

QUALITATIVE COMPARATIVE ANALYSIS GROWS UP: SURPRISING APPLICATIONS OF FUZZY SETS

By Wendy Olsen

University of Manchester, Discipline of Social Statistics
and Cathie Marsh Centre for Census & Survey Research

Workshop on Impact Assessment Methods

AIMS

- **1. Help you to interrogate data whether random or non-random samples, using mixed methods. I summarise methods of applying fuzzy set methods to the analysis of large scale and randomly sampled data, including data with control samples or treatment groups.**
- 2. Show how epidemiologists are beginning to use QCA and fuzzy set consistency
- 3. Illustrate multiple pathways of cause

1. THE FUZZY SET MEASUREMENT METHODS

Crisp sets are so simple they are quite useful

0=no, 1=yes

Fuzzy sets are so complex they are controversial

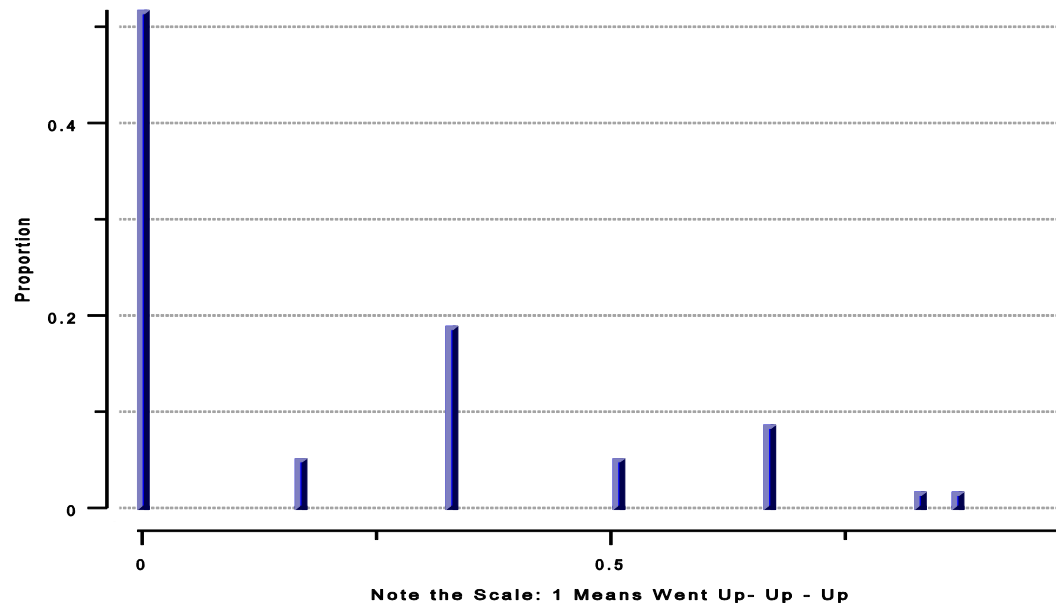
Range: 0 to 1.

Scaling: Degree of membership in the qualitative set defined by 0.

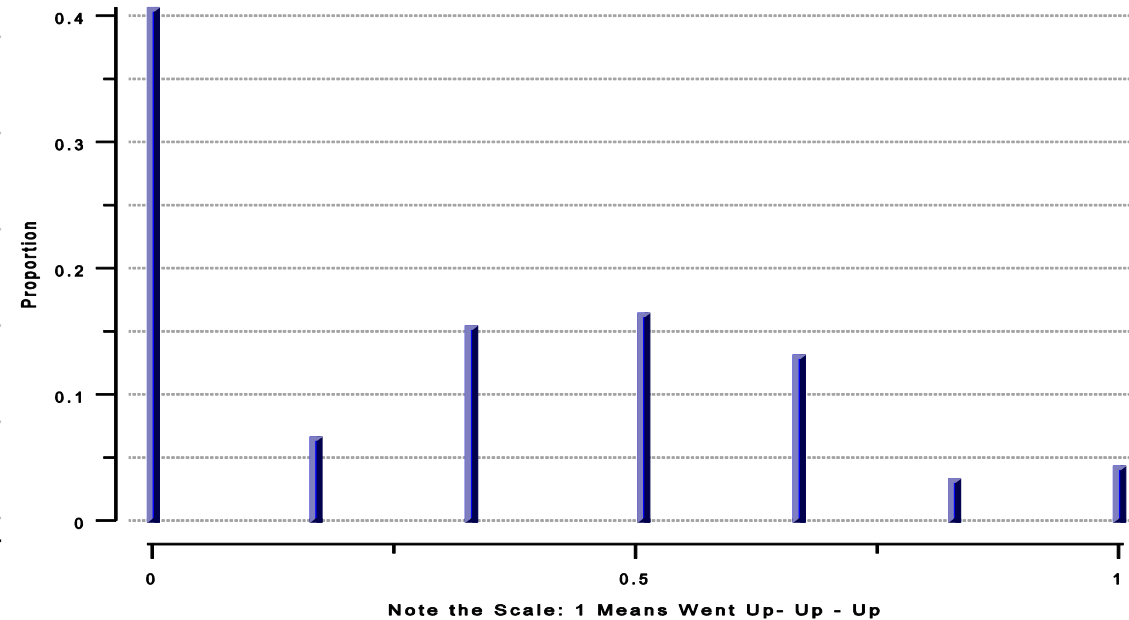
(or by 1, conversely).

