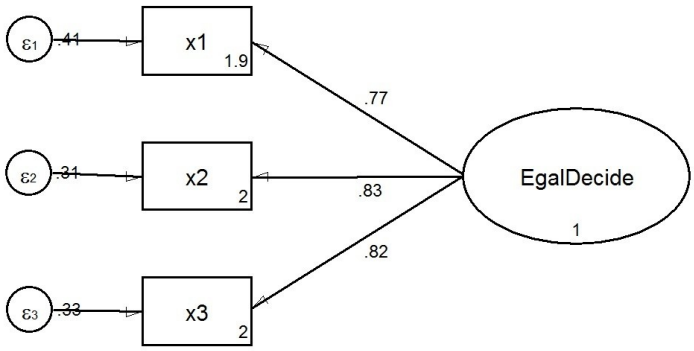
Research jointly supported by the ESRC and DFID

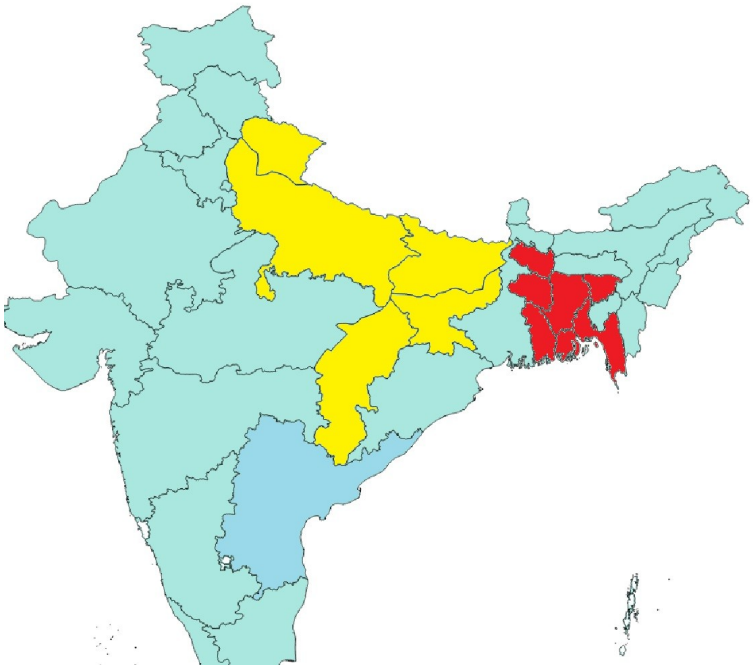
How To Statistically Test Attitudes Over Space: Answers From the Gender Norms Project

In this briefing paper you will find out how to:

- Understand the 'index' or scale created by factor analysis
- Design a factor based on 3+ indicator variables using Confirmatory Factor Analysis (CFA)
- Compare the average of a scale in two countries
- Test the average level of a scale in two or many states
- Find out how gender attitudes differ in India and Bangladesh

Introduction: This paper forms part of a project which investigates the gender aspect of the impact of poverty alleviation schemes in rural India and Bangladesh. Combining a variety of sources, we aim to offer a fresh view on the effect of anti-poverty interventions by focusing on how women's involvement in the labour market is mediated via local gender norms. To interpret our findings we draw on an innovative combination of approaches from a number of disciplinary backgrounds including sociology, economics and social policy.





Research jointly supported by the ESRC and DFID

How To Statistically Test Attitudes Over Space: Answers From the Gender Norms Project

MANCHESTER 1824
The University of Manchester

What is a factor?

A factor is a combination of the original variables. Factors are used to estimate values which are not measured directly in the data. **A factor is a latent construct.** Factors represent the underlying dimensions that summarise or account for the original set of variables (Bowen and Guo, 2012). Well-defined factors can then be used in the way normal variables are. For example we could calculate a regression between respondents having more egalitarian view towards women's work and other variables such as age.

How do we create a factor for gender attitudes?

Factor analysis is useful when we wish to research concepts which can not be measured directly such as norms. There are many types of factor analysis. In our project we use confirmatory factor analysis (CFA) and hence structural equation modelling (SEM). CFA involves exploring *a priori* assumptions about causality. In our case we indentified variables which we suspected to be related to attitudes towards women.

