Dual Calibration of Fuzzy Sets

It is often useful to create more than one fuzzy set from a single source variable. This is especially important given the asymmetric nature of set-theoretic analysis.

Consider, for example, the causal conditions linked to poverty versus the causal conditions linked to having a well-paying job. When thinking in terms of variables, it is straightforward to examine the correlation between “education” and these two outcomes. Education has a negative correlation with poverty and a positive correlation with having a well-paying job.

When thinking in terms of fuzzy sets, however, it is necessary to abandon abstract nouns like “education” and think in terms of sets and the adjectives that describe them. It is the “set of low-education people” who tend to end up in poverty and the set of “high-education people” who often end up with the well-paying jobs. To examine these connections, it is necessary to calibrate education in two different ways, focusing on the different ends of the education distribution.

An illustration:

The set of people with low education might be calibrated as follows:

	Membership in low education	**Membership in not-low education**
4 years	0.99	0.01
6 years	0.95	0.05
10 years	0.75	0.25
12 years	0.50	0.50
14 years	0.25	0.75
16 years	0.05	0.95
18 years	0.02	0.98
20 years	0.01	0.99

The set of people with high education, by contrast, might be calibrated as follows:

	Membership in high education	**Membership in not-high education**
4 years	0.01	0.99
6 years	0.01	0.99
10 years	0.02	0.98
12 years	0.05	0.95
14 years	0.25	0.75
16 years	0.50	0.50
18 years	0.75	0.25
20 years	0.95	0.05

The calibration shift just shown is subtle but important, especially considering the fact that set-theoretic relations are the key focus, and these in turn are defined by $x_i \leq y_i$ (x is a subset of y, a pattern consistent with sufficiency) and $x_i \geq y_i$ (x is a superset of y, a pattern consistent with necessity).

From a correlational viewpoint, by contrast, the difference between the two calibrations is inconsequential, for the correlation between these two different codings of years of education would be very strong, close to 1.0.

Dual Calibration and Crisp Sets

Usually, calibration and dual calibration are associated with fuzzy sets. However, crisp sets can also be coded so that they accomplish similar objectives. Consider “low education” and “high education” as crisp sets, using the following coding scheme:

	Membership in low education	**Membership in high education**
4 years	1	0
6 years	1	0
10 years	1	0
12 years	0	0
14 years	0	0
16 years	0	1
18 years	0	1
20 years	0	1

Possible solutions:

lowed: recipe applies to low education respondents
~lowed: recipe applies to middle and high education respondents
highed: recipe applies to high education respondents
~highed: recipe applies to middle and low education respondents

Notice also that the two middle categories in this scheme are coded as neither high nor low. Suppose we used these two crisp sets (“lowed” and “highed”) as causal conditions in fsQCA, and one of the solutions was: **~lowed*~highed**. It’s the intersection of not-low and not-high, which is mid-level education (versus low or high education).

Thus, another benefit of dual calibration is that it opens the door to causal configurations specific to the middle range of causal conditions that have been coded or calibrated in a dual manner. (This benefit applies as well to fuzzy sets.)

