cemmap

centre for microdata methods and practice

Semiparametric methods for the measurement of latent attitudes and the estimation of their behavioural consequences

Richard Spady

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP26/07

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

SEMIPARAMETRIC METHODS FOR THE MEASUREMENT OF LATENT ATTITUDES AND THE ESTIMATION OF THEIR BEHAVIOURAL CONSEQUENCES

RICHARD H. SPADY

European University Institute, Florence
Centre for Microdata Methods and Practice, London

Abstract. We model attitudes as latent variables that induce stochastic dominance relations in (item) responses. Observable characteristics that affect attitudes can be incorporated into the analysis to improve the measurement of the attitudes; the measurements are posterior distributions that condition on the responses and characteristics of each respondent. Methods to use these measurements to characterize the relation between attitudes and behaviour are developed and implemented.

1. Introduction.

'Factor analysis' has a long history in the social sciences, and one use of factor analysis has been to measure latent characteristics of individuals. These latent characteristics are supposed to explain a variety of related behaviours, and interest is centered on how the assumption of the existence of a small number of underlying characteristics explains the correlations of the related behaviours. Since the factors or characteristics are not in themselves directly observable, the measurement of the characteristics in particular individuals must to a large extent depend on the same observations as the behaviours or outcomes that are the object of explanation. This poses some special problems.

The purpose of this paper is two–fold. First, it develops a view of factor analysis that eschews inessential assumptions and applies this view to a situation in which the data that provide the basis for measurement are categorical or discrete. Second, it attempts to provide methods for using the product of the measurement exercise to (causally) explain phenomena

Date: June 23, 2007. (This version preliminary and incomplete.) This paper was prepared for the conference 'Measurement Matters' at the Centre for Microdata Methods and Practice, June 28-30, 2007.

Key words and phrases. Latent variables.

I am grateful to Mayssun El–Attar, Roger Klein, Roger Koenker, Byron Shafer, and Konrad Smolinski for useful discussions regarding this work.

that were not incorporated into the measurement process. By this explicit separation we sacrifice some efficiency in estimation in the hope of gaining credibility in explanation.

Because factor analysis does indeed have a long history, it is not possible or desirable to give a comparative account of the methods developed here in that context. But two recent threads of thought deserve special attention.

The first thread is a series of papers whose common co–author is James Heckman (cf. Carneiro, Hansen, and Heckman (2003), Cunha and Heckman (2006), and Cunha, Heckman, and Schennach (2006), and the references therein.) These papers give an account, in increasing generality, of the development of cognitive and noncognitive abilities in individuals, and the role of these abilities in determining important life outcomes.

The second thread is a series of papers by Karim Chalak and Halbert White. These papers have as one aim to give a comprehensive account of the causal analysis of observational data. In particular, the causal scheme of Graph 7a in Chalak and White (2006) is closely related to the scheme pursued here.

The methods developed here are motivated by the problem of explaining the political behaviour of individuals as depending on a small number of essential 'attitudes'. These attitudes are hypothesized to be decisive for political behaviour: individuals with similar attitudes behave similarly, regardless of their other characteristics.

The data consists of responses to attitudinal questions on politics and society and the demographic characteristics and self–reported voting behavior and party identification of the respondents. It is available as a time–series of cross sections, with sample sizes varying between 700 and 4000.

It is hypothesized that there exist two underlying attitudes that explain both the responses to the attitudinal questions (hereafter 'items' and 'item responses') and the political behavior. It is further thought that there are some items whose responses depend solely on one attitude or the other; in contrast, voting behavior depends on both. Item responses are discrete, typically having categorical responses corresponding to "strongly disagree", "disagree", "agree", "strongly agree". The convention for assigning numbers to responses is to assign higher integers to more 'liberal' responses; in practice there is little ambiguity about what constitutes a liberal response.

The problem before us is two–fold. First, can we 'measure' the attitudes that underlie the item responses? In what follows this means: "Can we reliably estimate a probability distribution for the distribution of an individual's attitudes without recourse to his political (i.e. voting and party affiliation) behaviour? Second, can we infer from the outcome from this process that "attitudes cause political behaviour," when we never directly observe attitudes, only their manifestation in item responses?

The remainder of this paper is organized as follows. Section 2 gives an informal account of an item response theory in which the effect of successively higher levels of an attitude is to induce stochastic domination relations in response probabilities. On the basis of the item responses alone, it is possible to infer attitudes in such a context via Bayes' Theorem. Section 3 extends this analysis to the case where the social background and experiences of the individual affect the attitudes, and discusses how to use the corresponding demographic characteristics in inferring individual attitudes. Section 4 discusses the particulars of estimation of item response models and the inference of attitudes of individuals and their systematic variation across social groups; several examples are examined. Section 5 develops methods for estimating the joint distribution of multiple attitude scales. In Section 6 we move on to the second of our concerns, inference of the causal effects of attitudes on voting behaviour. Several examples for the 1984 U.S. Presidential election are presented. Section 7 is a conclusion.

2. Stochastic Dominance and Monotonic Scale Representations.

By "an attitude determines an item response", we mean (at least) that the responses of more liberal respondents (first–order) stochastically dominate those of less liberal respondents. If two respondents have the same attitudes, the *probability distribution* of their responses is the same.

Attitudes are essentially ordinal. If attitude $a_2 > a_1$, and r is the response to a particular item that depends solely on scalar a, then $prob(r \leq k|a_2) \leq prob(r \leq k|a_1)$ or $F(k|a = a_2) \leq F(k|a = a_1)$ and this statement remains true for any monotonic transformation of a. Consequently, we can normalize a to have any distribution we like for any reference population we care to choose (subject to some trivial regularity conditions.) (Multivariate generalizations are entirely natural, even for responses that depend on several attitudes.)

2.1 An example with a picture.

We have the item "Congress should pass laws making abortions more difficult to obtain" with the liberal response being disagreement. The relation between some scalar attitude and the probability of each of four responses could be represented as in the following figure

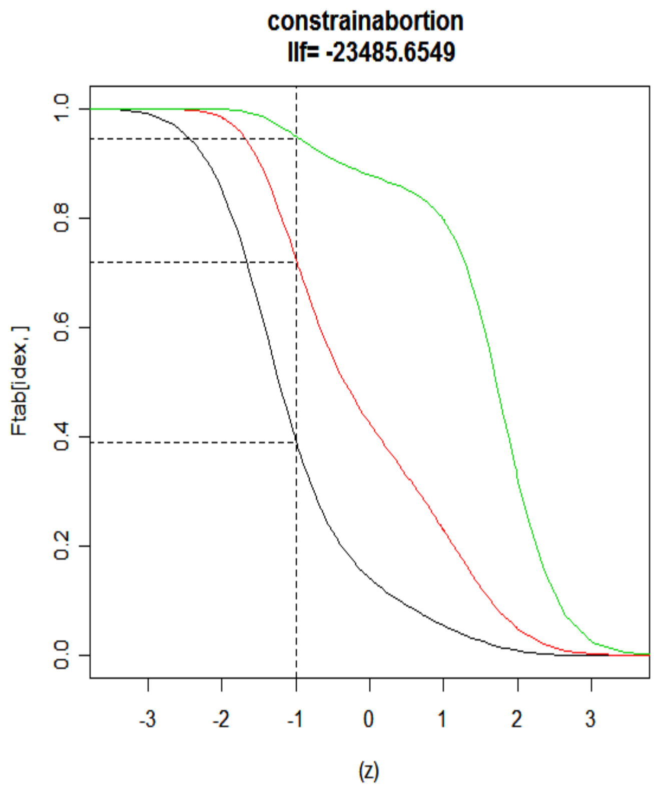


Figure 1.

where the attitude has been scaled on $(-\infty, \infty)$ and the lines represent probabilities of giving a response that is *at least as conservative* as the response to which they correspond. Thus the lowest line is the probability of responding "1", the second line the probability of responding "2" or "1", etc. Stochastic dominance requires that the all the lines slope downwards; the definitions of probability theory require that they do not cross. The figure illustrates the computation of probability distribution and corresponding probabilities at $a = -1$; this and values for other values of a are illustrated in the following two tables:

Response:	**1**	**2 or less**	**3 or less**	**4 or less**
-1	.391	.720	.945	1
-0	.144	.425	.878	1
1	.059	.230	.796	1

Table 1.