FUZZY SET MEASUREMENT – SUBJECTIVE WELLBEING EXAMPLE

Household Econ. Circ's Went Up 2008-9



Indiv. Econ. Circ's Went Up 2008-9



Households' and Individuals' Economic Circumstances Went Down (Low Fuzzy Score)



FUZZY SETS - – OBJECTIVE EXAMPLE 1⁵

QCAClassFuzzySheet39CasesFINAL.xls [Compatibility Mode] - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Add-Ins Team

Clipboard Font Alignment Number Styles Cells Editing

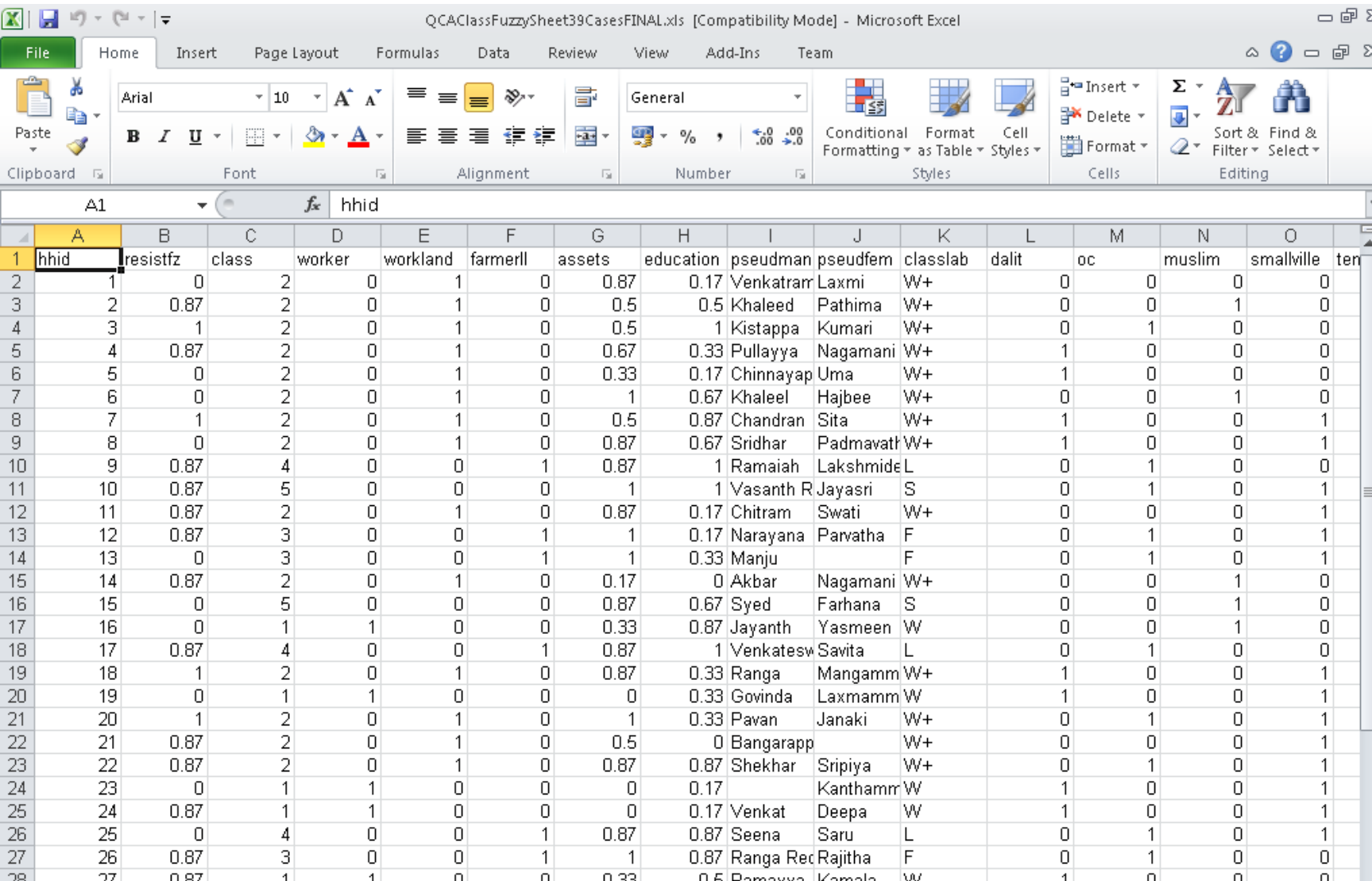
A1 hhid

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	ten
1	hhid	resistfz	class	worker	workland	farmerll	assets	education	pseudman	pseudfem	classlab	dalit	oc	muslim	smallville	ten
2	1	0	2	0	1	0	0.87	0.17	Venkatrarr	Laxmi	W+	0	0	0	0	0
3	2	0.87	2	0	1	0	0.5	0.5	Khaleed	Pathima	W+	0	0	1	0	0
4	3	1	2	0	1	0	0.5	1	Kistappa	Kumari	W+	0	1	0	0	0
5	4	0.87	2	0	1	0	0.67	0.33	Pullayya	Nagamani	W+	1	0	0	0	0
6	5	0	2	0	1	0	0.33	0.17	Chinnayap	Uma	W+	1	0	0	0	0
7	6	0	2	0	1	0	1	0.67	Khaleel	Hajbee	W+	0	0	1	0	0
8	7	1	2	0	1	0	0.5	0.87	Chandran	Sita	W+	1	0	0	0	1
9	8	0	2	0	1	0	0.87	0.67	Sridhar	Padmavath	W+	1	0	0	0	1
10	9	0.87	4	0	0	1	0.87	1	Ramaiah	Lakshmidel	L	0	1	0	0	0
11	10	0.87	5	0	0	0	1	1	Vasanth R	Jayasri	S	0	1	0	0	1
12	11	0.87	2	0	1	0	0.87	0.17	Chitram	Swati	W+	0	0	0	0	1
13	12	0.87	3	0	0	1	1	0.17	Narayana	Parvatha	F	0	1	0	0	1
14	13	0	3	0	0	1	1	0.33	Manju		F	0	1	0	0	1
15	14	0.87	2	0	1	0	0.17	0	Akbar	Nagamani	W+	0	0	1	0	0
16	15	0	5	0	0	0	0.87	0.67	Syed	Farhana	S	0	0	1	0	0
17	16	0	1	1	0	0	0.33	0.87	Jayanth	Yasmeen	W	0	0	1	0	0
18	17	0.87	4	0	0	1	0.87	1	Venkatesw	Savita	L	0	1	0	0	0
19	18	1	2	0	1	0	0.87	0.33	Ranga	Mangamm	W+	1	0	0	0	1
20	19	0	1	1	0	0	0	0.33	Govinda	Laxmamm	W	1	0	0	0	1
21	20	1	2	0	1	0	1	0.33	Pavan	Janaki	W+	0	1	0	0	1
22	21	0.87	2	0	1	0	0.5	0	Bangarapp		W+	0	0	0	0	1
23	22	0.87	2	0	1	0	0.87	0.87	Shekhar	Sripiya	W+	0	1	0	0	1
24	23	0	1	1	0	0	0	0.17		Kanthamm	W	1	0	0	0	1
25	24	0.87	1	1	0	0	0	0.17	Venkat	Deepa	W	1	0	0	0	1
26	25	0	4	0	0	1	0.87	0.87	Seena	Saru	L	0	1	0	0	1
27	26	0.87	3	0	0	1	1	0.87	Ranga Rec	Rajitha	F	0	1	0	0	0
28	27	0.87	1	1	0	0	0.33	0.5	Romayya	Kemala	W	1	0	0	0	0

