The econometrics of inequality and poverty
Lecture 11: Happiness econometrics
Work in progress, very likely to change a lot

Michel Lubrano

January 2011

Contents

1 Introduction

We have assumed that well being can be measured with a single indicator, either income or consumption. A second dimension, the size of the household was introduced, but just as a mean to measure scale effects and to approximate individual well being when only household income or consumption were observable. Well being is by nature multidimensional. So the health status, level of education for instance enter in the composition of well-being. This multidimensional aspect has motivated the promotion of the Human Development Index as an alternative the more simple notion of GNP per capita when comparing countries. At the microeconomic level, this direction of research is illustrated by the field of multidimensional inequalities and multidimensional stochastic dominance. A central point is to determine how income can enter the utility function and if more income necessarily induces more utility and more well-being.

In this lecture, we introduce a part of the literature called *happiness economics* which proceed by direct questioning about the state of well-being of individuals instead of proceeding indirectly by looking at income data and inferencing about preference using the approach of *revealed preferences*. The first author to make a great use of happiness data was Richard Easterlin (1974). He became famous following the Easterlin paradox: despite an increase in average personal income over time, people do not report an increase in average happiness. This paradox is at the basis for questioning the way income is introduced in utility functions.

Data about reported well-being exist for many countries and over a quite long period of time. The core of the lies in relating first *reported* well being (the answers to a questionnaire) to an unobserved level of wellfare or utility and then that unobserved utility to a set off observed socio-economic variables. The link is made thanks a particular econometric model, the ordered logit or probit model. The empirical results can give some insight on the way economic growth, income inequality and social organisation are perceived by different populations (Europe versus USA for instance) and how an index of well-being can be devised.

2 Data sets and questions

Different types of qualitative data exist or can be collected. They correspond the general wording of *satisfaction data*. They come from interviews and surveys and correspond to happiness, life satisfaction, household budget constraint, minimum income needed. These data attempts at measuring a level of individual well-being.

2.1 Happiness surveys

The US General Society Survey (GSS), started in 1972. For the last thirty years it has been monitoring societal change and the growing complexity of American society. It is a widely source of information and data for teaching and research. It contains standard demographic information and of course attitudinal questions. Many of the core questions have remain unchanged since 1972 to facilitate time trend studies as well as replication of earlier findings. One of the questions is "*Would you say that you are very happy, pretty happy, or not too happy?*

Since 1985, the GSS took part in the International Social Survey Program (ISSP) which covers 47 countries.

Eurobarometer is a series of surveys regularly performed on behalf of the European Commission since 1973. It produces reports of public opinion of certain issues relating to the European Union across the member states. It is a unique tool for discerning public opinions and trends on a wide variety of issues relating to the EU. Its database since 1973 is one of the largest in the world. The University of Mannheim has collected and harmonised these data to obtain the Mannheim Eurobarometer Trend File 1970-2002 which is time series collection of socioeconomic data that can be accessed by the web, after registration: `http://zacat.gesis.org`. The available variables are among other: satisfaction life and happiness for socio-political variables and education, sex, age, income for demographic variables.
`http://zacat.gesis.org/webview/index.jsp?object`
`=http://134.95.45.58:80/obj/fStudy/ZA3521`

In the Eurobarometer, both type of questions are eventually used. Graham (2005) notes that psychologists have a preference for life satisfaction questions.

Blanchflower and Oswald (2004) used in their Appendix B, data over Britain 1975-1986 coming from the Eurobarometers, Cumulative File (ICPSR #9361).

German socio-economic panel: ”How satisfied are you with your life?” The scale for answering is over 0-10. The survey covers over 20 years 14,000 German individuals.

The BHPS contains questions about life satisfaction.

2.2 Equivalence scales

Questions about The revealed preference approach for estimating equivalence scales is based on the estimation of demand systems. Danziger, van der Gaag, Taussig, and Smolensky (1984) have proposed a subjective method which is based on direct questioning about subjective welfare level associated to money income and family composition. In France, a question about how difficult it is to make both ends meet is asked in the *Enquête sur le budget des familles*. Answering this question corresponds in fact to determine indirectly the level of income needed to reach a level for welfare for a given family composition.