To get the response probabilities in each category, we take differences between the columns of Table 1 to obtain:

Response:	**1**	**2**	**3**	**4**
Attitude				
-1	.391	.329	.225	.055
-0	.144	.281	.453	.122
1	.059	.171	.566	.204

Table 2.

2.2 Multiple items.

When there are multiple items, each possible combination of item responses constitutes a 'cell' such as those in the following table:

Abortion→	1	2	3	4	Total
Military↓					
1	.1125	.0597	.0671	.0385	.2778
2	.0360	.0267	.0435	.0165	.1227
3	.0364	.0308	.0491	.0158	.1321
4	.1302	.0991	.1510	.0870	.4674
Total	.3151	.2163	.3108	.1579	

The probabilities in the table are from a sample of 3,829 responses. It turns out that they can be exactly reproduced by two 'box models' such as that in Figure 1, one for each item, with the probability of the joint response $\{i, j\}$ conditional upon attitude a being the probability of $p(r_1 = i|a) * p(r_2 = j|a)$, which is to say that after conditioning on the scalar attitude a, the two responses are independent. In such a case we will say that the item responses have a *monotonic scale representation.* When multiple items are determined by a single attitude, the probability of a particular response pattern (or cell) conditional upon a is simply the product of the constituent item probabilities. That is, by saying that a collection of items have a 'monotonic scale representation' we assume

$$p(r_1, r_2, ..., r_m|a) = p_1(r_1|a)p_2(r_2|a)...p_m(r_m|a)$$

Consequently the probability of the response $r_1, r_2, ..., r_m$ is given by:

$$p(r_1, r_2, ..., r_m) = \int p_1(r_1|a)p_2(r_2|a)...p_m(r_m|a)f(a)da \tag{1}$$

If the 'true' or 'population' cell probabilities generated by a set of item responses can be exactly reproduced for a given choice of $f(a)$ by suitable choice of the functions $p(r_i|a)$ then a strictly monotonic transformation of a yields $\tilde{f}(\tilde{a})$ and new functions $\tilde{p}(r_i|\tilde{a})$ that also exactly reproduce the population cell probabilities. Thus the choice of $f(a)$ is largely if not entirely a matter of convenience.

Since the conditions for a collection of items to have a monotonic scale representation refer to *population* cell probabilities, the fact that *sample* cell frequencies cannot be reproduced by a monotonic scale representation does not imply that the items are not 'scalable.' In particular, in most applications, a very large number of cells, (corresponding to unlikely response patterns), will be empty. Zero probabilities cannot be reproduced easily by any reasonable item response theory, nor should they be.

All of the foregoing has an interpretation for multidimensional a, though our applications will largely estimate item response functions for items that depend on a single attitude.

Implicit in the foregoing is the view that factors should generate relations of stochastic dominance, including possibly higher–order stochastic dominance, and that e.g. questions of factor independence or orthogonality, linearity of responses, maximization of explained variance, etc., are secondary or irrelevant.

2.3 Inference of scale position based solely on item responses.

By Bayes' Theorem:

$$f(a|r) = \frac{f(a,r)}{p(r)} = \frac{p(r|a)f(a)}{p(r)} \tag{2}$$

If we estimate equation (1), we have estimates of the components of the last expression in equation (2): the denominator is the integral of the numerator. Since $f(a)$ can be chosen for convenience if $p(r|a)$ is estimated 'flexibly', we can, for example, set it to be $U[0,1]$ or $N(0,1)$. This already provides us a 'measurement' of a (as a distribution posterior to observing the response), and, with the information that is available to this point, it is hard to see how any further measurement of a can be made.

It is important to keep in mind that when $p(r|a)$ and $p(r)$ are estimated, not known, that the posterior distribution obtained by substituting these estimates into equation (2) is itself an estimate.

3. Attitude Measurement in the Presence of Demographic Characteristics

Estimates of the attitude position of an individual can possibly be made more precise when in addition to item responses there are data on individual characteristics that affect the attitude underlying the response. Let W be a vector of such variables and write

$$p(r_1, r_2, ..., r_m|W) = \int p_1(r_1|a, W)p_2(r_2|a, W)...p_m(r_m|a, W)f(a|W)da \quad (3a)$$

$$= \int p_1(r_1|a)p_2(r_2|a)...p_m(r_m|a)f(a|W)da, \quad (3b)$$

where the first equality follows from definition and the second from an assumption that W affects the response solely through its affect on attitude. Clearly various relaxations of this assumption are possible.

We can choose to normalize a by assigning a a distribution for a fixed value of W, say $\bar{W}$, which then, under mild regularity conditions, will determine a complete distribution for a. Alternatively, the normalization can be chosen so that the entire population has a particular distribution for a. (In our application we follow both these approaches: for estimation of equation (3b), we pick $\bar{W}$ and then normalize $f(a|\bar{W})$ to be $N(0, 1)$; for expositional purposes and in estimating causal effects of a we scale it to have uniform marginals for the whole population, based on estimates of the distribution of a for the sample in hand.)

3.1 The specification, estimation, and interpretation of $f(a|W)$.

If we specify $f(a|\bar{W})$ to be $N(0, 1)$ then this implies both particular representations of $p(r_1|a), p(r_2|a), ..., p(r_m|a)$ and a particular distribution for a in the population. If we choose to model $f(a|W)$ parametrically–a natural (or at least convenient) choice being $N(\mu(W), \Sigma(W))$, then we are—for the first time in the exposition—making a restrictive assumption. One avenue for relaxation of such an assumption, not to be pursued here, would be to estimate $f(a|W)$ with an explicit sieve strategy once $f(a|\bar{W})$ has been chosen.

In the examples to follow we model $f(a|W)$ as $N(W\delta, \Sigma(W))$, with $\Sigma(W)$ having unit diagonal and $\bar{W}$ defined so $\bar{W} = 0$; this is convenient and seems not to do much violence to the data, at least as measured by our own subjective inspection. (It is easy to relax the restriction that $\Sigma(W)$ has unit diagonal.)

If equation (3b) is estimated with e.g. a semiparametric or sieve specificaton of the item response functions $p(r_k|a)$ and a normal additive location shift model for $f(a|W)$, then the resulting estimates of $f(a|W)$ are similar to the normal linear regressions of a on W that would be possible were a to be observed. The estimates obtained in this way are advantaged in that, *provided the item response models* $p(r_k|a)$, $k = 1, ..., m$ *have been 'well–specified',* a has been scaled so that $f(a|\bar{W})$ is normal. It is of course an empirical matter whether, if the reference subpopulation $\bar{W}$ is normal, other subpopulations are similarly normal. However, this is a distinctly more flexible approach than simultaneously assuming a specific distribution for the latent attribute *and* a specific ordered response model for the items.

Estimates of $f(a|W)$ inform us of the distribution of the attitude a in various sub-populations; an interpretation that W *causes* a may be reasonable but requires further assumptions.

3.2 Inference of scale position based on item responses and demographic characteristics.

Equation (2) can be repeated with conditioning on W throughout:

$$f(a|r,W) = \frac{f(a,r|W)}{p(r|W)} = \frac{p(r|a,W)f(a|W)}{p(r|W)} = \frac{p(r|a)f(a|W)}{p(r|W)}, \tag{4}$$

where the last equality follows from the *assumption* that $p(r|a,W) = p(r|a)$. Again the integral of the numerator is the denominator, and the components of the last expression are available from ML estimation of equation (3b).

3.3 Inference of scale position based on demographic characteristics only.

Since $f(a|W)$ is directly estimated, we already have an estimate of a respondents' scale position once we know his or her W.