From Interviews, we calculate how many instances of **Exits**, **Resistance**, **Innovation**, and **Conformity** were mentioned in semi-structured interviews around the landlord-tenant relationship: **Count leads to a Fuzzy Score**



FUZZY SETS - – OBJECTIVE EXAMPLE 2⁶



QCAClassFuzzySheet39CasesFINAL.xls [Compatibility Mode] - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	hhid	resistfz	class	worker	workland	farmerll	assets	education	pseudman	pseudfem	classlab	dalit	oc	muslim	smallville	ten
2	1	0	2	0	1	0	0.87	0.17	Venkatrarr	Laxmi	W+	0	0	0	0	0
3	2	0.87	2	0	1	0	0.5	0.5	Khaleed	Pathima	W+	0	0	1	0	0
4	3	1	2	0	1	0	0.5	1	Kistappa	Kumari	W+	0	1	0	0	0
5	4	0.87	2	0	1	0	0.67	0.33	Pullayya	Nagamani	W+	1	0	0	0	0
6	5	0	2	0	1	0	0.33	0.17	Chinnayap	Uma	W+	1	0	0	0	0
7	6	0	2	0	1	0	1	0.67	Khaleel	Hajbee	W+	0	0	1	0	0
8	7	1	2	0	1	0	0.5	0.87	Chandran	Sita	W+	1	0	0	0	1
9	8	0	2	0	1	0	0.87	0.67	Sridhar	Padmavath	W+	1	0	0	0	1
10	9	0.87	4	0	0	1	0.87	1	Ramaiah	Lakshmid	L	0	1	0	0	0
11	10	0.87	5	0	0	0	1	1	Vasanth R	Jayasri	S	0	1	0	0	1
12	11	0.87	2	0	1	0	0.87	0.17	Chitram	Swati	W+	0	0	0	0	1
13	12	0.87	3	0	0	1	1	0.17	Narayana	Parvatha	F	0	1	0	0	1
14	13	0	3	0	0	1	1	0.33	Manju		F	0	1	0	0	1
15	14	0.87	2	0	1	0	0.17	0	Akbar	Nagamani	W+	0	0	1	0	0
16	15	0	5	0	0	0	0.87	0.67	Syed	Farhana	S	0	0	1	0	0
17	16	0	1	1	0	0	0.33	0.87	Jayanth	Yasmeen	W	0	0	1	0	0
18	17	0.87	4	0	0	1	0.87	1	Venkatesw	Savita	L	0	1	0	0	0
19	18	1	2	0	1	0	0.87	0.33	Ranga	Mangamm	W+	1	0	0	0	1
20	19	0	1	1	0	0	0	0.33	Govinda	Laxmamm	W	1	0	0	0	1
21	20	1	2	0	1	0	1	0.33	Pavan	Janaki	W+	0	1	0	0	1
22	21	0.87	2	0	1	0	0.5	0	Bangarapp		W+	0	0	0	0	1
23	22	0.87	2	0	1	0	0.87	0.87	Shekhar	Sripiya	W+	0	1	0	0	1
24	23	0	1	1	0	0	0	0.17		Kanthamm	W	1	0	0	0	1
25	24	0.87	1	1	0	0	0	0.17	Venkat	Deepa	W	1	0	0	0	1
26	25	0	4	0	0	1	0.87	0.87	Seena	Saru	L	0	1	0	0	1
27	26	0.87	3	0	0	1	1	0.87	Ranga Rec	Rajitha	F	0	1	0	0	0
28	27	0.87	1	1	0	0	0.33	0.5	Romayya	Kemala	W	1	0	0	0	0

From background survey of a random sample, N= 187

Education is a fuzzy set (years, rescaled to the 0-1 range).

