

Panel session at CEPREMAP Conference, Paris, March 2012

Directions for future research:
**Using data on subjective well-being
in social policy evaluation**

Martin Ravallion
World Bank

Subjective data are questionable for monitoring long-term changes in welfare

- Implicit standards in scales will clearly evolve over time with the structural changes in the economy.
- Urbanization changes reference groups and scales.
- For example, what it means to be “poor” in China today is a much higher standard of living than in 1980.
 - Thus the real value of China’s national poverty line has been revised upwards to reflect rising standards.
- Economic development changes scales, which leads one to question subjective welfare data for long term monitoring.
- However, there may well be more promising uses ahead in development policy....

A new window on welfare for the purposes of policy analysis?

- Some formulation of the idea of human welfare is at the foundation of almost all policy analysis.
- Standard “objective” data sources provide an incomplete picture of welfare.
- Subjective self-assessments provide valuable new information.
- Also, a lot easier to collect this type of data compared to standard consumption, income or wealth data.

But these new data come with problems too!

- Self-assessed welfare on some scale, or its money metric, is unlikely to be acceptable as a basis for deciding who is more deserving of public assistance.
 - Few governments would ever accept that a neurotic, lazy or depressed person is deserving of (say) favorable tax treatment.
- Subject to confounding effects of idiosyncratic personality traits and differences in the interpretations of scales.
- However, there is scope for using these data to understand the trade-offs people adopt, especially concerning non-market goods, for which prices are absent or market goods for which prices are unreliable.

Future research:

We need more tests of the reliability of subjective data for policy analysis

- Heterogeneity in scales also looms large as a problem in cross-sectional comparisons, or short time periods.
- The core tool for policy analysis: the SW regression.
- Our tests using vignettes in Tajikistan confirm frame-of-reference bias.
- However, the extent of bias in SW regressions is modest.
- Good news for future applications, but more tests are needed.

:

How reliable is subjective recall of welfare in evaluating assigned public policies?

- Much demand from practitioners for “**shoestring methods**” of impact evaluation—sometimes called “**quick and dirty methods**.”
- Probably the thing that practitioners would most like to avoid is the need for baseline data collected prior to the intervention.
- Imagine how much more we could learn about development impact if we did not need baseline data!

Can we trust self-assessed changes in welfare for policy evaluation?

- Long-term recall of welfare changes has been used in a number of studies, but few tests.
- Also, here we are interested in potential biases in the differences in recall between those receiving the policy intervention and the comparison group.
- Unclear on a priori grounds: Possibly the intervention changes the meaning of the scales used in subjective welfare questions.

Test for a poor-area development program in China

- At the end of a 10 year evaluation I designed a subjective recall “report card” on 30 dimensions of welfare. Both treatment and comparison group.
- The subjective recall of the household’s overall standard of living contained only a weak signal of changes in consumption based on contemporaneous surveys.
- It takes large changes in objectively-assessed consumption to make a difference to subjective, qualitative, assessments of how the standard of living has changed. This is clearly a relatively blunt indicator.

There are also signs of significant bias

- Controlling for the actual change in consumption, the recalled improvement in living standards tended to be higher for initially richer households.
- There were clear signs of telescoping in the recall responses, but the bulk of the benefits occurred in the earlier half of the recall period, which was given too little weight by respondents in treatment villages.
- Recall was clearly also affected by many idiosyncratic factors not accountable to consumption.

This test is not encouraging, but more such tests are needed

- This is just one study, and (to my knowledge) the only one to date in the context of policy or program evaluation. Further tests are needed.
- Thankfully, the marginal cost of doing such tests in the context of a full-scale evaluation is not too high.
- **Maybe we can learn more about shoestring methods and when they might be reliable.**