4. Estimation Methods

Our basic strategy is to approach the estimation of an 'item response box' such as Figure 1 by flexibly specifying each line, subject to the constraint that they are monotonically decreasing and do not cross. This looks forward to a strategy of sieve ML estimation: the item response curves are estimated by a suitable sieve; the function $f(a|W)$ can similarly be estimated by a sieve; the resulting likelihood function, as found in equation (3b) is then estimated by maximum likelihood. If a is scalar (or even bivariate), numerical quadrature can be used for the integration without too much trouble.

4.1 Sieve specifications of item response curves that preserve stochastic dominance properties.

One difficulty with this strategy is to enforce the monotonic declining, no crossing properties of the item response curves: clearly estimation will be much easier if these properties can be enforced by parameterization within a sieve family. We solve (or at least address) this problem with the following steps

(1) Transform a to take values on $[0,1]$. Call the transformed variate u.
(2) A distribution function on $[0,1]$ will be monotonic increasing. If $G(u)$ is such a distribution function, then the hazard function $1-G(u)$ will monotonically decreasing (the 'decreasing' requirement is an artifact of the specification of the normalizations that higher values of u mean higher values of the attitude and higher ordered responses are more likely for those with a higher amount of the attitude.) So the problem is to find distribution functions for each item with k responses $G_1(u), ..., G_{k-1}(u)$ such that $G_{k-1}(u) \leq G_{k-2}(u) \leq ... \leq G_1(u)$.
(3) If e.g. $G_2(u) \leq G_1(u)$ then $G_2(u) = h(u)G_1(u)$ for some $h(u)$ with $1 \geq h(u) \geq 0$, with $h(1) = 1$. Now $h(u)$ can itself be a distribution function, but it need not be since it is not required that $h'(u) \geq 0$.[1] Nonetheless we will require $h(u)$ to be a distribution function (discussion below) and from a collection of distribution functions $F_1(u)$, $F_2(u)$,...,$F_{k-1}(u)$ form the sequence:

$$\begin{aligned} G_1(u) &= F_1(u) \\ G_2(u) &= G_1(u)F_2(u) \\ &\dots \\ G_{k-1}(u) &= G_{k-2}(u)F_{k-1}(u) \end{aligned} \tag{5}$$

(4) We specify $F_1(u)$, $F_2(u)$,...,$F_{k-1}(u)$ to be integrals of regular exponential family densities on the unit interval formed by modeling log densities with a polynomial basis,

[1]The relevant bound on $h'(u)$ is given by the simple calculation:

$$\begin{aligned} G_2(u) &= h(u)G_1(u) \\ log(G_2(u)) &= log(h(u)) + log(G_1(u)) \\ \frac{\partial log(G_2(u))}{\partial log(u)} &= \frac{\partial log(h(u))}{\partial log(u)} + \frac{\partial log(G_1(u))}{\partial log(u)} \geq 0, \end{aligned}$$

hence

$$\frac{\partial log(h(u))}{\partial log(u)} \geq -\frac{\partial log(G_1(u))}{\partial log(u)},$$

which can be satisfied for $h'(u) \leq 0$, i.e. $h'(u)$ can be negative but not too negative.

as in Barron and Sheu (1991). That is,

$$F_i(u) = \int_0^u \frac{e^{t_1 x + t_2 x^2 + \ldots t_m x^m}}{M(t)} dx = \int_0^u f_i(x) dx$$

where $M(t) = \int_0^1 e^{t_1 x + t_2 x^2 + \ldots t_m x^m} dx$.

(5) With $G_{k-1}(u) \leq G_{k-2}(u) \leq \ldots \leq G_1(u)$ in hand, use $1 - G_j(u)$ as $prob(r = j|u)$. Transform u back to a, i.e. compute $prob(r = j|a)$, where a is the attitude measured on a $[-\infty, \infty]$ scale, as e.g. $prob(r \leq j|a) = prob(\{r \leq j\}|\{u = \Phi(a)\}) = 1 - G_j(\Phi(a))$, where $\Phi(\cdot)$ is the standard normal distribution function. (The use of $\Phi(\cdot)$ is not essential: any similar monotonic function from $[-\infty, \infty] -> [0, 1]$ will do, since in principle the $F_i()$'s are flexible.)

(6) Use the resulting probabilities to compute

$$p(r_1, r_2, \ldots, r_m|W) = \int p_1(r_1|a) p_2(r_2|a) \ldots p_m(r_m|a) f(a|W) da$$

(equation 3b) for each observation and maximize the log–likelihood.

The six 'steps' are phrased in terms of a scalar attitude a, but there is no apparent obstacle to interpreting a as multivariate and modifying the exposition accordingly. However, the computational burden of numerically integrating over multiple dimensions can be significant. In our application, we identify items that are assumed to be dependent on a single attitude, and thus we need only compute for scalar a for much of what follows (though estimating the joint distribution of the two attitudes will require a bivariate integration.)

The recursion in equation (5) forces the required first–order stochastic dominance and a little more. In particular, it forces the following condition. Divide the population into those who agree (or more, i.e. strongly agree or agree, etc.) with a liberal proposition and those who disagree. Then consider only those who agree. Then system (5) implies that the probability of strongly agreeing within the "agreers" is montonic increasing in a (and also that the probability of strongly disagreeing among the "disagreers" is monotonically decreasing in a.) This seems a natural and desirable property.[2]

[2]Given a collection of distribution functions on the unit interval $\{F_i(),\ i = 1, \ldots, k-1\}$ then system (5) can be written:

$$\begin{aligned} log\ G_1(u) &= log\ F_1(u) \\ log\ G_2(u) &= log\ F_1(u) + log\ F_2(u) \\ &\ldots \\ log\ G_{k-1}(u) &= log\ F_1(u) + \ldots + log\ F_{k-1}(u) \end{aligned}$$

4.2 An example: Estimates of a 'cultural' scale

From a Pew Foundation survey taken in 1987 we have observations of 3,829 individuals' responses to a variety of political and social preference items, and also the respondents' demographic characteristics and self–reported Presidential vote in the 1984 election, among other variables. It is one of a set of similar surveys taken at various points between 1987 and 2002. Since the primary aim of the present discussion is to fix ideas through an example, we give a cursory description of the data.

We suppose the following five items, with responses in the four catgories 'strongly agree', 'agree','disagree', 'strongly agree' form a cultural scale:

(1) women should resume their traditional role in society
(2) which is the best way to assure peace: (a) military strength, (b) diplomacy
(3) police should be allowed to search known drug dealers without a warrant
(4) dangerous books should be banned from public school libraries
(5) school boards should be permitted to fire homosexual teachers

Standardizing Protestant non–evangelical white male high school graduates of middle income and average age to have attitudes that are distributed $N(0,1)$ we obtain the item response curves shown in Figure 2:

which suggest an obvious linear in logs system $log\ G = A\ \ (log\ F)$ which in the $k = 4$ case that characterizes our examples has

$$A = \begin{matrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{matrix},$$

whereas other choices of A would also be suitable. This is a question about sieve construction (since any $\{F, A\}$ structure without unit lower diagonal A could be reexpressed as $\{F', A'\}$ where A' would have the unit lower diagonal structure).