We examine **class** with a strong theoretical background.

Assets, labour relations lead to **class structure as Categorical.**



POTENTIAL MICRO INDEPENDENT VARIABLES

- First group are brought in from theories
 - Gender
 - Treatment Group
 - You can use survey weights, design weights, grossing weights
- Demographic variables
 - Age; mean/median age of the household;
 - Hhold size
- Crisp set social class: industrial sectors and work status groups, or asset fuzzy
- Wealth (one can use the Wealth Index. Factor analysis + FUZZY fine)
- Multiple livelihoods (difficult to measure; most had 2 but few had 3 different livelihood strands in monetised sectors)



IND. VARS. CONTINUED

- Treatment Group
 - Example of Self-Help Group Membership: Non-random joining date, early joiners are likely to be special people who are well networked, vocal, leaders or Hhold-credit-constrained
 - NOTICE especially that the wealth are not credit-constrained so DON'T JOIN
- Amount of the treatment
- Be sure to notice that the mechanisms of cause need to be parsed out
- Qualitative research and mixed methods helps in **process tracing**
 - **Example: Membership SHG + Savings + Debt(s) + Invest? + Leadership/voice**
 - And amount of debt, amount of repayment, whether default,
 - Finally whether husband has utilised the loan for an existing profitable activity
- **Sequence analysis** is also useful but needs a lot of data.

2. EXAMPLES FROM EPIDEMIOLOGY

Longest, K.C., and P. Voyts (2012) Gender, the Stress Process, and Health: A Configurational Approach. *Society and Mental Health*, Nov. 2012. 2: 187-206.

Table 2. Configurations' Consistency with Distress versus Not Distress and Reduced Solution Sets for the Total Sample ($n = 528$)

Set ^a	Outcome consistency	Negation consistency	<i>F</i>	Best fit
v·t·m·c·s	.837	.905	8.72	23
v·t·m·c·S	.853	.920	7.77	10
v·t·m·C·s	.854	.914	8.45	17
v·t·m·C·S	.849	.919	9.76	10
v·t·M·c·s	.790	.932	31.16	24
v·t·M·c·S	.767	.935	39.58	28
v·t·M·C·s	.781	.941	40.44	28
v·t·M·C·S	.771	.930	37.56	27
v·T·m·c·s	.897	.832	5.99**	26
v·T·m·c·S	.897	.872	1.09	9
v·T·m·C·s	.884	.862	0.84	25
v·T·m·C·S	.895	.883	0.29	8
v·T·M·c·s	.871	.906	2.21	12
v·T·M·c·S	.885	.905	0.83	9
v·T·M·C·s	.867	.908	3.37	13
v·T·M·C·S	.864	.906	3.31	14
V·t·m·c·s	.893	.862	1.78	12
V·t·m·c·S	.888	.890	0.01	9
V·t·m·C·s	.894	.871	1.03	9

SOME QCA CONCEPTS IN THIS SCENARIO

- necessary cause;
- sufficient causal mechanism

PROTOCOL

- For testing for necessity of a cause X in relation to outcome W , you only need to test individual X variates. It is not important to test complex configurations at this stage.
- Also test for the necessity of X for NOT- W .
 - Advice of Rihoux and Ragin, 2008.

NEXT TEST FOR SUFFICIENCY OF CAUSE

- The vectors differ in complexity.
- As a result there are many, many permutations. Software helps ease the decision problem.

CONSISTENCY FOR SUFFICIENCY

- Consistency is a *measure of the extent to which membership strength in the causal configuration is consistently equal to or less than membership in the outcome* (Epstein et. al, 2007: 10).
- = the *inclusion ratio*, Smithson & Verkuilen (*Fuzzy Set Measurement*, London: Sage, 2006). They do not claim that an inclusion ratio implies causality.
- For each configuration, the joint membership scores ($X_1 \cap X_2 \cap Y$) are added for all **cases**. This number is divided by the sum of all minimum membership scores in the causal combination $X_1 \cap X_2$. The general formula for consistency is:
- Consistency ($\underline{X}_j \leq Y_j$) = $\Sigma(\min(\underline{X}_j, Y_j)) / \Sigma(\underline{Y}_j)$.

IMPORTANT POINT ON METHOD ¹⁵

- Our method acknowledges points made by post-structuralists
- We can do discourse analysis as well as hypothesis testing.
- Knowledge is always fallible. We achieve a warranted argument strongly underpinned by the mixed methods data.
- Triangulation is also helpful.
- But QCA focuses most strongly on the structural factors which are causal mechanisms.
- Structural background factors can enhance, or limit, the success of an intervention.
 - See Byrne and Ragin, eds., *Handbook of Case-Based Research*, London: Sage.