For example when we looked at the norm that men and women can participate equally in the economy we considered four variables:

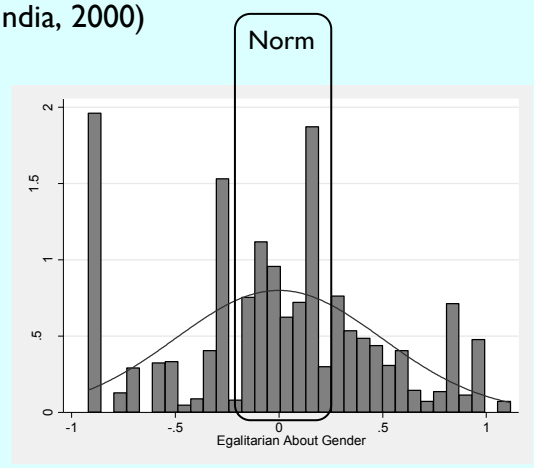
- Who has the final say on health care?
- Who has the final say on making large household purchases?
- Who has the final say on making household purchases for daily needs?
- Who has the final say on visits to family or relatives?

(See Figures and data on pages 6-7 using DHS.)

Science and Gender Norms

The scientific way to study gender is to allow for change, and allow disapproval to be shown. This does not mean to agree with either the 'approval' or 'disapproval' of the social norm, but to be aware that there is a spread of attitudes around the norm itself. In some cases the amount of dispersion in two societies can be compared by measuring the variance of the factor.

Figure 1: Distribution of Attitudes About Women (World Values Survey, India, 2000)



Using the answers to these questions, we can see if there is an underlying norm structure (Figure 1, which uses World Values Survey data). Here we have chosen variables we suspect to be linked to women's ability to make decisions in the economy. Our village-based research also examines attitudes about work roles.

Creation of the Measure of Attitudes and Social Norms For Figure 1

Using the World Values Survey, with N=2000 male and female respondents in the year 2006, we are able to combine five attitude questions into one factor. The five are worded as follows:

- Egal1: (WVS D018, a yes/no variable) "If someone says a child needs a home with both a father and a mother to grow up happily, would you tend to agree or disagree?"
- Egal2: (WVS D022, a yes/no variable) "Do you agree or disagree with the following statement: "Marriage is an out-dated institution."
- Egal3: (WVS C059, a yes/no variable) "Imagine two secretaries, of the same age, doing practically the same job. One finds out that the other earns considerably more than she does. The better paid secretary, however, is quicker, more efficient and more reliable at her job. In your opinion, is it fair or not fair that one secretary is paid more than the other?" (Fair/Not Fair)

And 2 Likert Scale variables:

"For each of the following statements I read out, can you tell me how strongly you agree or disagree with each. Do you strongly agree, agree, disagree, or strongly disagree?"

- Egal4: (WVS D057) "Being a housewife is just as fulfilling as working for pay.
- Egal5: (WVS D059)"On the whole, men make better political leaders than women do."

We used the MPLUS method of categorical variables to combined the Likert scales with the dummy variables shown above. Because some respondents who said 'No' to all three dummy variables also had the lowest response option on two Likert Scales, the resulting index is very lumpy with strong skewness. The scaling placed more egalitarian views about men and women at the righthand side of the scale. Measures of goodness of fit 'combined fit index' (CFI) and the root mean squared error of estimation (RMSEA) performed well for this model. These depend on the sample size and the complexity of the model, or degrees of freedom (Kaplan, 2009).

In Figure 1, the mean of the scale is around zero, as shown by the overlaid 'normal curve' peak. The dispersion of the scale was not set by us, and instead reflects the situation for these variables without making a standardising adjustment. Commonly when doing a CFA one adjusts the standard deviation of the factor to 1, giving a scale generally ranging from about –3 to +3. Because we are creating several indices we wish to compare the distributions without setting them all to the same dispersion *a priori*. Therefore the scale shown has a much narrower range.

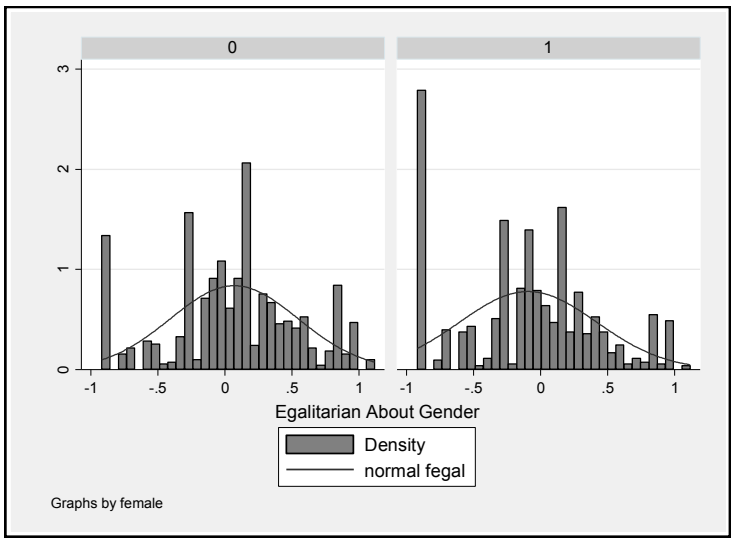


Figure 2: Male and Female Groups Show Small Differences in a Scale

(Note: Left graph is 0 for Males, and right one is 1 for Females; Source is WVS 2006 for India)

Testing Group Differences of Norms

To compare any two social groups, e.g. men vs. women, we can first compare the mean of the factor within each social group.