2.3 A common goal

All these surveys have a common goal: provide an alternative way of measuring individual welfare. We have answers to a question given by z_i and we try to relate this subjective well-being to a set of socio economic variables

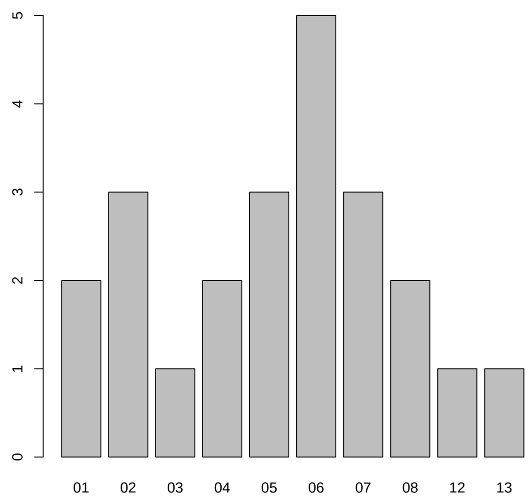
$$z_i = f(y_i, fc_i). \tag{1}$$

For instance y_i money income and fc family composition. Of course this equation might be biased by unobserved heterogeneity, which is most of the time not taken into account, and also by cultural habits.

2.4 A very local survey

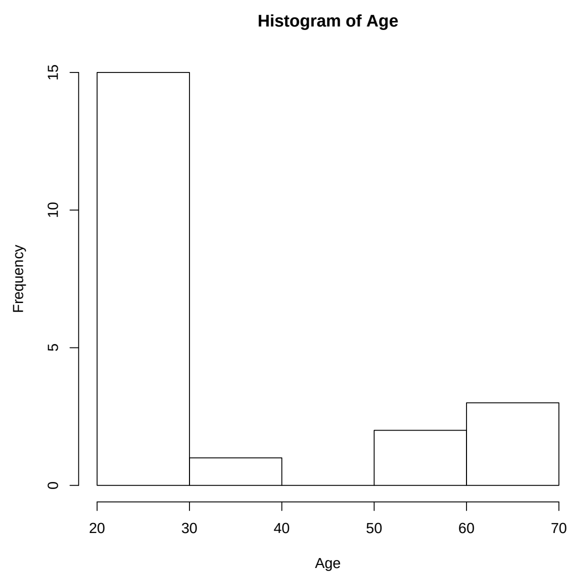
The sample is composed of 30 people, which is very small for such surveys. There were roughly as many women as men (52%). People came in a majority from the center of the town. Marseille is roughly divided in two parts: the north which is poor and the south which is richer. 65% came from the richer part of the city (Figure 1). The age composition was very peculiar because there

Figure 1: Location distribution



are young students, retired people and no people in the forties as shown in Figure 2.

Figure 2: Age distribution



3 Happiness values, optimism

Various types of questions concerning happiness can be formulated. Before taking about the general topic of happiness economics, we have to analyse those questions, see if they give equivalent results and select one. In the Eurobarometer, there are two questions: 1) Life satisfaction, 2) Happiness. The first question concerned the way life has been going on. It is a kind of long term question. The second question ask directly if people were happy. It is more related to the present situation, influenced by recent events. For the little survey made in Marseille, the first question was: *Vous trouvez que pour vous la vie se déroule manière:*

1. Très satisfaisante
2. Plutôt satisfaisante
3. Juste satisfaisante
4. Pas très satisfaisante
5. Pas du tout satisfaisante

There are 5 alternatives, when the Eurobarometer contains only 4 alternatives very satisfied, fairly satisfied, not very sat., not at all sat. The happiness question was: *En ce moment, avez-vous le sentiment d'être:*

1. Très heureux
2. Assez heureux
3. Ni heureux, ni malheureux
4. Pas trop heureux
5. Pas du tout heureux

In the Eurobarometer, there are only 3 alternatives: very happy, pretty happy, not too happy.