• CONFIGURATIONAL LOGIC OF QCA

- Byrne (e.g. in 2002 *Interpreting Quantitative Data*) argues that configurations are really different from each other. Context really matters.
- It is an empirical question to what EXTENT and HOW they differ.
- As a realist we see configurations as having **causal powers and liabilities**. A case has its own tendencies by virtue of what **configuration it lies within**.
- **Realists are not post-structuralist in deep methodological terms: structuralist instead.**

BOOLEAN INTERSECTION

- AND means the intersection of two sets.
- We interpret AND to mean 'does the configuration of X and Y involve membership in both X and Y? To what extent?'
- X and Y implies taking the minimum of X, Y

BOOLEAN NOTATION

- OR is often presented as \cup
 - OR is also presented by $+$
- AND is presented as \cap
 - AND is also presented by $*$
- NOT is presented by \sim
 - NOT is also shown by small letters.

3. ILLUSTRATE MULTIPLE PATHWAYS OF CAUSE

- This is not merely a deductive exercise.
- Developing the argument does not reside merely in the domain of social statistics.
- One applies disciplinary expertise and one uses existing theory.

RESEARCH METHODOLOGY

- Mixed Methods Data Collection
- *Variables from the survey method + Interview texts in NVIVO, or Focus Groups, to provide in-depth insight and case-studies)*
- Mixed Methods Triangulation at the Analysis Stage is Recommended

Data is a kind of evidence

Use other evidence about correlation of the treatment with income, etc.

What reasoning links the treatment with the outcome? Depends on the actual, varied mechanisms. QCA finds the various PATHWAYS OF CHANGE.

DATA

- Lam and Ostrom analysed watersheds in Nepal.
- They analysed them historically so time is embedded in the variables, not in panel.
- The data was adapted qualitatively so LONGTERM GOOD WATER SUPPLY was the final outcome, carefully defined.
- **0=Not good water supply. 1 = Good longterm W.**
- **Also designated W and ~W or W and w.**
- SHORTTERM GAINS were shown to vary from the longterm gains.

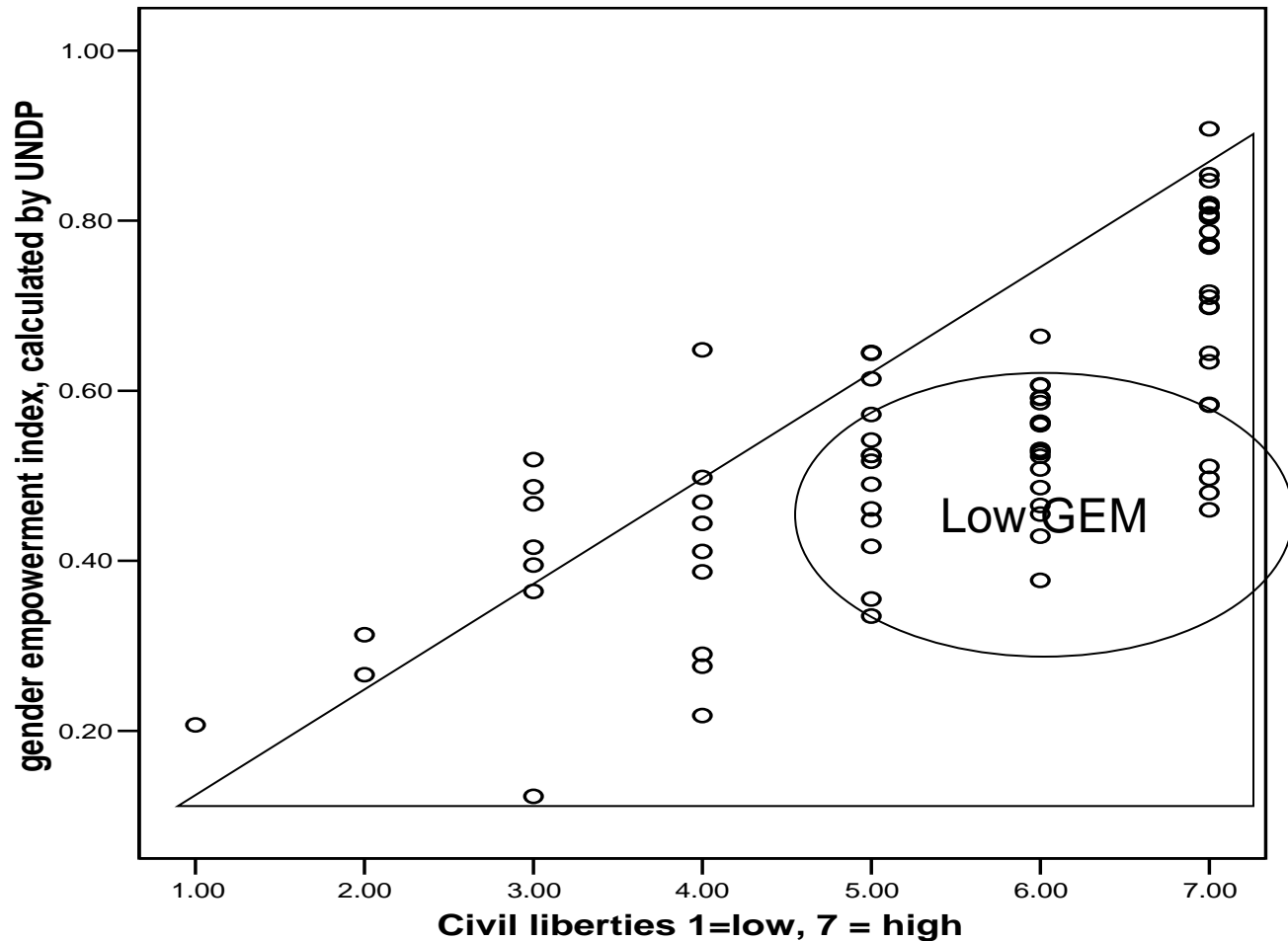
THREE TYPES OF HYPOTHESIS

- 1. E is necessary for the outcome to be enhanced.
- 2. A combination ABfG is sufficient for the outcome to be enhanced.
- 3. A series of combinations all qualify, using the consistency cutoff of 0.8, and these can be summarised in an equation.