Then we make an inference after allowing for a 95% confidence interval around one of these means. If the second mean lies within this interval, we could conclude that "in repeated samples of this kind, from this population, there is a 95% chance that Group 2's mean would lie near the Group 1 mean in

the sense of being within this confidence interval." A shorthand test of such a hypothesis is offered by either the t-test or the ANOVA procedure (Field, 2013, Chapter 11).

Table 1: T-test of a Difference of Means By Sex for India Egalitarian Index, 2006

```
 Group    |    Cases    Mean    Std.Err. Std.Dev. [95% Conf. Interval]
---------+--------------------------------------------------------------------
   Males  |     1139    .068     .014     .477    [.041,    .096]
  Females |      861   -.093     .017     .512    [-.127, -.056]
---------+--------------------------------------------------------------------
combined  |     2000   -.001     .011     .498    [-.023 , .021]
---------+--------------------------------------------------------------------
    diff  |             .162     .022             [.118,    .205]
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   7.28
Ho: diff = 0                                     degrees of freedom =   1998

    Ha: diff < 0                 Ha: diff ≠ 0                 Ha: diff > 0
 Pr(T < t) = 1.00         Pr(|T| > |t|) = 0.00          Pr(T > t) = 0.00
```

Source: WVS, 2006, anova results from Stata v13, CFA based on runmplus using MPLUS version 7. See also Cramer, 2003, chapter 10-11, which explains the adjustment made when allowing for age differentials in the Groups for an ANOVA as an alternative to the t-test.

Critical Assessment of T-test

The World Values Survey has sampling problems in taking just 2000 respondents for India. In using a t -test, we are assuming random sampling across the whole country. There is no guarantee that this sample is random. Instead WVS used quota sampling methods. They got clusters of cases in specific geographic areas. The sample **is big enough in absolute terms** to support inference to all-India.

We usually assume that the variance of the factor for Group 1 is the same as the variance of the factor for Group 2. We could make an adjustment to allow for different estimated group variances.

Finally we notice the difference between a *statistically significant* t-test result, and a **substantial difference in attitudes.** Here, the difference in the means is very small but is statistically significant. Figure 2 shows a small leftward shift of the women's attitude distribution relative to that of men. Ironically, the statistical test has minimal stringency. If we require effect sizes, we will need a regression to be done. Otherwise the supposed 'male/female disagreement' over gender norms is misleading. The statistician should not merely *deduce* using inference, but should draw a warranted conclusion overall looking at the substantive size of a difference of the mean. The mean represents the social norm.

Taking the Analysis Further

Using regression or analysis of covariance we can discover whether age and having children are intervening factors which moderate or mediate the apparent relationship of sex with attitude. MacKinnon (2008: Chapter 6) explains how mediation analysis is done.

Our analysis shows that the attitude is relatively less egalitarian among older people than among the youth in India in 2006. Furthermore, an interaction of sex and age is non-significant, so this was not only the case among women—who are commonly assigned key roles in child care—but also among men in this survey.

Research jointly supported by the ESRC and DFID

The University of Manchester

Gender Norms

When we approach issues of stereotypes about gender roles, we place ourselves outside the social norms for a moment. We describe the social norms, and then allow for our own and each other person's opinion – known as their 'attitude' – to lie outside or near the norm. A scale would allow a person to deviate, either in very strongly approving, or disapproving ways, relative to the social norm. The histogram shows the norm about holding egalitarian views as a range of views located in the middle. Those to the right of this zone are strongly tending to disapprove of patriarchal inequality of the genders. Those to the left are strongly tending to approve of patriarchal inequality of the genders.

This complex definition allows for change in the gender norm over time.

Diversity in the society is seen (Figure 1) as holding views further away from the mean.`

There may be important differences in wording from one survey to another. In particular, the wording may be 'about the society' in one survey and 'about one's own views' in another. In the first case, Respondent reports on social norms. In the second case, Respondent reports on personal attitudes. In either case, we can argue, the Latent Factor represents a social norm, or a normal medium point, at its mean. However, the dispersion means a different thing and will tend to respond to different aspects of knowledge: knowledge about diversity in the first case; and willingness to admit deviation from a norm, in the second case.

