zimmerman, K. (1996). Pseudo-R^2 measures for some common limited dependent variable models. *Journal of Economic Surveys*, **10**, 241–59.

Veenhoven, R. (1996). Developments in satisfaction research. *Social Indicators Research*, **37**, 1–46.

—— with J. Ehrhardt, M. Sie Dhian Ho and A. de Vries (1993). *Happiness in Nations: Subjective Perceptions of Life in 56 Nations 1946–1992*. Rotterdam: Erasmus University.

Winkelmann, L. and Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica*, **65**, 1–15.

World Bank (1998). *World Development Report*. New York: Oxford University Press.