ILLUSTRATION

- Use Lam and Ostrom to Illustrate. Convert their table to numbers. In fsQCA set the S-consistency cutoff at 0.80 and frequency ≥ 1 .

EXAMPLE: CIVIL LIBERTIES APPEAR TO BE NECESSARY BUT NOT SUFFICIENT FOR GENDER EMPOWERMENT

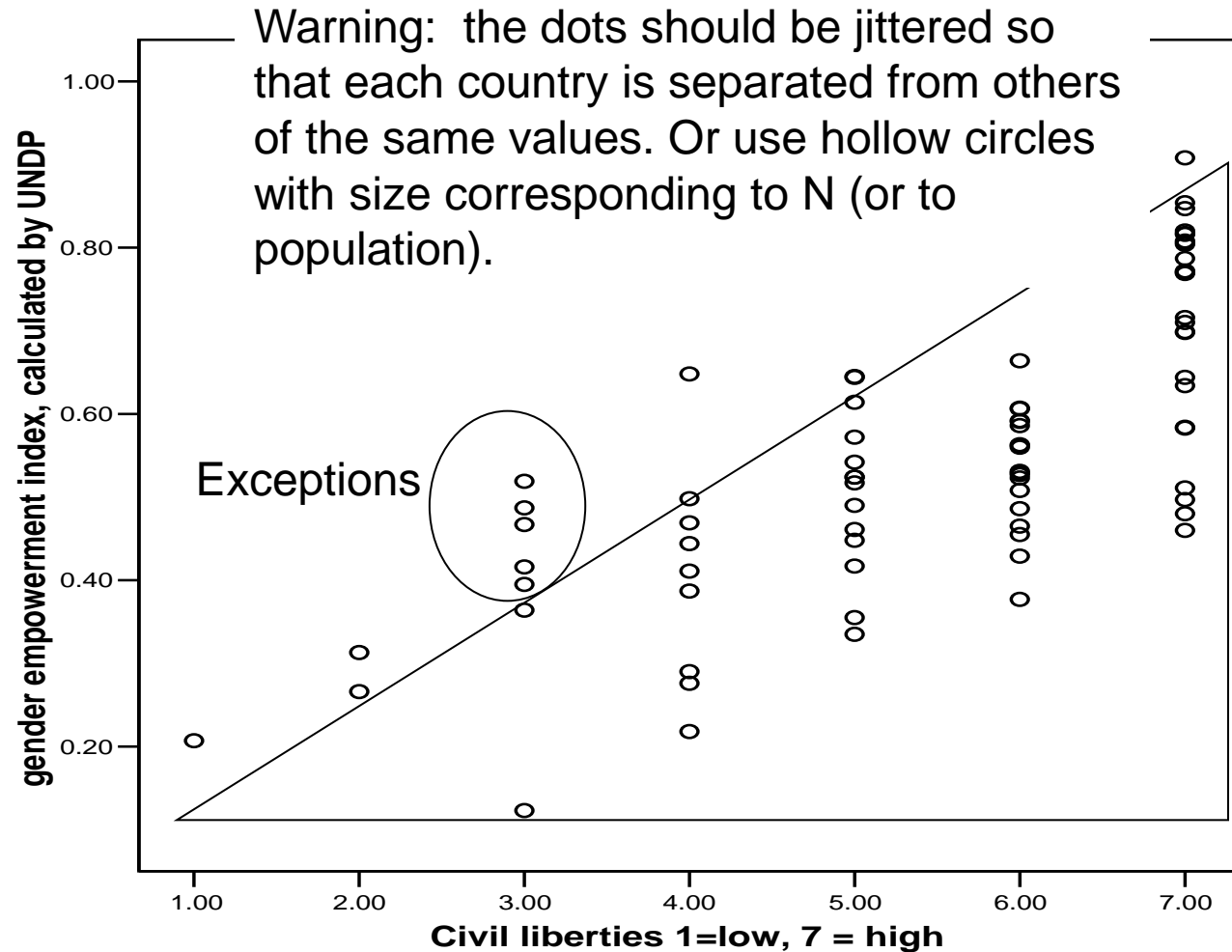


Key: Each case is one country.

Horizontal: freedomhouse **Civil Liberties Index** (Gastil index)