Further Analysis of the Parameters Is Possible

Secondly we can test whether the *measurement model* is the same in Groups 1 and 2. For instance, is the measure constructed with the same parameters (the factor scores for each measured variable) in rural India and rural Bangladesh? This group test is done using a statistical package such as MPLUS. This software can test all possible models and rank them, allowing us to see whether a model with separate sets of parameters is better than one with a single set of parameters.

A Final Test for Group Differences: Panel Data

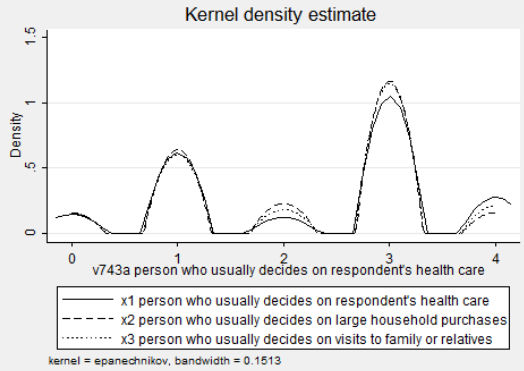
Thirdly, we can test for change over time. It is necessary for a test over time that the sampling and question wordings should be invariant over time. With WVS, which is a multiple-period cross-sectional survey, analysis of change over time may be useful. In MPLUS the group test can be done by pooling the data and defining as Groups the two time periods.

How Many Parameters Are There in a 'Factor'?

Figure 3 shows a measurement model with three manifest variables. The estimate has measurement parameters which tell us exactly how many scale points go onto the new factor measure from each of the underlying variables. (A parameter is an estimated quantity. There are other parameters, e.g. the variance of the factor.) Three possibilities exist, as shown in Figure 4. Consider one more example before interpreting Figure 3 in terms of the number of parameters estimated.

To illustrate, three manifest categorical indicators have been laid out on an ordinal measure in Figure 3.

Figure 3: Women's Views on Women's Control Over Decision Making



Source: Bangladesh DHS, 2011. Primary data analysed with a Likert scale manually constructed for each variable v743a, b and d.We obtain CFI=1.0, or with survey weights on, RMSEA=0.00, with a tight fit.

The methods raise two key issues.

Is Each Manifest Variable a Trace of a Latent Normal Distribution?
Clearly the distributions are not normal as coded. The STATA default method will be to take them ***as if*** there were a latent normal distribution, and estimate one parameter for the slope in regressing each X on F, where X is the manifest variable and F is the new factor. The MPLUS methods more easily allow independent estimates of each of 4 categories (k-1) for the two variables (v-1), ie eight parameters in the measurement model instead of just two.

The key to MPLUS methods is that the covariance matrix is the 'dataset' on which the estimates

Source: Bangladesh DHS, 2011. Primary data analysed with a Likert scale manually constructed for each variable v743a, b and d.

An Index of the Women Making Decisions About Spending In the Household

From such data a factor can usually be constructed.

Gender Norms in the Demographic and Health Surveys

The DHS contains information well suited to a factor for measuring how far a woman is permitted to (or takes upon herself to) make decisions about spending money.

The first element in the factor is whether she is the person who usually decides on her own health care. We code as more egalitarian the people who respond that she 'alone' decides this; as least egalitarian those who say 'husband/partner alone' decide it. In between an ordinal scale is set up with her -joint-with-husband coming second, her-joint-with-someone-else next, another person coming next, and of course the husband alone coming last.

Further manifest variables are available for 2 other types of purchases for Bangladesh 2011 in DHS:

- Who decides on spending for a large household purchase
- Who usually decides on a visit to friends or relatives.

A PhD on the permission granted women to make decisions about their own physical mobility was written recently by Nikhila Menon. We acknowledge that research which used a specific method of factor analysis known as item-response theory, 'IRT', and specifically a Rasch model, which requires 10 or more manifest measures.

are based. This matrix is enlarged if CATEGORICAL measures are introduced. The covariance matrix is then re-estimated using the new latent factor values which give correlations (and covariances) similar to the original correlations. This method gives the user the freedom with a CATEGORICAL subcommand to instruct that X1 and its friends X2 and X3 are multinomial. Each is then granted a latent normal curve of its own (effectively), and each value gets a threshold of its own, excepting the reference values.

Is the DHS Able to Compare Men's With Women's Views on Decision Making?
In the Bangladesh DHS 2011, women have an exhaustive survey, but men form a small subsample and men have only a short survey. Here it is not possible to test the group of men against the group of women.

How Do You Test For Group Differences in STATA?

We can use the analysis of variance command, 'oneway', to get an unadjusted Anova test for all 8 regions of Bangladesh (Table 2). We find statistically significant differences between the regions. Another option is to adjust this test for age of the woman, since age is negatively associated with egalitarian views. After adjustment, the tests still show significant differences by region. Table 2 shows that degree of differentiation by region is small in substantive terms (ie on average). This analysis of variance needs to be supplemented by a table of means.