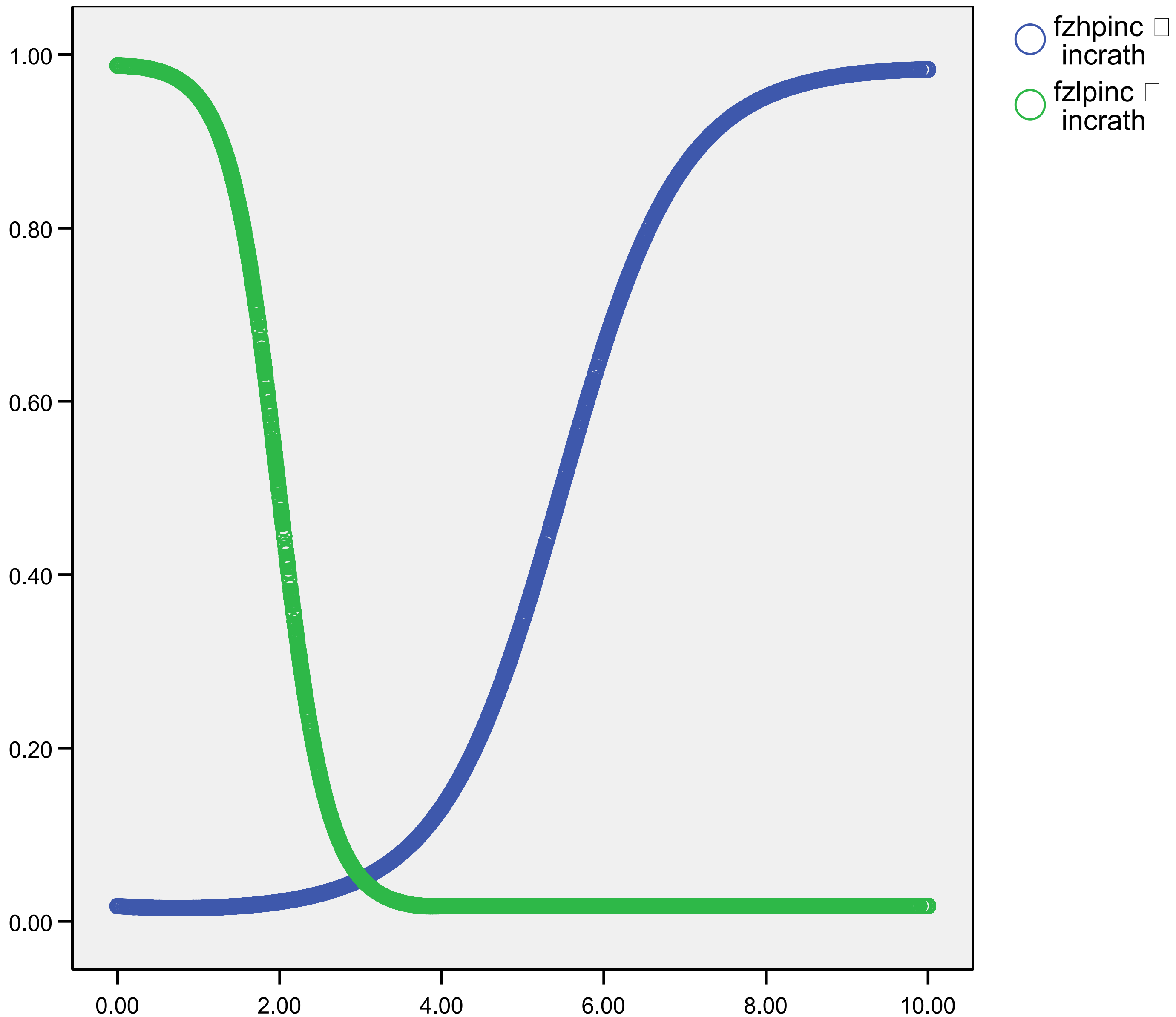
Dual Calibration and Causal Mechanisms

Dual calibration is also useful for gaining greater insight into the nature of causal mechanisms. For example, when it comes to life chances, is it more important to have rich parents or is it OK simply to have parents who are at least middle class? This is not a question that can be answered very easily with correlational methods. You can see if the slope for the variable income flattens and, if so, where in the income distribution it does flatten, and this provides some important signals about whether you have to be rich or just middle class. But the linear model is strongly biased in favor of linear relationships and, in any event, curvilinearity is always specification dependent (now you see it, now you don't).