Vertical: UNDP **Gender Empowerment Index**

EXAMPLE: SHOWING THE EXCEPTIONS



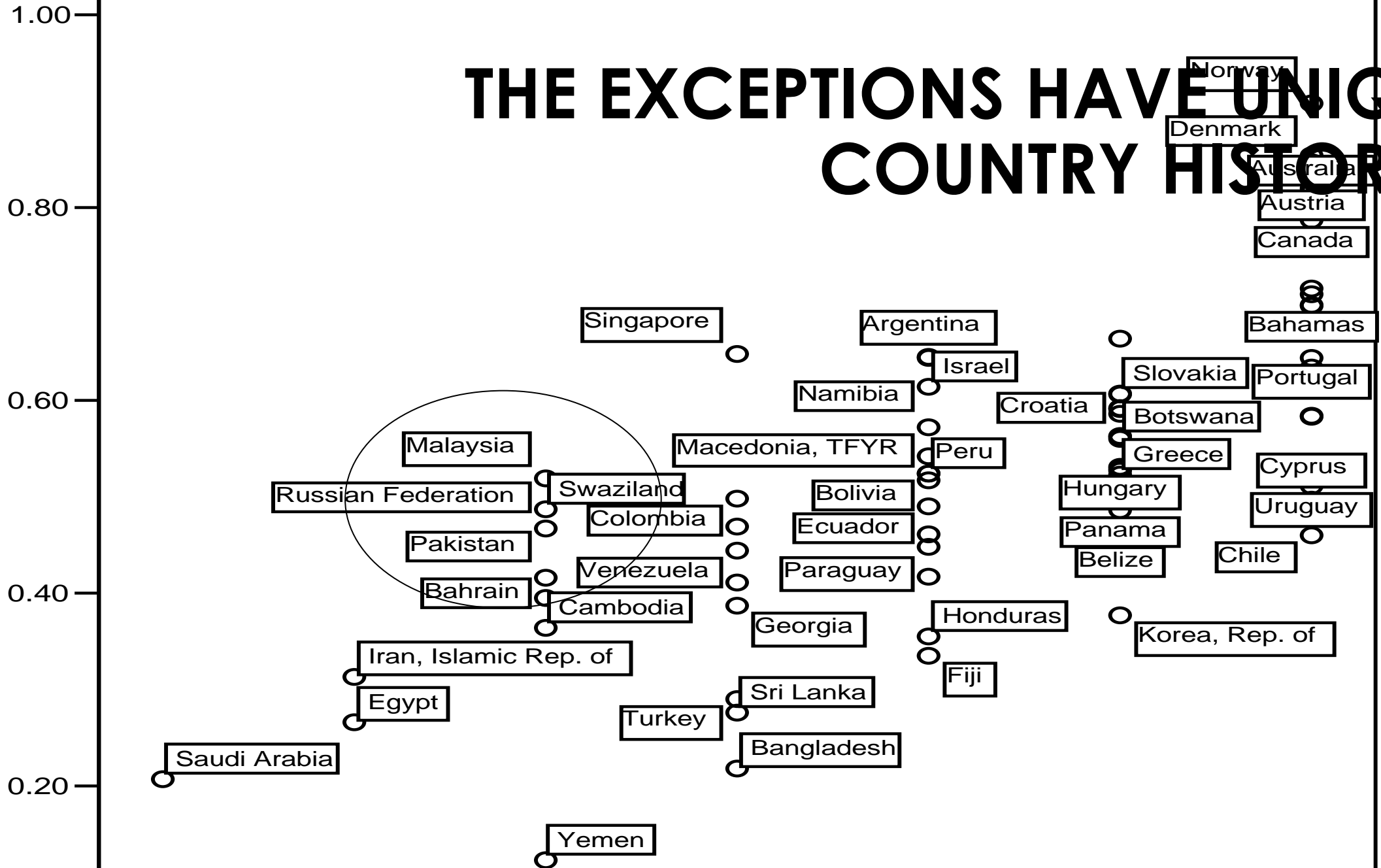
Key: 1 case is one country.

Horizontal:
freedomhouse civil liberties index (Gastil index)

Vertical: UNDP Gender Empowerment Index

THE EXCEPTIONS HAVE UNIQUE COUNTRY HISTORIES

gender empowerment index, calculated by UNDP



DATA – RAW – NEPAL IRRIGATION SUCCESS IN W

A	R	F	L	C	Success in W	NOT W
0	1	0	1	1	1	1
0	1	1	1	1	2	0
1	1	0	1	1	2	0
1	1	1	1	1	2	0
0	0	0	0	1	1	0
0	1	0	0	1	1	0
0	0	1	1	1	0	1
1	0	0	0	0	0	1
1	1	1	0	0	1	0
1	1	0	0	1	1	0
1	1	1	0	1	1	0

The number of **rows** here is 11. This is the number of configurations, allowing for some contradictory configurations, i.e. counting as a single configuration both the aRfLC combination with W and NOT-W outcome. That is, aRfLCW and aRfLCw.

FSQCA RESULTS

- This is a simple CSV or XLSX file.
- Model: $w = f(a, r, f, l, c)$
- No factor was necessary overall.
- Test for sufficiency.
- Lam & Ostrom report that several combinations of factors were sufficient.

RESULTS, CRISP SET QUINE ALGORITHM

(HERE - WE HAVE STILL SPECIFIED THAT CONTRADICTIONS ARE EXCLUDED
ALSO CODED A CONTRADICTION AT 0.5 AS A 0 I.E. EXCLUDE/NO.)

	raw coverage	unique coverage	consistency
$a*r*c$	0.500000	0.250000	1.000000
$\sim a*\sim f*\sim l*c$	0.166667	0.166667	1.000000
$a*r*f*\sim l$	0.166667	0.083333	1.000000
$r*f*l*c$	0.333333	0.166667	1.000000
solution coverage: 0.916667			
solution consistency: 1.000000			

This tells you: The four pathways' results are all highly consistent (1.0) with Sufficiency. A and R and C were sufficient to achieve a good water supply outcome.