Again issues crop up: *Do women live where they say they live?* In Bangladesh, a good proportion of people report their residence in one place, but have a de jure (legal) residence elsewhere. This will affect attitudes not only in the measurement model… but over time, through evolution of attitudes and gradual change of rural social norms in particular. *Do the variances differ by region because of diversity of attitudes being greater in the urban areas?* There is evidence of variation being higher in some regions, for which Bartlett's measure is a good starting point. Retroductive thinking then asks: why, and how do we explain, the regional disparity in the diversity of attitudes around the broad national social norm? The Group Test in MPLUS then allows an empirical test of whether the **measurement model should be region-specific**. This will be used in our research, but is not shown here.

Table 2: Bangladesh Regional Differentation of Gender Egalitarian Norms

```
     ANOVA                  Analysis of Variance
    Source                SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between groups       64.6736207        6   10.7789368     15.08     0.0000
 Within groups       12750.1759    17835   .714896324
------------------------------------------------------------------------------
    Total            12814.8496    17841   .718280901

Bartlett's test for equal variances:  chi2(6) =  46.6978  Prob>chi2 = 0.000
```

Source: STATA 13 based on the factor constructed with sem command, subcommands as shown:

```
sem (FEgal -> x1 x2 x3) [pweight = weight], stand latent(FEgal) // use sampling weights
estat gof, stats(all)
predict FEgal, latent(FEgal)
label var FEgal "CFA latent variable representing egalitarian decision making"
```

Figure 4: How The Number of Parameters Differs, Depending on Measurement Levels

v Ordinal Measures, Each Taken with k Separate Categories	v Ordinal Measures Taken as Continuous Measures	v Continuous Measures (e.g. a scale from 0 to 10)
*k-1 parameters * v-1 measures*	*v-1 parameters*	*v-1 parameters*

(note: v is the number of attitude questions on a related topic in the questionnaire. k is the number of categories, thus for example 2 for yes/now, or 5 for a Likert Scale {1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree}.)

We have underlined the three levels of measurement of the manifest variables. For example, if there are five questions on the survey, there would be 16 parameters, or 4 parameters, depending on what assumptions were made. With 16 parameters a more flexible model can achieve a better overall fit. Such models may be more sensitive to groupwise deviation in the measurement model. MPLUS software or manual adjustments are used to create the estimate with 16 parameters. SPSS defaults are to have just 4 parameters.

The Usefulness of Each Factor Component

We recommend using confirmatory factor analysis, so the decision about which manifest variables to include/exclude will rest heavily on theory/ies about reality. Theory is strongly influential in CFA. A useful method of getting a 'modification index' helps show the impact of each X. In combination, retroduction can be used to work out—perhaps with qualitative or historical evidence—whether to leave all variables 'in' or remove one (Olsen and Morgan, 2005). Full models and their variants can be tested as nested models (Bowen and Guo, 2012; Kaplan, 2008). In general, CFA is preferred, but within that school of thought, adjustments can be made by researchers in response to evidence.

Factors Reflecting Background Cultural Factors in India and Bangladesh

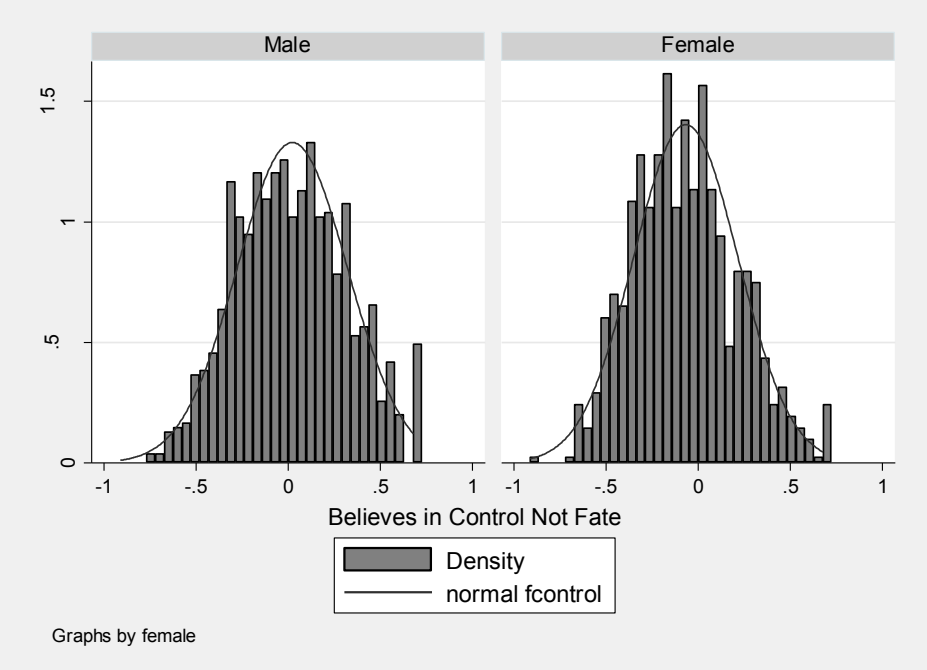
Research on patriarchy and gender require background knowledge about cultural differences. Selected measures gained from the WVS can be compared across both India and Bangladesh using the World Values Survey (WVS, 2009). Furthermore, the Demographic and Health Surveys (DHS) ask relevant questions about domestic violence and women's autonomy in many countries such as Bangladesh (NIPORT, 2007, 2011). India's National Family Health Survey (IIPS, 2006) also covers similar topics to the DHS. A limiting factor is that the DHS questions are selectively omitted by each country when they prepare the survey. In the WVS, too, a question was omitted if a country believed answers would be too homogenous, so there is not easy comparability of coverage.