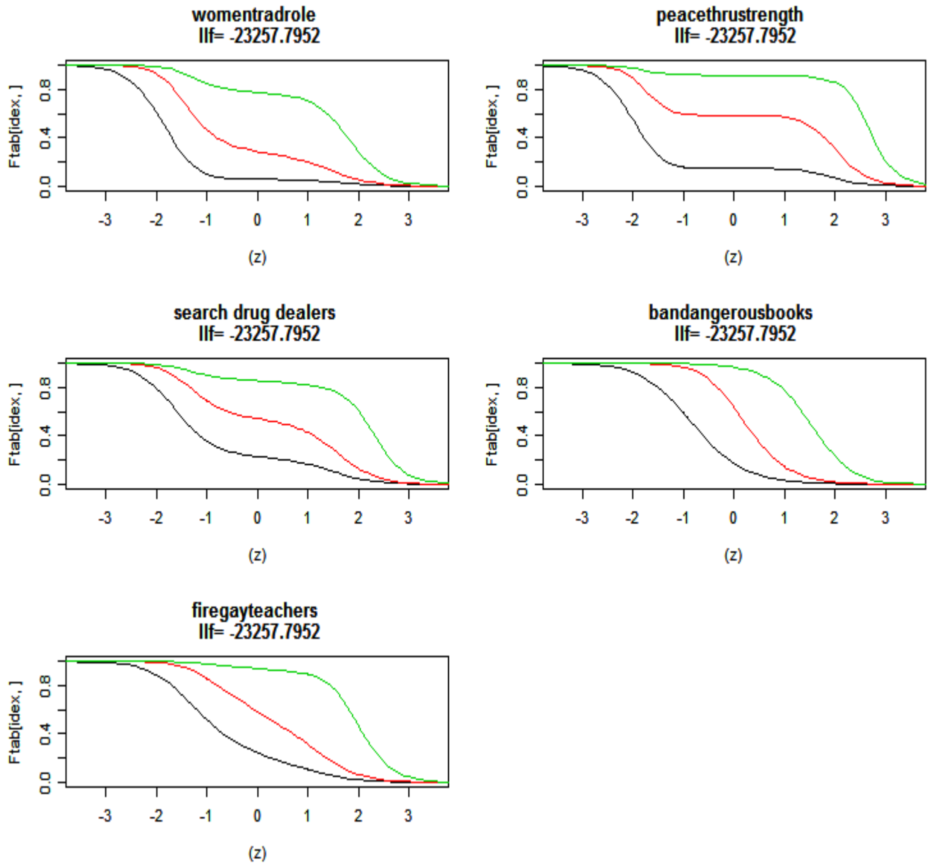


Figure 2. An item response model for a cultural scale.

In addition we obtain the following coefficient estimates for $f(a|W)$:

bmu.c10-14-1987.H	coefficient	OPG s.e.	Hessian s.e.	Robust s.e.
age	-0.0130	0.0015	0.0014	0.0013
agesq.01	0.0081	0.0069	0.0066	0.0067
black	-0.0745	0.0987	0.0909	0.0872
bornagain	-0.3972	0.0524	0.0505	0.0500
blackbornagain	0.0851	0.1353	0.1349	0.1399
rel.catholic	0.1075	0.0516	0.0493	0.0486
rel.nonchr	0.2509	0.0652	0.0626	0.0614
attend.1	-0.3249	0.0562	0.0546	0.0549
attend.2	-0.0998	0.0675	0.0627	0.0618
attend.3	-0.0520	0.0709	0.0636	0.0597
attend.5	0.2111	0.0598	0.0590	0.0592
ed.dropout	-0.1530	0.0575	0.0551	0.0561
ed.somecoll	0.5204	0.0549	0.0530	0.0522
ed.collgrad	1.0193	0.0587	0.0552	0.0552
income.1	-0.0347	0.0540	0.0509	0.0497
income.3	0.1793	0.0526	0.0510	0.0538
income.4	0.2286	0.0737	0.0687	0.0686
income.dk	-0.0856	0.1052	0.0986	0.0947
parent	-0.1260	0.0471	0.0442	0.0429
hispanic	-0.0996	0.0943	0.0831	0.0740
female	0.1807	0.0403	0.0392	0.0424

Table 3.

Notes on the variables appearing in Table 3:

(1) *age, agesq.01*: *age* is age in years minus the sample mean; *agesq.01* is .01*age^2.
(2) *black, hispanic, female*: *hispanic* includes only whites.
(3) *bornagain, blackbornagain: bornagain* comprises whites who self–identify as both Protestant and born–again Christian, plus Mormons; *blackbornagain* is similar, but for blacks.
(4) *rel.catholic and rel.nonchr: rel.nonchr* includes Jews and other non–Christians, including those professing no religion.
(5) *attend:* five categories of frequency of religious service attendance: *attend.1* is once a week or more, *attend.5* is seldom or never.
(6) *ed:* four categories of education; *ed.collgrad* is university degree or more.

(7) *income:* four categories of income; *income.dk* is "don't know" or "refused". *income.1-3* are three categories from low to high each covering about 30% of the population; *income.4* is the highest 10%.
(8) *parent:* the respondent has at least one dependent child.

The 'missing coefficients', *attend.4, ed.hsgrad, income.2* are zero as these correspond to the characteristics of the population that is standardized to be $N(0,1)$.

Since our main interest here is methodological, we will not discuss the results except to say that they are very much in accord with prior expectations.

4.3 Estimating the position of individuals on the cultural scale.

The posterior distribution of a respondent's attitude is, as given in equation (4) above:

$$f(a|r,W) = \frac{f(a,r|W)}{p(r|W)} = \frac{p(r|a,W)f(a|W)}{p(r|W)} = \frac{p(r|a)f(a|W)}{p(r|W)}, \tag{4}$$

which for ease of interpretation can be expressed on a scale that is $U[0,1]$ by transforming a from the mixture of normals implied by the estimate of $f(a||W)$ as given in Table 1 and the empirical distribution of W. It is a consequence of equation (4) that respondents with the same r but different W will generally have different posteriors.

In the figures immediately following we have calculated $f(a|r,W)$—that is, the estimate of the posterior for attitude based on responses and characteristics measured on a semi–normal scale—and $f(u(a)|r,W)$, the same quantity measured on a scale that is $U[0,1]$ for the population—for the first ten respondents having the response patterns {4,3,3,4,4} and {3,3,3,2,3}. The first of these response patterns is very liberal and there are 35 (of 3,829) observations with this response; the second pattern is moderate and coincidentally there are also 35 with this response.