We decided to ask a third question concerning the budget constraint which is not contained in the Eurobarometer, but in the BHPS or the French survey Enquete sur le Budget des familles. We want to link happiness data and income data. The budget question was also a subjective question. How do budget difficulties relate to happiness and utility?

Arrivez-vous à "joindre les deux bouts" à la fin du mois ?

1. Très facilement
2. Assez facilement
3. En faisant attention
4. Assez difficilement

	Happiness				
Life	1	2	3	4	5
1	0	0	0	0	0
2	0	2	1	0	0
3	0	1	2	0	1
4	0	0	4	14	1
5	0	0	0	2	2

Table 1: Comparing Life satisfaction and happiness answers

5. Très difficilement

All these questions were proposed with five modalities with a central neutral point.

There are similarities in the answers to these questions. The two questions about life and happiness received very similar, but not identical answers. Table 1 shows that the answers are very well correlated, the answers being on the main diagonal and on both sides of it. Figure 3 shows that people give more dispersed answers for happiness than for life satisfaction. Answers concerning the budget constraint are even more scattered. These answers show the usual optimistic bias. People have a tendency to declare that they are relatively happy. There is a clear modal value at 4 (the scale is 1 for unhappy and 5 for very happy).

4 Econometric model

The dependant variable y_i we have to explain is ordinal on a scale between 1 and 5. Usual regression techniques are inappropriate first because of the discrete nature of y_i and second because the difference between 1 and 2 need not be the same as the difference between say 3 and 4. It is best to express the model introducing a latent variable z_i so that

$$z_i = x_i\beta + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2).$$

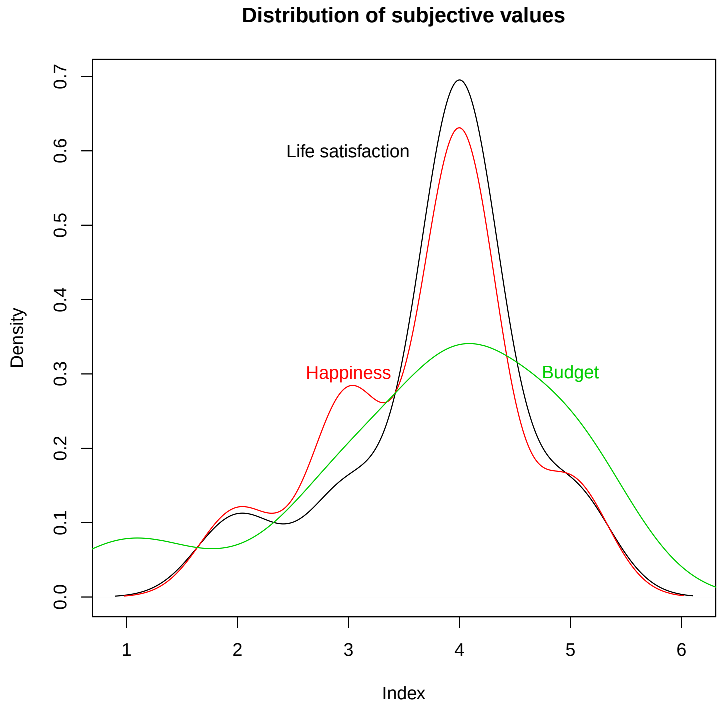
We then map this latent variable to the observed ordinal answers y_i

$$\begin{aligned} y_i = 1 &\Leftrightarrow z_i \leq \tau_1 \\ y_i = 2 &\Leftrightarrow \tau_1 < z_i \leq \tau_2 \\ \ldots \\ y_i = 5 &\Leftrightarrow \tau_4 < z_i \end{aligned}$$

The generic probability to obtain the answer j is given by

$$Prob(y_i = j) = \Phi((\tau_j - x_i\beta)/\sigma) - \Phi((\tau_{j-1} - x_i\beta)/\sigma).$$

Figure 3: Distribution of subjective values



This is the ordered probit model. Albert and Chib (1993) proposed a Bayesian treatment of this model based on the Gibbs sampler. The parameter to estimate are the regression parameter β and the variance σ^2 together with the four thresholds τ_j.