Retroduction

The logic of asking why the data take the pattern they do.

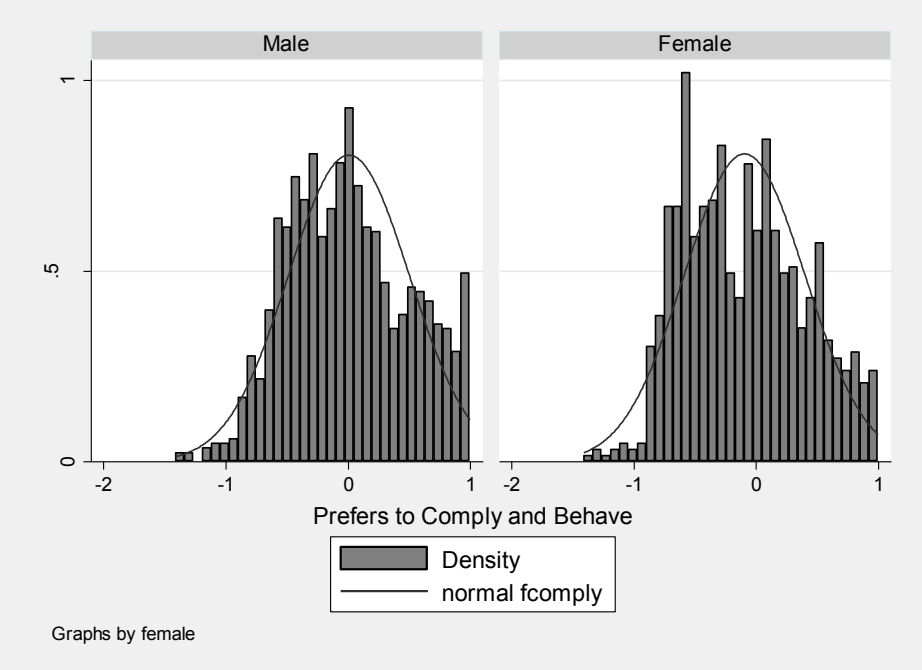
In general, asking why society has developed particular empirical evidence as a reflection of its underlying structure.

Figure 5: India, Socially Normal Level of 'Control' by Gender



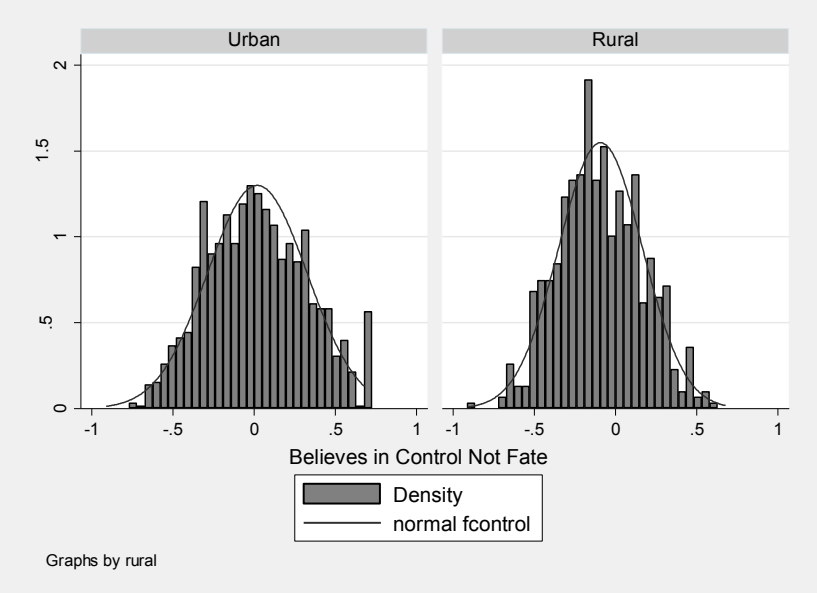
Source: World Values Survey, India, 2006.