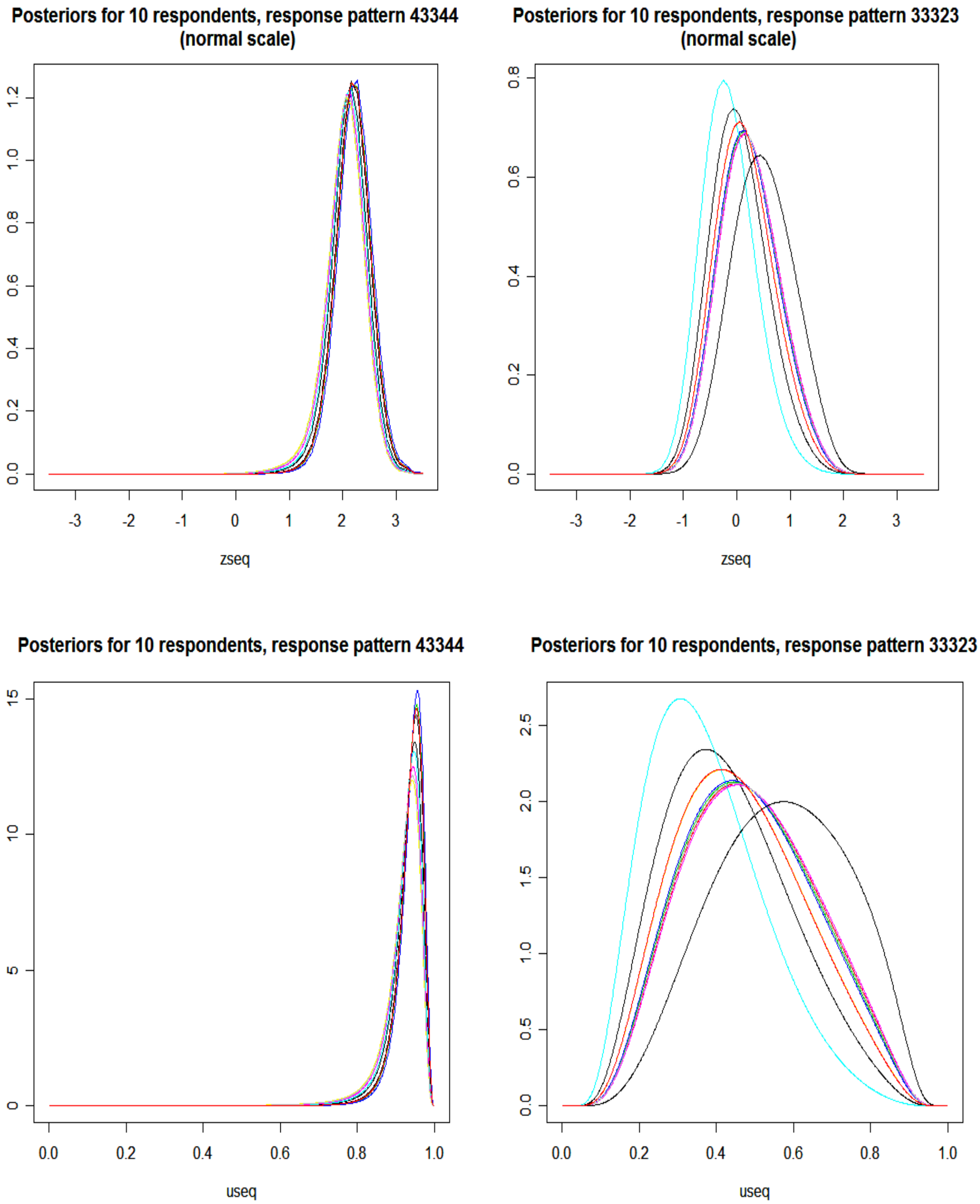

Posteriors for 10 respondents, response pattern 43344 (normal scale)
zseq
Posteriors for 10 respondents, response pattern 33323 (normal scale)
zseq
Posteriors for 10 respondents, response pattern 43344
useq
Posteriors for 10 respondents, response pattern 33323
useq

Figure 3.

These figures show *what could be inferred* about the attitudinal position of these respondents *if our model is correctly specified and without estimation error.* Clearly the precision of this (hypothetical) inference depends both on the response pattern and the demographic or W characteristics of the respondents. Constructing a measure of attitudes as a single number and using this number in subsequent analysis will generally give rise to a measurement error that depends on the attitude of the respondent, the response pattern he manifests, and (perhaps most importantly) his value of W. The figures above show, as it were, the limits of measurement—under the assumptions of correct specification and no estimation error they represent the *correct* measurement of the attitude based on the information being used.

It is important to emphasize that although we cannot know the attitude of an individual because of the limitations of measurement, we can, at least in principle for large samples, know the *distribution* of attitudes for a particular social group. This latter possibility is derived from the consistent estimation of $f(a|W)$.

5. Multiple Attitudes

The estimates in the previous section have assumed a scalar a, that is, that a single attitude underlies the observed response pattern. The reasonableness of this assumption is subject–matter dependent, but in general we will want to assess the impact of multiple attitudes on behaviours. If there is a collection of items with a monotonic scale representation with a single attitude, proceeding as in Section 4 for each such collection produces consistent estimates for each individual attitude, but to analyze responses or outcomes that are dependent on multiple attitudes we will (apparently) require at least a simple model of how attitudes are related, that is, jointly distributed. This section sketches one possible approach.

In order to have a second attitude to help fix ideas, the Appendix shows estimates for an 'economic values' scale akin to the 'cultural values' scale estimated in Section 4. 'Economic values' relate to the desire to assure equitable outcomes in the distribution of material goods among individuals and social groups.

Let us divide the total set of item respones so that the first m are cultural items and responese $m+1, ...M$ are economic. If each class of items has a monotonic scale representation in a single attitude we can write:

$$\begin{aligned} p(r_1, r_2, ...r_M|W) &= \int\int p(r_1, r_2, ...r_M|a_C, a_E) f(a_C, a_E|W) da_C da_E \qquad (6) \\ &= \int\int p(r_1|a_C)p(r_2|a_C)...p(r_m|a_C)p(r_{m+1}|a_E)...p(r_M|a_E) f(a_C, a_E|W) da_C da_E \end{aligned}$$

Now from the scale models already estimated we have $p(r_1|a_C)...p(r_m|a_C)$ and $p(r_{m+1}|a_E)...p(r_M|a_E)$, and we also have already specified $f(a_C, a_E|W)$ as having normal margins. Consequently we can completely specify $f(a_C, a_E|W)$ by simply supplying the missing covariance between a_C and a_E, which can be made to depend on W. A convenient specification is:

$$\rho(W) = \frac{\exp(W\delta) - 1}{\exp(W\delta) + 1}, \qquad (7)$$

where W is augmented from its previous definition by the inclusion of an intercept so that the 'standard' respondent may have correlation between a_C and a_E. More generally we could specify the margins as uniform (i.e. work in terms of what would be called u_C, u_E) and estimate the density corresponding to the copula $F(u_C, u_E)$.

An important aspect of the foregoing is that we have *not* assumed that attitudes or factors are independent. The basic question in our exposition is the dimensionality of a required for the existence of a monotonic scale representation. In principle, if the dimensionality of a has been underspecified model specification tests derived from the ML or sieve ML framework should detect this. If in contrast we treat two item sets as arising from different factors when there is in fact one (so e.g. there is not 'cultural liberalism' and 'economic liberalism' but just 'liberalism') then when estimating the above equations (or their more general analogs) ρ should be estimated to be large (or the copula should high association.)

To supply the raw material for the following section we estimate equation (joint 1) using the parameters derived from the separately estimated C and E models so that only δ from equation (7) is being estimated. The following table gives the estimates of δ (the standard errors are not correct as they do not reflect the two–step nature of the estimation process; they are likely to be underestimates.)

brho.CE-1987.HH	coefficient	OPG s.e.	Hessian s.e.	Robust s.e.
constant	0.0487	0.1553	0.1756	0.2007
age	-0.0033	0.0031	0.0035	0.0039
agesq.01	0.0289	0.0155	0.0170	0.0188
black	-0.6352	0.2215	0.2282	0.2376
bornagain	-0.0499	0.1112	0.1263	0.1450
blackbornagain	0.9594	0.2738	0.2974	0.3269
rel.catholic	0.0300	0.1137	0.1284	0.1461
rel.nonchr	-0.1339	0.1338	0.1489	0.1678
attend.1	-0.3627	0.1156	0.1296	0.1465
attend.2	-0.2958	0.1481	0.1629	0.1807
attend.3	-0.0448	0.1539	0.1698	0.1890
attend.5	0.0087	0.1350	0.1486	0.1653
ed.dropout	-0.1824	0.1232	0.1389	0.1588
ed.somecoll	-0.0191	0.1146	0.1294	0.1482
ed.collgrad	0.5419	0.1206	0.1351	0.1530
income.1	-0.2588	0.1138	0.1273	0.1441
income.3	-0.2409	0.1135	0.1310	0.1534
income.4	-0.0862	0.1575	0.1767	0.2000
income.dk	0.1020	0.2273	0.2333	0.2419
parent	0.1543	0.1020	0.1131	0.1278
hispanic	-0.4034	0.1912	0.2127	0.2390
female	0.1577	0.0857	0.0962	0.1092

Table 4.