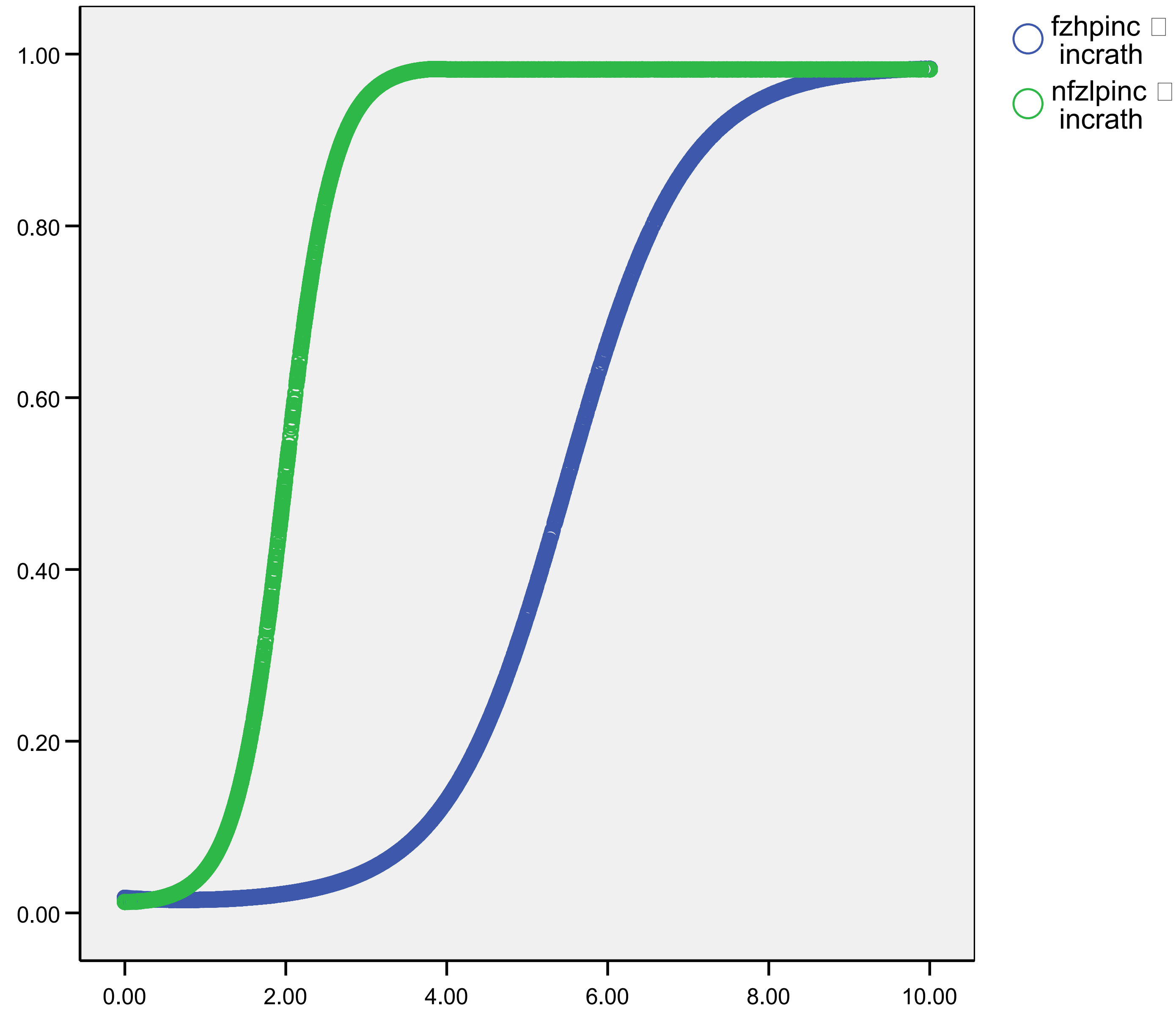
With fuzzy sets, answering questions like the one just posed is more straightforward because you can directly calculate degree of membership in different sets, using the same source variable, and you can enter these different calibrations of the same source variable into the same analysis. (It is important to note as well that multicollinearity is not an issue in set theoretic analysis.)

For illustration, consider the source variable, the ratio of parental income to the household-adjusted poverty level and two fuzzy sets: the set of respondents with low-income parents and the set of respondents with high income parents:

PLOT of Membership in Low Parental Income and High Parental Income Against Ratio of Parent's Income to Poverty Level



PLOT of Membership in Not-Low Parental Income and High Parental Income Against Ratio of Parent's Income to Poverty Level



Let's take a look at the impact of these two different calibrations on staying out of poverty for white males and black females.

Outcome: Degree of membership in the set "not in poverty"

Causal conditions:

- Degree of membership in low-income parents (fzlpinc)
- Degree of membership in high-income parents (fzhpinc)
- Married versus not (crisp set: mar)
- Children versus no children (crisp set: kid)
- Degree of membership in college educated (fzcled)
- Degree of membership in high-school educated (fzhsed)

Data set: National Longitudinal Survey of Youth

Outcome: Not in Poverty
Conditions: FZLPINC, FZHPINC, MAR, KID, FZHSED, FZCLED

White Males Only--Parsimonious Solution:

	raw coverage	unique coverage	consistency
fzcled+	0.396070	0.067727	0.911020
fzhpinc*~kid+	0.397651	0.114321	0.901613
mar*~kid+	0.241772	0.063425	0.928265
fzhpinc*mar	0.403286	0.162043	0.915832

solution coverage: 0.804818
solution consistency: 0.876410

Black Females Only--Parsimonious Solution:

	raw coverage	unique coverage	consistency
~kid*fzcled+	0.217991	0.155730	0.869889
fzhpinc*mar+	0.135728	0.049523	0.922754
mar*fzcled+	0.182578	0.073477	0.908326
~fzlpinc*mar*~kid	0.102084	0.030031	0.961687

solution coverage: 0.433405
solution consistency: 0.877275

Notice that the recipes for black females are subsets of the recipes for white males, which is to say simply that the conditions for staying out of poverty are tougher for black females:

White Male Recipes	Black Female Recipes
fzcled	fzcled(~kid +mar)
fzhpinc(~kid +mar)	fzhpinc*mar
mar*~kid	~fzlpinc*mar*~kid

The key point, however, is that because parental income is calibrated in two different ways, we are able to see important differences between the two groups that otherwise would have been obscured (without dual calibration).