This model is not identified. Any change in σ can be immediately be compensated by a change in τ_j and β. The easiest identification rule is to constrain $\sigma = 1$ and drop the constant term from β. But other rules are possible, see Jackman (2009), chap. 8.

4.1 Bayesian inference for polychotomous regression models

This section is based on Albert and Chib (1993). The model was already introduced in a previous chapter. An individual is surveyed and he has to answer a question with predefined possibilities given on an ordered scale. This was used for the severity of the budget constraint. It can be related to health status which can be *poor, fair, good, excellent*. In this chapter, the survey relates to a degree of well-being. The problem is the same. We want to make inference about the underlying level of welfare which is of course not observed.

We observe an ordered variable Y which can have values $y_i = 1, 2, 3, ...$ for individual i. It corresponds to the ordered answers to the survey. This observed polychotomous variable is a function of the unobserved level of welfare z_i for individual i. The model says that the responder answers 1 if his unobserved level of welfare is below the threshold κ_1, answers 2 if his unobserved level of welfare is between the thresholds κ_1 and κ_2 and so on. For the case of four item question, we have

$$
\begin{array}{rcll}
Y_i & = & 1 & \text{if } z_i \leq \kappa_1 \\
Y_i & = & 2 & \text{if } \kappa_1 \leq z_i \leq \kappa_2 \\
Y_i & = & 3 & \text{if } \kappa_2 \leq z_i \leq \kappa_3 \\
Y_i & = & 4 & \text{if } z_i \geq \kappa_3
\end{array}
$$

This is a simple extension of the ordinary probit model where the answer has got only two modalities 0 or 1, so that $J = 2$.

The unobserved level of welfare z_i is supposed to be determined by a series of observed exogenous variables using a linear regression model of the form

$$z_i = x_i'\beta + u_i$$

where u_i follows a $\mathrm{N}(0, \sigma^2)$ normal distribution. The variance will be assumed to be 1 for identification reasons. If the z_i were observed, we would have an ordinary regression model. But, we observe only the y_i. As $y_i = j$ if $\kappa_{j-1} < zi_i \leq \kappa_j$, the probability that $y_i = j$ is given by

$$\mathrm{Pr}[y_i = j] = p_{ij} = F(\kappa_j - x_i'\beta) - F(\kappa_{j-1} - x_i'\beta)$$

where $F(.)$ is the cumulative distribution of u_i. So the likelihood function is given by

$$L(\beta, \kappa; y) = \prod_{i=1}^{n} \sum_{j=1}^{J} \mathbb{I}(y_i = j)\, p_{ij}.$$

For identifications reasons, it is necessary to impose one restriction on the bin boundaries which is usually $\kappa_1 = 0$ and the restriction $\sigma^2 = 1$. So the model has got two identification restrictions. Note that is we exclude the constant term from β, we could relax the restriction $\kappa_1 = 0$. The parameter to estimate are the β and $J - 1$ threshold parameters in κ.

In a classical approach the log-likelihood is maximised

$$l(\beta, \kappa; y) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log p_{ij},$$

as explicated in chapter 15.3.1 of Cameron and Trivedi (2005). The marginal effect of of a variable in the probabilities is given by:

$$\frac{\partial p_{ij}}{\partial x_i} = [F'(\kappa_j - x_i'\beta) - F'(\kappa_{j-1} - x_i'\beta)]\beta$$

In a Bayesian framework, inference can be conducted quite simply as soon as we notice that we are in a simple regression model if the z_i were observed. The idea is of course to simulate the z_i using a Gibbs sampler with data augmentation. We have to put a prior on the parameter β and κ, but we lave that question aside for the while to concentrate on the Gibbs sampler following Koop (2003). Let us examine the necessary conditional posterior densities:

- $p(\beta|y, z, \kappa)$ is a Normal density if the prior is either noninformative or normal is used for β.
- $p(z|y, \beta, \kappa)$ is a truncated normal density with

$$p(z_i|y_i = j, \beta, \kappa) \sim N(x_i'\beta, 1)\mathbb{I}(\kappa_{j-1} < z_i \leq \kappa_j).$$

- The conditional posterior density of the κ_j is less intuitive. In the case of the Probit model, we simply have $\kappa_1 = 0$. Here, we must draw value of κ_j when $J > 2$. Following intuitive arguments, Koop (2003), chapter 9, explains why this conditional posterior density has to be uniform. We have

$$\kappa_j \sim U(\bar{\kappa}_{j-1}, \bar{\kappa}_{j+1}).$$

Suppose that we have a starting value for the z_i. We start by drawing the β from N$(\hat{\beta}, (X'X)^{-1})$ if we have chosen a non-informative prior. $\hat{\beta} = (X'X)^{-1}X'z$. These formulae come the Bayesian inference in the linear regression model with known variance.

We must have a starting value for he κ_j. We get draws for the z using a truncated normal density with mean $x_i'\beta$ and variance 1.0. The truncation boundaries are given by the κ_{j-1}, κ_j, j being given by the observed value of y_i.

Finally, we draw sequentially the $J-2$ values of κ, starting from κ_2, till κ_{J-1}. Remember that $\kappa_1 = 0$, that the lower bound is$-\infty$ and the upper bound $+\infty$. The parameters of the uniform density have to be determined sequentially. Given the previous draws, we determine

$$\begin{aligned}
\bar{\kappa}_{j-1} &= \max(\max(z_i|y_i = j), \kappa_{j-1}) \\
\bar{\kappa}_{j+1} &= \min(\min(z_i|y_i = j+1), \kappa_{j+1})
\end{aligned}$$

The process is iterated until convergence.

4.2 Using R

This model can be estimated using `R` and `MCMCoprobit` in the package `MCMCpack` for Bayesian inference and the function `polr` of the package `MASS` for classical inference. Due to the very small sample size, it is not always possible to compute classical standard errors when the Bayesian approach provides more robust results.

4.3 Using the local sample for ordered probit

We have run ordered probit regression to explain the relationship between the answers to the happiness question and the answer to the question about how life has been going and how strong the budget constraint is. The happiness variable has five modalities, but the responders used only four, so there are three estimated constant terms in Table 2. These constant terms corresponds to the cell boundaries of the ordered probit. This model can be estimated in two ways in R. First, the dependent variable has to be transformed into a factor using;

Coefficients	Value	Std. Error	t value
Vie	1.296588	0.4002	3.23955
Budget	-0.006118	0.2166	-0.02825
2—3	2.9378	1.1659	2.5199
3—4	4.4778	1.3884	3.2250
4—5	6.2501	1.4771	4.2312

Table 2: Explaining happiness

Coefficients	Value	Std. Error	t value
Budget	0.5502	0.1987	2.768
2—3	0.5446	0.7288	0.7472
3—4	1.2305	0.7339	1.6765
4—5	3.4863	0.9068	3.8448

Table 3: Life and budget

```
fVie = as.factor(Vie)
fBonheur = as.factor(Bonheur)
fBudget = as.factor(Budget)
```

Then, we have the choice between a Bayesian approach using a Metropolis algorithm or a classical approach.

```
out1 <- MCMCoprobit(fBonheur~Vie+Budget, tune=0.3,mcmc.method = "AC")
summary(out1) plot(out1)

summary(polr(fBonheur ~ Vie + Budget, Hess=T,method = "probit"))
```

We give the results for the classical approach in Table 2. There is a clear and positive relation between *how life has been going* on and *how happy you are*. Blanchflower and Oswald (2004) in their Appendix B made a comparison with the Eurobarometer series.

However, the answer to the budget constraint question had no impact. his is in accordance with some of the results found in the literature.