The corresponding estimate of the joint density of attitudes with uniform marginals is:

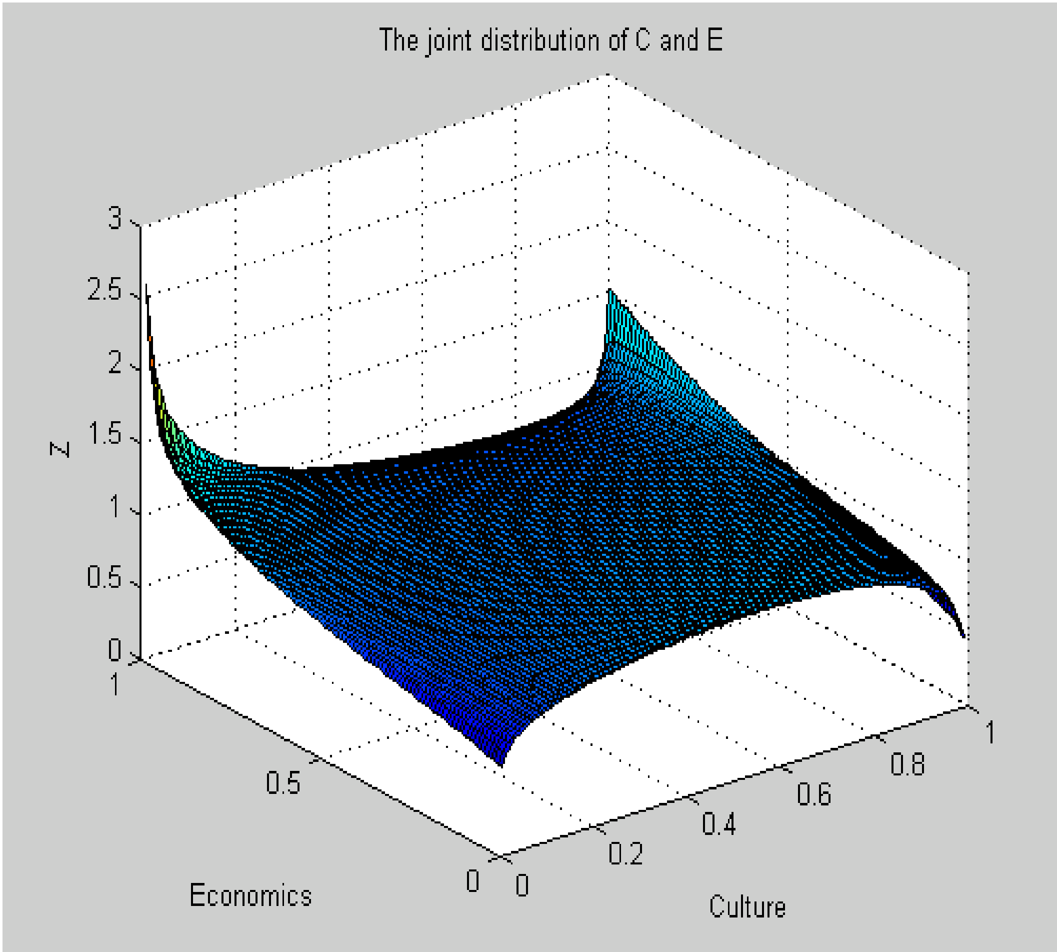


Figure 5.

As is evident from Figure 5, the joint distribution of the two scales in the population is largely 'flat', with some additional density in the corners corresponding to (culturally) conservative–(economically) liberal and liberal-liberal. Correspondingly there are slight "dips" corresponding to economically conservative at both cultural extremes. The evidence of this picture suggests that the two scales are largely independent of each other.

[It is also possible to compute figures similar to Figure xx for individuals, i.e. to show $f(u_C, u_E|W_i, r_i)$.]

6. Inferring the Causal Effects of Attitudes: First Steps

In our application, and of course more generally, there is interest in inferring whether *attitudes* cause particular behaviors. The behaviors of interest in our application is U.S.

Presidential voting and party affiliation. These events are discrete, and some aspects (such as not voting or not affiliating) are not apparently ordered.

Let V be the event of interest which takes categorical values, and use $p(V|a)$ to mean $p(V = V_k|a)$ when V takes generic value V_k and thus no confusion arises. How should we calculate $p(V|a)$ and when should we give it a causal interpretation? We consider several methods.

6.1 Method 1: Brute Force

The most direct approach would be to treat V in a fashion parallel to the item responses, but without imposing monotonicity of the V response in a. Thus building on equation (joint1) :

$$p(V, r_1, r_2, ...r_M|W) =$$

$$\int\int p(V|a_C, a_E)p(r_1|a_C)p(r_2|a_C)...p(r_m|a_C)p(r_{m+1}|a_E)...p(r_M|a_E)f(a_C, a_E|W)da_C da_E$$

Estimation could for example proceed by fixing the already estimated functions (as in Section 5) at their estimated values and specifiying an (unordered) multivariate discrete outcome model, possibly again using a sieve or other semiparametric approach. If V really is only determined by a and not by elements of W or r this will apparently work. If some elements of W, say X, affect V 'directly' then these may be included by replacing $p(V|a_C, a_E)$ with $p(V|X, a_C, a_E)$.

The advantage of this method is that it is transparent; a disadvantage is that it is cumbersome.

6.2 Method 2.

Invoking Bayes' Theorem:

$$p(V|W, a) = \frac{f(V, a|W)}{f(a|W)} = \frac{f(a|V, W)p(V|W)}{f(a|W)} \tag{8}$$

If we assume $p(V|W, a) = p(V|a)$ (again various intermediate versions of this assumption are possible, such as there exist components of W, say X, for which $p(V|W, a) = p(V|X, a) \neq p(V|a)$), we can write this as:

$$p(V|W,a) = p(V|a) = \frac{f(V,a|W)}{f(a|W)} = \frac{f(a|V,W)p(V|W)}{f(a|W)} \tag{9}$$

Inspecting equation (9) we see that *if* $p(V|W,a) = p(V|a)$ *then* the last expression in equation 9 cannot depend on W.

One method for exploiting equation (9) is to estimate the $f(a|V,W)$ component in the same way that $f(a|W)$ can be estimated, by formally treating V as a component of W and estimating using equation (3b). That leaves $p(V|W)$ to be estimated, which can be done by a suitable discrete choice method such as multinomial logit or a semiparametric analog. The denominator of equation (9) is:

$$f(a|W) = \sum_k f(a|V = V_k, W)p(V = V_k|W),$$

i.e. the sum of the numerators corresponding to the different possible V outcomes.

6.3 Method 3.

We can also write:

$$p(V|a) = \frac{f(V,a)}{f(a)} = \frac{f(a|V)p(V)}{f(a)} \tag{10}$$

which would suggest estimated $f(a|V,W)$ as in Method 2 and then (1) estimating $f(a|V)$ by averaging the estimate of $f(a|V,W)$ across each class of V and (2) using the sample analog to estimate $p(V)$. (An estimate of $f(a)$ is obtained by averaging $f(a|V,W)$ across the entire sample.)

An important aspect of both methods 2 and 3 is that both *require* that $f(a|V,W)$ be correctly estimated. That is, if a causes V, then $f(a|V,W) \neq f(a|W)$ and indeed consistent estimates of $f(a|W)$ averaged across V groups will *not* converge to $f(a|V)$. Explicit formulation of $f(a|V,W)$ is required, or to offer another interpretation, the informativeness of V for a must be recognized and exploited.

6.4 Implementation of Method 3.

Method 3 requires an explicit expression for $f(a|V,W)$, which can be obtained by including V among W in the methods used to estimate the components of equation (3b). Rewriting (3b) to make this perspicuous yields:

$$p(r_1, r_2, ..., r_m|V, W) = \int p_1(r_1|a)p_2(r_2|a)...p_m(r_m|a)f(a|V, W)da \qquad (11)$$