Figure 6: India, Socially Normal Level of 'Compliance' by Gender



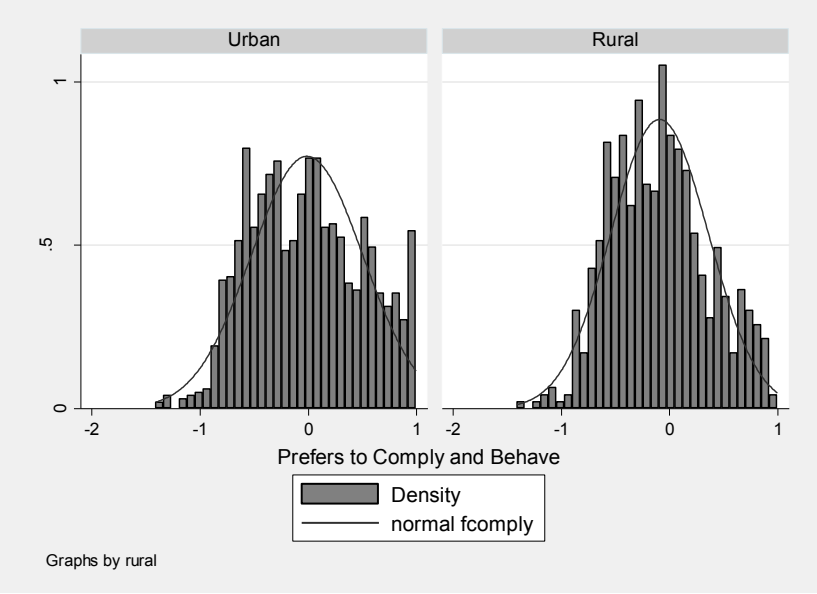
Within India, the latent factor for 'control', ie self-efficacy, is highly significantly different for men and women ($t=7.9$, $p<0.01$), with women lower ie feeling less in control and more fatalistic.

Figure 7: India, Socially Normal Level of 'Control' by Rural/Urban Location



Source: World Values Survey, 2006.

Figure 8: India, Socially Normal Level of 'Compliance' by Rural/Urban Location



Comparing rural with urban within India, rural people feel less 'control', ie self-efficacy, and the difference is highly statistically significant, ($t=7.9$, $p<0.01$).

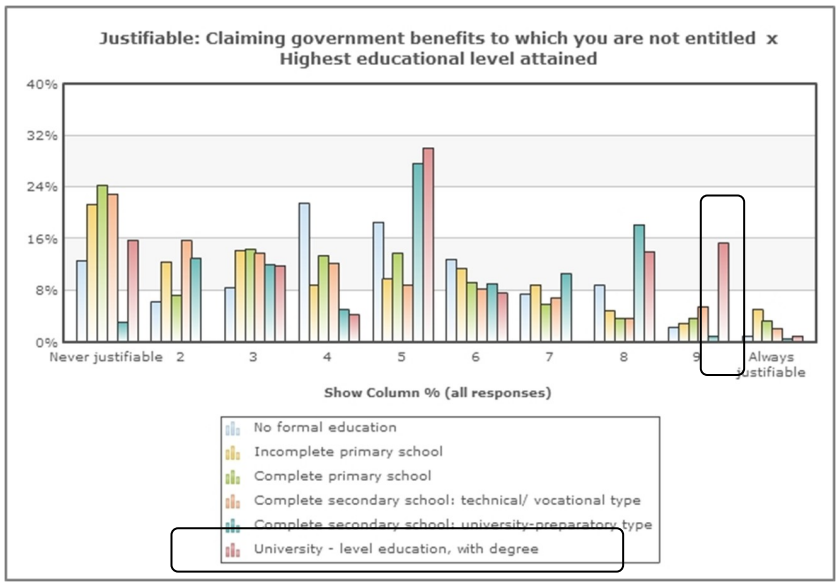
Comparing rural with urban within India, rural people are also less 'compliant', and again the difference is highly statistically significant ($t=3.9$, $p<0.01$). N=2000 and df 1998 throughout. As Figures 5 to 8 show, however, the differences are relatively small. Figure 8 shows ***greater homogeneity in rural India on the issue of being 'compliant'.*** This is a measurable parameter: the variance of EgalDecide.