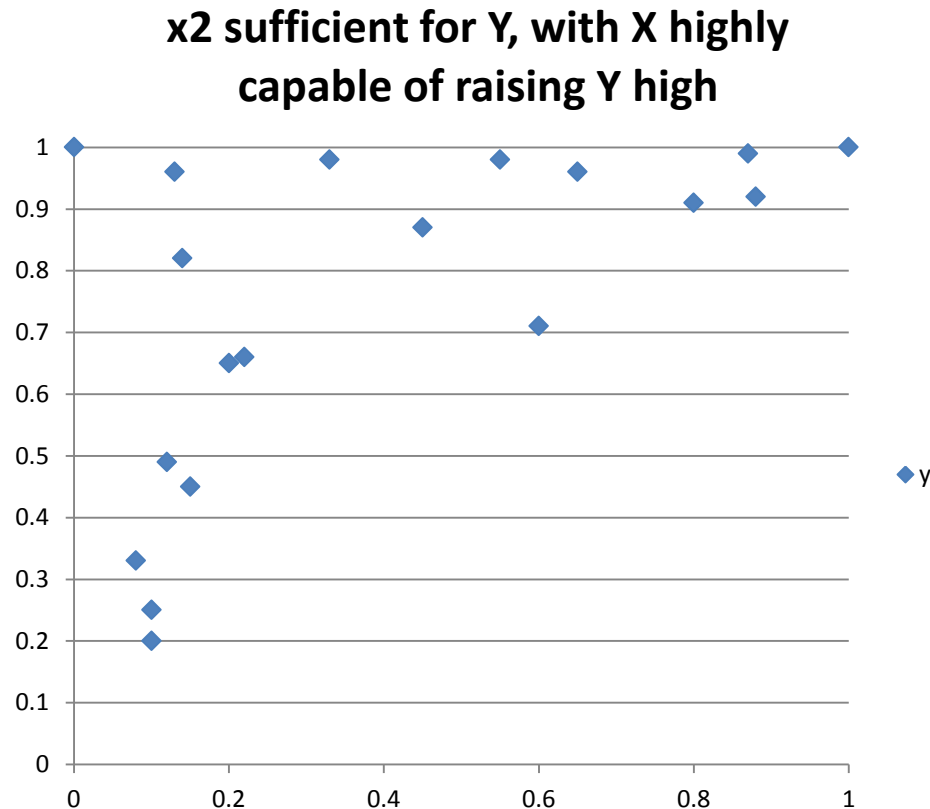
But not-A combined with absence of F, L and the presence of C was also sufficient ; more rarely; a particular pathway.

R is invoked in 3 out of 4 sufficient pathways in Nepal watersheds. So is C.

HOW THIS IS REPORTED:

- ARC or aflC or ARFI or **RFLC*** → outcome.
- ARC + aflC + ARFI + RFLC → outcome.
- AR (C + FI) + ... → W
- AR (C + FI) + C (afl + RFL) → W
- These steps use Boolean algebra.
- The four terms represent 13 cases of HIGH W and all the other case offers a contrast.
- These are four pathways.
 - The necessity of F occurs only within two pathways.
 - The absence of F is necessary within one pathway.
 - A also plays a role that varies from pathway to pathway.
 - * a short appendix uses this configuration to define 'consistency'.

TAKING THIS FORWARD WITH BOOTSTRAPPING



- Calculate C which is 1.0 in the diagram shown.
- Exceptions reduce C below 1.0.
- Bootstrap the sample by taking 1000 samples *with replacement*.
- Permutate through each possible configuration.
- Calculate C for each permutation.
- Calculate Upper and Lower bounds, which are asymmetrical.
- Compare and rank the pathways.

REFERENCES

- Lam and Ostrom (2010). "Analyzing the dynamic complexity of development interventions: lessons from an irrigation experiment in Nepal, *Policy Science*, 43:1, pp. 1-25. DOI 10.1007/s11077-009-9082-6
- **Ragin, C.C. (2008). Redesigning social inquiry: Set relations in social research. Chicago: Chicago University Press.**
- **Ragin, C. C. (2000). Fuzzy-set social science, Chicago; London, University of Chicago Press. (One only needs to read the first half to cover QCA; the second half covers Fuzzy Set Analysis.)**
- **Byrne, D., and C. Ragin, eds. (2009), Handbook of Case-Centred Research Methods, London: Sage.**
- Rohwer, Gotz (2010). Qualitative Comparative Analysis: A Discussion of Interpretations. *European Sociological Review*. Online version. July 19. DOI: 10.1093/esr/jcq034
- **Rihoux, B., & Ragin, C. C. (2009). Configurational comparative methods. Qualitative Comparative Analysis (QCA) and related techniques (Applied Social Research Methods). Thousand Oaks and London: Sage.**
- HUNT, SHELBY D., ***A Realist Theory of Empirical Testing: 'Resolving the Theory-Ladenness/Objectivity Debate'***, *Philosophy of the Social Sciences*, 24:2 (1994:June), p.133

MORE REFERENCES

- Byrne, D. (2005). "Complexity, Configuration and Cases", *Theory, Culture and Society* 22(10): 95-111.
- Ragin, C. C. (1987). *The Comparative Method: moving beyond qualitative and quantitative strategies*. Berkeley ; Los Angeles ; London, University of California Press.
- Snow, D. and D. Cress (2000). "The Outcome of Homeless Mobilization: the Influence of Organization, Disruption, Political Mediation, and Framing." *American Journal of Sociology* **105**(4): 1063-1104.
- Chapter 15 of R. Kent (2007), *Marketing Research: Approaches, Methods and Applications in Europe*, London: Thomson Learning.

Thank you for your attention.

Comments to me via email wendy.olsen@manchester.ac.uk

Twitter address: @Sandhyamma