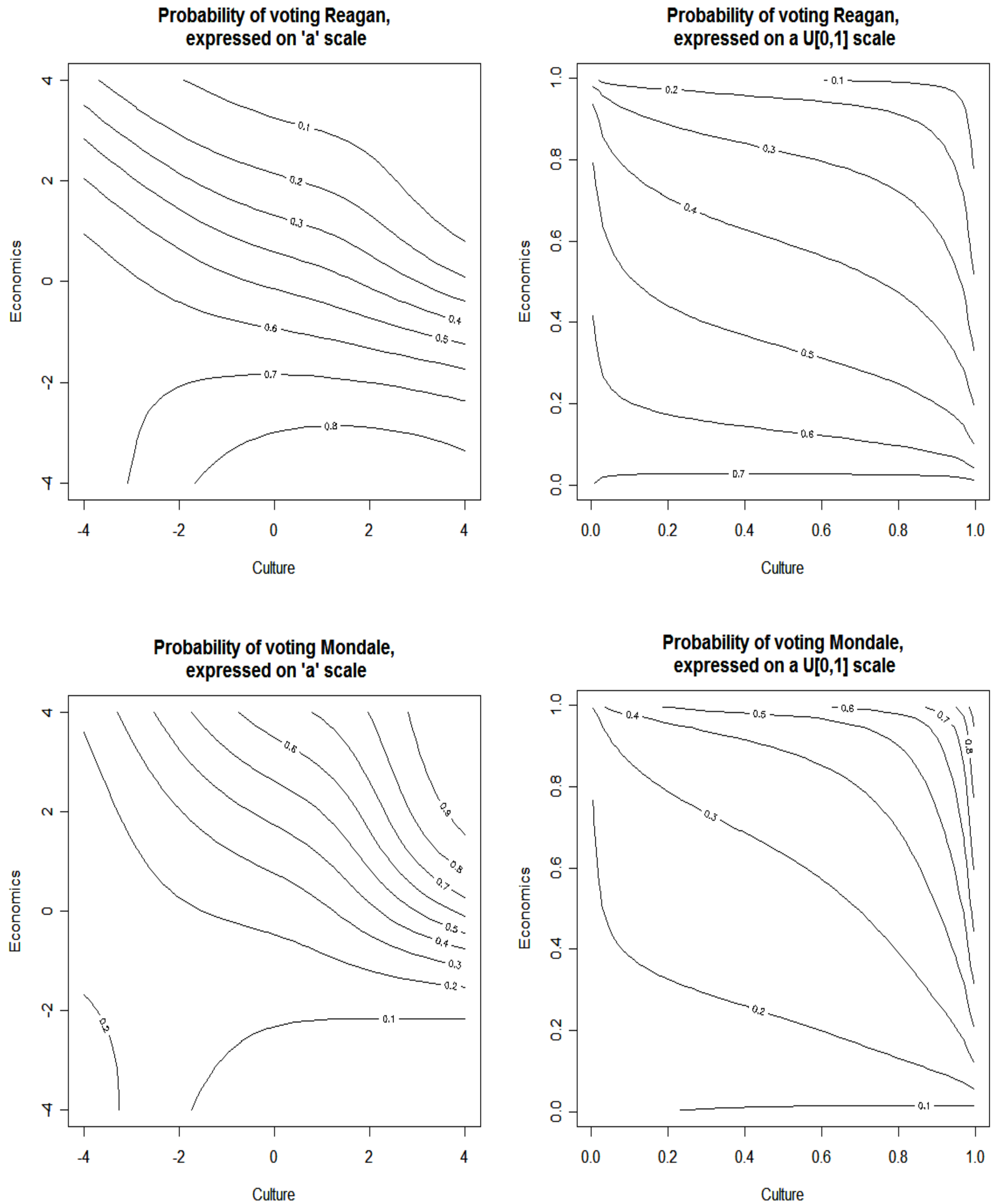
Consequently we can include V (and its interactions with components W as we wish) and re–estimate the model for a cultural scale as given in Table 3 to obtain the following estimates:

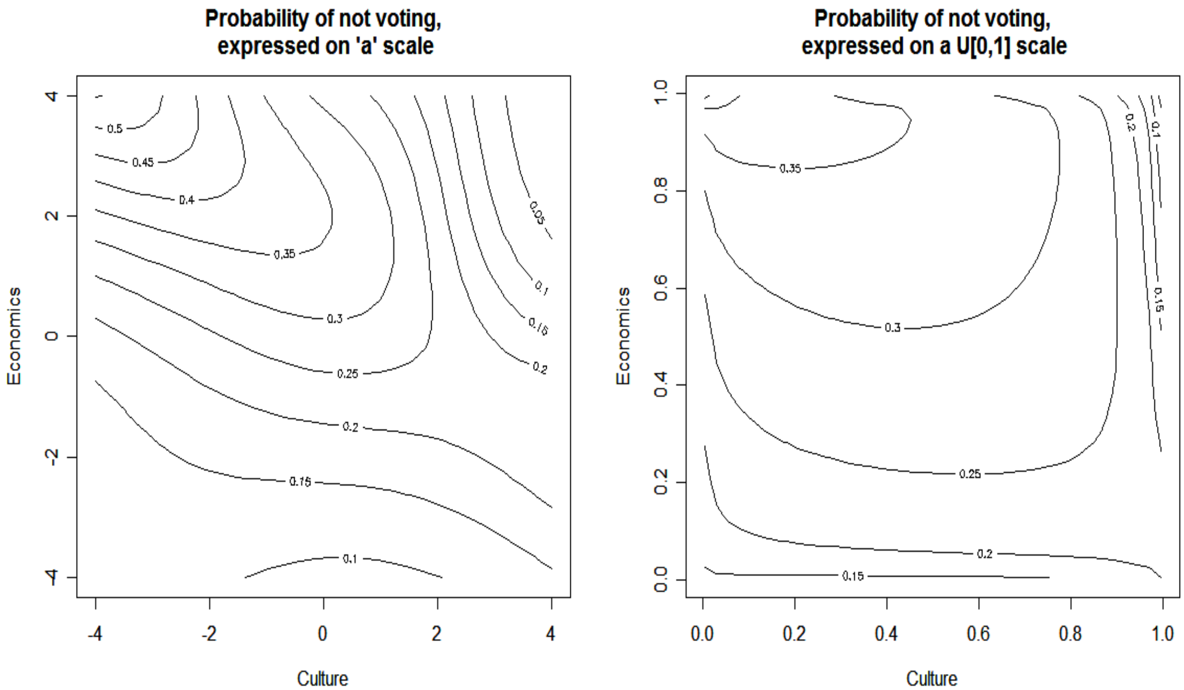
bmu.c10-14-1987V.A	coefficient	OPG s.e.	Hessian s.e.	Robust s.e.
age	-0.0138	0.0016	0.0014	0.0013
agesq.01	0.0087	0.0071	0.0068	0.0067
black	-0.2298	0.1016	0.0931	0.0882
bornagain	-0.3963	0.0534	0.0508	0.0500
blackbornagain	0.0593	0.1359	0.1358	0.1412
rel.catholic	0.0555	0.0529	0.0500	0.0484
rel.nonchr	0.1948	0.0664	0.0632	0.0613
attend.1	-0.3192	0.0573	0.0550	0.0549
attend.2	-0.1141	0.0690	0.0633	0.0622
attend.3	-0.0848	0.0714	0.0642	0.0602
attend.5	0.1855	0.0612	0.0597	0.0593
ed.dropout	-0.1757	0.0592	0.0560	0.0561
ed.somecoll	0.5157	0.0553	0.0534	0.0527
ed.collgrad	0.8278	0.0744	0.0647	0.0605
income.1	-0.0656	0.0550	0.0516	0.0500
income.3	0.1734	0.0538	0.0514	0.0541
income.4	0.2338	0.0753	0.0691	0.0685
income.dk	-0.0775	0.1056	0.0993	0.0957
parent	-0.1273	0.0481	0.0446	0.0429
hispanic	-0.0991	0.0962	0.0838	0.0739
female	0.1437	0.0414	0.0396	0.0421
mondale	0.3368	0.0560	0.0542	0.0548
novote	0.0840	0.0554	0.0512	0.0493
mondale.collgrad	0.4427	0.1114	0.0999	0.0928

Table 5

The results for the economics scale is shown in the Appendix.

After carrying out the joint estimation procedure corresponding to equation (6), the calculations indicated by equation (10) give estimates of $p(V|a)$. These can be expressed either on the a scale (which has the marginal of a distributed as $N(0, 1)$ for a reference population) or on a u scale (which has the marginals uniform on $[0, 1]$ for the population.)





These figures, among other things, show the hazards of displaying information on a 'normal' scale: The figures on the left are graphically pleasing but overemphasize tail behaviour. The right panels more clearly represent the most salient features, namely:

(1) That the probability of voting Reagan is very responsive to economic attitudes and much less so to cultural attitudes;
(2) That the probability of voting Mondale is somewhat more responsive also to cultural attitudes;
(3) That the reconciling factor to these two observations is that the probability of not voting at all is highest among those who are economically liberal but culturally conservative; and finally
(4) That the nonmonotonicities evident in the left side panels are extreme tail behaviours that are barely evident in the right side panels.