A Realist Approach to Latent Factors

We perceive each factor as a latent variable: that is, in reality there is a social norm and we have measures which tap into that norm. Other ways of saying the same thing are: "we have traces of the real in the data" (Olsen and Morgan, 2005); "we can remove measurement error by using four measures of one thing, as long as they all indicate that one thing" - a common statement—or as "entity realism" or "Latent variables are measured indirectly through multiple observed variables" (Bowen and Guo, 2012, page 18). Entity realism was discussed by Borsboom, et al.

With realism, the thing measured is real, and it is not just a variable.

Figure 8: Justifiability of Cheating the Government



Authors who have discussed the use of a measure of personal effectiveness as part of a notional latent factor for control include Parboteea et al. (2005) and Lee and Guven (2013). These studies have also measured the degree of 'compliance' of people in different countries. We were unable to compare India and Bangladesh on either point.

Source: World Values Survey online data tool. URL http://www.worldvaluessurvey.org/WVSOnline.jsp.

Interesting findings from 40 countries excluding both India and Bangladesh were that, internationally, those women who were relatively strong risk-takers were also more likely to accept that sometimes a corrupt or cheating practice, such as not paying a bus fare, is justifiable (Parboteea, et al., 2005). Risk-taking tendencies helped explain a raw gender difference in the justifiability of cheating. The risk-taking is not a personal characteristic, but a new social norm.

In the WVS data, Bangladesh people unanimously rejected any justifiability of cheating. In India, however, factor analysis showed a relatively stronger willingness to accept or express that a justification may be given for cheating. An online data tool suggests that saying cheating can be justified tends to occur among those with higher education.

Research jointly supported by the ESRC and DFID

How To Statistically Test Attitudes Over Space: Answers From the Gender Norms Project

Project Partners:

The project 'Gender Norms, Labour Supply and Poverty Reduction in Comparative Context: Evidence from India and Bangladesh" runs from January 2014 –December 2016. The Cathie Marsh Centre for Census and Survey Research (CCSR) at the University of Manchester is working on this in partnership with the BRAC Development Institute, Bangladesh, the Indian Institute for Dalit Studies, Delhi, and the German Institute of Global and Area Studies (GIGA).

See the web site http://www.cmist.manchester.ac.uk/research/projects/norms-labour-supply-and-poverty-reduction/labour-supply-additional-

References:

Bowen, N.K., and S. Guo (2012) *Structural Equation Modelling*, Oxford: Oxford University Press.

Borsboom, D., G.J. Mellenbergh, and J. van Heerden (2003), The Theoretical Status of Latent Variables, *Psychological Review*, 110:2, 203–219, DOI 10.1037/0033-295X.

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*, London: Sage, 4th ed.

International Institute for Population Sciences (IIPS) and Macro International. 2007. *National Family Health Survey* (NFHS-3), 2005–06: India: Volume II.Mumbai: IIPS. pp.18-19.

Kaplan, David (2009). *Structural Equation Modeling : foundations and extensions*. 2nd ed. Los Angeles, London: Sage.

Lee, Wang-Sheng, and Cahit Guven (2013), *Engaging in Corruption: The Influence of Cultural Values and Contagion Effects at the Micro Level*, IZA Working Paper No. 7685.

McKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis,* NY: Lawrence Erlbaum and Associates.

National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Macro International. 2009. *Bangladesh Demographic and Health Survey 2007.*

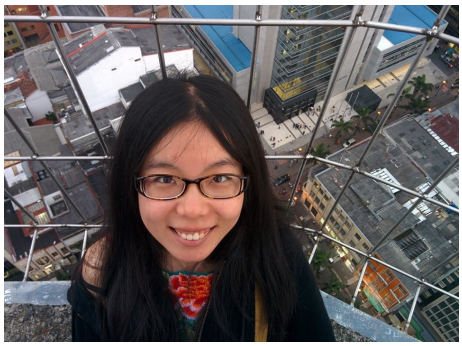
National Institute of Population Research and Training (NIPORT), Mitra and Associates, and ICF International. 2013. *Bangladesh Demographic and Health Survey 2011.*

Olsen, W.K., and J. Morgan (2005) A Critical Epistemology Of Analytical Statistics: Addressing the sceptical realist, *Journal for the Theory of Social Behaviour*, 35:3, September, pages 255-284.

Parboteea et al (2005), "Does National Culture Affect Willingness to Justify Ethically Suspect Behaviours?" , *International Journal of CrossCultural Management,* 5:2, 123-138.

WORLD VALUES SURVEY 1981-2008 OFFICIAL AGGREGATE v.20090901, 2009. World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid.

Wendy Olsen



Min Zhang

Briefing Papers:

This briefing paper is the third in a series. To join a mailing list about this project, and get our updates and Briefing Papers, please send an email to wendy.olsen@manchester.ac.uk.