We can further investigate this relation by running a regression on the *how life* question and the budget question. From Table 3, clearly, a softer budget constraint makes life easier. But the equation shows that two classes could be merged.

Coefficients	Value	Std. Error	t value
Age	-0.192137	0.065772	-2.921
Age2	0.002591	0.001060	2.445
Sex	-1.689788	1.009956	-1.673
2—3	-6.3618	0.0521	-122.0572
3—4	-5.2055	0.6513	-7.9930
4—5	-1.6311	1.1144	-1.4637

Table 4: Life and cycle

5 Utility, happiness and income

By looking at coefficient when there is income, we can put a figure on the cost of a chock like divorce, singleness, unemployment to price these events. See for instance Blanchflower and Oswald (2004).

The Easterlin paradox. Happiness depends on income, but not in a linear way. Once basic needs are satisfied, there are other factors which are more important than money. Individuals are sensitive to changes in income. But once this change has occurred,they get used to it.

Role of low or high expectations.

Study poverty. Below the poverty line and happiness. Use of budget constraint. Who are those with a high constraint on budget. Role of economic insecurity.

Inequality: a chance for upward mobility in rich countries. A disadvantage for the poor who have no chance of an upward mobility. Threat of downward mobility is causing social unrest. In Europe, inequality is said to have a low impact on happiness.

6 Life cycle

The dependent variable is *how life is going on*. The life cycle tries to see if there is an age effect on happiness. In Table 4, there is a life cycle effect as happiness decreases with age and then increases. The effect is well marked. We found also the fact that women on average are more happy than men. But this effect is not well marked here.

A graph of the life-cycle effect.

7 Happiness and personal characteristics

Sex, children, married, divorced, profession. Sex, divorce, marriage, unemployment

Insecurity due to unemployment spells, divorce. Security with retirement and good pensions?

Coefficients	Value	Std. Error	t value
North	0.9599	0.942	1.019
South	0.8817	0.879	1.003
S07	2.6000	1.101	2.362
2—3	-0.5183	0.8342	-0.6213
3—4	0.5622	0.8320	0.6758
4—5	2.5441	0.9447	2.6930

Table 5: Happiness and location

8 Geography

The city of Marseille is roughly divided in two parts: the north of the city is rather poor. But only three districts were represented in the sample : 1rst, 2nd and 3rd. The south of the city is much richer. All districts between the 4th and the 8th were represented. The outskirts of the city will be taken as the reference. We want to test if there is a significant difference between North and South. Table 5 gives the results when the endogenous variable is Happiness. Location does not seem to have an influence. There is no difference between north and south of the city. However, people in the 7th are significantly more happy. Similar results obtain when the endogenous variable is life. We could check this using another model, where districts are exchangeable cells.

9 Education and earnings

10 Budget constraint and scale equivalence

References

ALBERT, J. H., AND S. CHIB (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88(422), 669–679.

BLANCHFLOWER, D. G., AND A. J. OSWALD (2004): "Well-being over time in Britain and the USA," *Journal of Public Economics*, 88(7-8), 1359– 1386.

CAMERON, C. A., AND P. K. TRIVEDI (2005): *Microeconometrics*. Cambridge University Press, New-York.

DANZIGER, S., J. VAN DER GAAG, M. K. TAUSSIG, AND E. SMOLENSKY (1984): "The direct measurement of welfare levels: How much does it cost to make ends meet?," *The Review of Economics and Statistics*, 66(3), 500–504.

EASTERLIN, R. A. (1974): "Does economic growth improve the human lot? Some empirical evidence.," in *Nation and Households in Economic Growth*, ed. by P. A. David, and M. W. Reder. Academic Press, New York.

GRAHAM, C. (2005): "The Economics of Happiness, Insights on globalization from a novel approach," *Worls Economics*, 6(3), 41–55.

JACKMAN, S. (2009): *Bayesian Analysis for the Social Sciences*. Wiley, Chichester, UK.

KOOP, G. (2003): *Bayesian Econometrics*. Wiley, Chichester, England.