7. Conclusion

This paper has developed and implemented factor analytic methods for the measurement of latent variables that are arguably more general than existing methods. The focus is on the representability of the true but unknown population cell frequencies when there are stochastic orderings among responses at different factor levels.

In contexts where measurement is of necessity crude it is probably necessary to explicity reflect the uncertainty of the measurement at the individual level in the subsequent analysis. This is not inconsistent with being able to give explanations and draw conclusion in terms of latent variables, provided sample sizes are large enough. The political example analyzed here is particularly favourable for this sort of analysis, since it is somewhat credible that the behaviour of interest depends almost entirely on the latent attitudes being characterized. Nonetheless, there is no obvious obstacle (other than computational complexity) to employing these factor analytic methods in other contexts.

References

Barron, Andrew R., and Chyong-Hwa Sheu (1991): "Approximation of Density Functions by Sequences of Exponential Families," *Annals of Statistics,* 19, 1347–1369.

Chalak, Karim, and Halbert White (2006): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," UCSD Department of Economics Discussion Paper.

Carneiro, P., K. Hansen, and J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review,* 44, 361–422.

Cunha, Flavio, and James Heckman (2006): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," mimeo, University of Chicago.

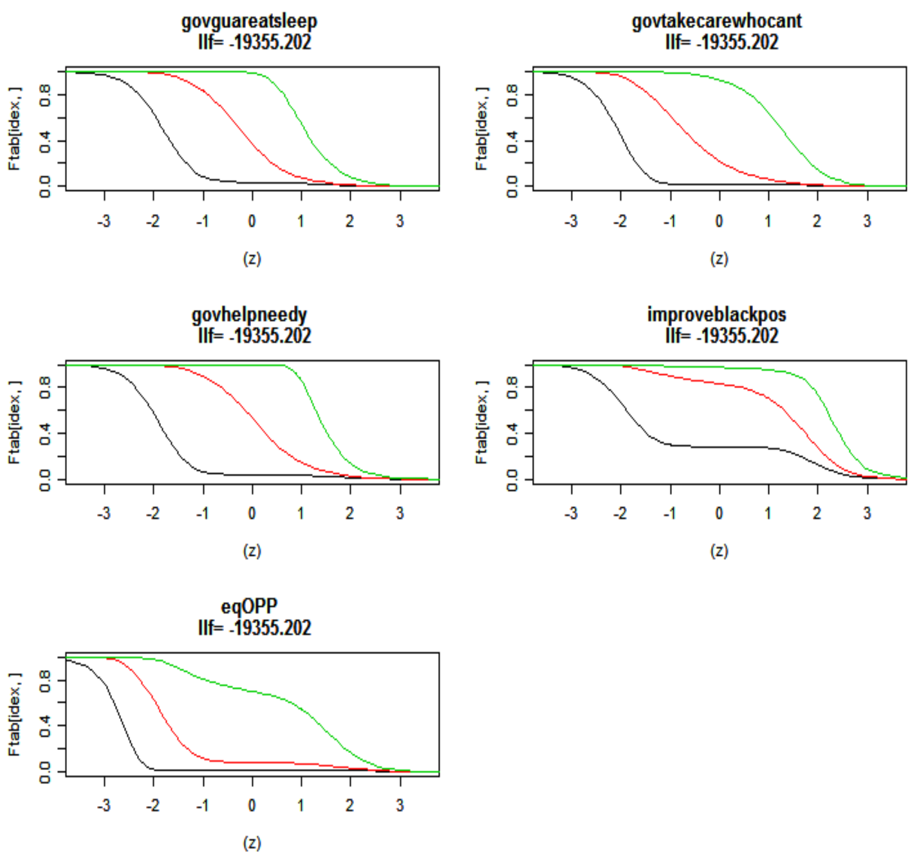
Cunha, Flavio, James Heckman, and Susanne Schennach (2006): "Estimating the Elasticity of Substitution between Early and Late Investments in the Technology of Cognitive and Noncognitive Skill Formation," mimeo, University of Chicago.

White, Halbert and Karim Chalak (2006): "A Unified Framework for Defining and Identifying Causal Effects," UCSD Department of Economics Discussion Paper.

Appendix: An Economics Value Scale

The five items of the economics scale concern the role of the government in assuring an equitable distribution of material resources for individuals and social groups; they are:

(1) The government should guarantee every enough to eat and a place to sleep.
(2) The government should take care of those who can't take care of themselves.
(3) The government should do more to help the needy, even if it means running bigger deficits.
(4) More should be done to improve the position of black people in this country, even if it means giving them preferences.
(5) The government should assure equal opportunity for everyone.



bmu.e12345-1987.H	coefficient	OPG s.e.	Hessian s.e.	Robust s.e.
age	-0.0021	0.0014	0.0013	0.0013
agesq.01	-0.0005	0.0065	0.0060	0.0113
black	0.7339	0.0873	0.0852	0.1014
bornagain	-0.0735	0.0495	0.0462	0.0702
blackbornagain	0.2900	0.1245	0.1269	0.1362
rel.catholic	0.2685	0.0491	0.0467	0.0510
rel.nonchr	0.2388	0.0612	0.0601	0.0610
attend.1	-0.2117	0.0529	0.0499	0.0865
attend.2	-0.0948	0.0629	0.0571	0.1013
attend.3	-0.0034	0.0626	0.0591	0.0868
attend.5	-0.0673	0.0579	0.0533	0.1102
ed.dropout	0.1975	0.0536	0.0502	0.0771
ed.somecoll	-0.1482	0.0506	0.0474	0.0893
ed.collgrad	-0.1421	0.0524	0.0489	0.0914
income.1	0.1811	0.0500	0.0477	0.0674
income.3	-0.1045	0.0505	0.0451	0.1028
income.4	-0.0927	0.0686	0.0638	0.1066
income.dk	-0.1382	0.0931	0.0928	0.1161
parent	0.0077	0.0431	0.0399	0.0856
hispanic	0.3311	0.0810	0.0786	0.0800
female	0.0722	0.0372	0.0354	0.0510

After inclusion of V related variables, as explained in Section 6, this becomes:

bmu.e12345-1987V.A	coefficient	OPG s.e.	Hessian s.e.	Robust s.e.
age	-0.0021	0.0014	0.0013	0.0013
agesq.01	-0.0019	0.0066	0.0064	0.0065
black	0.5734	0.0875	0.0879	0.0935
bornagain	-0.0698	0.0501	0.0476	0.0463
blackbornagain	0.2493	0.1233	0.1274	0.1338
rel.catholic	0.2151	0.0496	0.0473	0.0459
rel.nonchr	0.1799	0.0631	0.0606	0.0591
attend.1	-0.1977	0.0536	0.0523	0.0528
attend.2	-0.0961	0.0633	0.0600	0.0591
attend.3	-0.0323	0.0640	0.0608	0.0597
attend.5	-0.1081	0.0590	0.0569	0.0559
ed.dropout	0.1645	0.0540	0.0524	0.0518
ed.somecoll	-0.1528	0.0513	0.0501	0.0498
ed.collgrad	-0.2541	0.0633	0.0607	0.0593
income.1	0.1447	0.0505	0.0489	0.0486
income.3	-0.1056	0.0509	0.0488	0.0494
income.4	-0.0813	0.0697	0.0660	0.0651
income.dk	-0.1404	0.0958	0.0940	0.0937
parent	0.0072	0.0438	0.0427	0.0429
hispanic	0.3233	0.0811	0.0792	0.0793
female	0.0314	0.0379	0.0365	0.0359
mondale	0.4569	0.0560	0.0515	0.0490
novote	0.2428	0.0489	0.0487	0.0496
mondale.collgrad	0.2226	0.1007	0.0958	0.0925

European University Institute, Florence

E-mail address: Richard.Spady@yahoo.com