AMERICAN SOCIOLOGICAL REVIEW

OFFICIAL JOURNAL OF THE AMERICAN SOCIOLOGICAL ASSOCIATION

ONLINE SUPPLEMENT
to article in

AMERICAN SOCIOLOGICAL REVIEW, 2006, VOL. 71 (OCTOBER:778–807)

How Cultural Tastes Shape Personal Networks

Omar Lizardo
University of Notre Dame

MATCHING ANALYSIS

While the instrumental variables technique has allowed us to obtain estimates of the effect of cultural taste on the size of the personal network, the validity of the results hinge on the validity of the exclusion restriction and the dependence on strong instruments for parameter identification. Furthermore the standard errors of the nonlinear IV Poisson model are probably too low due to the presence of overdispersion (Barron 1992). One way to assure ourselves of the robustness of the results is to attempt to replicate the results using a different set of techniques designed to estimate casual effects from observational data. I follow Winship and Morgan's (1999:703) advice to researchers encouraging the implementation of "…a variety of methods…to determine how robust the treatment effect estimate is to alternative methods." To this end, I resort to *matching*. As noted by Winship and Morgan (1999: 673–74), matching has several advantages over other approaches of estimating casual effects in non-experimental data: 1) matching "…makes no assumption about the functional form of the dependence between the outcome of interest and the other Xs" and 2) "…matching ensures that only those portions of the distribution of the Xs in the observed data that contain individuals in both the treatment and control groups enter the estimation of the treatment effect" and 3) because fewer parameters are estimated than in a regression model, matching is more efficient."

In this section, I report the results of a matching analysis in which I estimate the casual effect of being in one of the two culture consumption groups on the expected size of the strong and weak tie components of the personal network. Matching analysis is an econometric technique initially designed to estimate casual effects from non-experimental data (Cochran and Rubin 1973; Rubin 1974; Rosembaum and Rubin 1983, 1985). As noted by Winship and Morgan (1999; see also Morgan and Harding 2005) in matching, the notion of a casual effect is boiled down to its most basic definition, taken from experimental methodology: the average casual effect (δ_i) is the arithmetic difference in the average outcome score between the group that was exposed to the potential casual agent (treatment) and the group which was not (control):

EQUATION 1

$$E[\delta_i] = E[Y_i^T] - E[Y_i^C]$$

Where *i* is a subscript that denotes each individual in the two groups. The "fundamental problem of casual inference" (Holland 1986) in observational data is centered on the fact that the observed assignment to the "treatment" and "control" groups is not random, and that therefore the estimation of casual effects depends on making inferences about at least two unknown quantities: the counterfactual expected score on the outcome for the group that was exposed to the treatment *if* the individuals in that group *had not been* exposed to the treatment ($E[Y_i^T | T_i = 0]$) and the expected score on the outcome among the untreated *if* that group *had been* exposed to the treatment ($E[Y_i^C | T_i = 1]$), where T_i is a dummy variable that equals one if individual *i* was exposed to the treatment and zero otherwise.

The idea behind matching is fairly simple and it revolves around a "second best" solution to having observed the actual values of the outcome under the two counterfactual conditions. Using relevant sociodemographic variables in the data set, it is possible to construct groups that are as close as possible to the treatment and control groups but which have the *opposite* value on the candidate casual factor of interest. These groups can then serve as stand-ins for the counterfactual groups that would be just like the treatment and control groups in all relevant respects (i.e., average age, educational attainment, etc.) *except* for the variable that indexes the treatment. We can then measure the average score for the outcome variable among the unmatched and matched versions of the control and treatment groups. Thus, under the assumption that *conditional on having been matched on the relevant sociodemographic variables* assignment to treatment—i.e. popular culture consumer—is random, then the casual effect of belonging to each of the culture consumption groups on the outcome, in this case the expected size of the strong and weak ties components of the personal network, can be estimated. In this sense, the expected score on the outcome for the group that is just like the group that received the treatment in all relevant sociodemographic respects *except for the fact of not having received the treatmen*t, is an estimate of the counterfactual score on the outcome among the treated had they not received the treatment ($E[Y_i^T | T_i = 0]$). In a similar way, the expected score on the outcome for the group that is the "sociodemographic twin" of the group that did not receive the treatment *except for the fact of having received the treatment*, can be taken as an estimate of the counter-factual score on the outcome among those who did not receive the treatment had they received the treatment ($E[Y_i^C | T_i = 1]$).

In the present case, matching can allow us to answer the following questions: 1) what is the expected change in the number of weak and strong ties that we would observe if—holding all relevant sociodemographic covariates constant—a popular culture consumer was transformed into a non-consumer? 2) What is the expected change in the number of weak and strong ties that we would observe if a highbrow culture consumer was transformed into a non-consumer? If hypothesis 1 is correct, we should observe a significant difference in the expected number of strong ties when looking at matched respondents across the two levels of highbrow consumption (consumer and non-consumer), but there should be no significant difference in the expected number of *weak* connections across the highbrow culture consumer/non-highbrow culture consumer divide. In a similar way, if hypothesis 2 is correct, we should observe a significant change in the expected number of weak ties when looking at the difference between matched respondents who engaged in popular culture consumption versus those who did not. However, the expected number of *strong* ties should not differ across the two levels of popular culture consumption status when respondents are matched on all other relevant sociodemographic factors.

I use propensity score matching (Rosembaum and Rubin 1985) on the following sociodemographic variables: age, years of education, SES, average parental SES and the expected (binary) culture consumption score for the non-treatment variable. These are all variables that have shown to be the primary determinants of both highbrow and popular culture consumption, and should thus be able to do a good job of explaining the bulk of the heterogeneity in culture consumption status. This means that respondents are matched on their expected degree of highbrow consumption when looking at the casual effects of popular consumption and vice versa. A propensity score is simply the expected probability taken from a logistic regression model of all of the matching covariates on dichotomous versions (binarized at the 75th percentile) of the popular and highbrow consumption scales. Rosembaum and Rubin (1983) show that this single dimension contains all of the information relevant to estimating the bias in treatment assignment that comes from individual heterogeneity. Individuals with the same propensity score are then assumed to have similar values on the relevant matching variables. I use the *nearest neighbor matching* algorithm implemented in the *psmatch2* routine written by Leuven and Siansesi (2003) for version 9.1 of the Stata statistical package (Statacorp 2005).[1]

[1] In nearest neighbor matching the counterfactual the treatment and control cases are chosen by taking a random respondent from each group and matching it with a "nearest neighbor" from the opposite group as defined by their propensity scores differences. For instance, a respondent from the control group is matched with a random respondent from the treatment group with the probability of match being a linear function of the inverse of their propensity score difference. Statistical tests showing the mean differences for the pre-match and post-match samples in the relevant variables of interest are shown in Table 2.

RESULTS

Table 1 shows the results of the matching procedure applied to the 2002 General Social Survey Data. In Table I show the expected number of weak and strong ties for respondents in different culture consumption groups along with the average difference and an estimate of the standard error of the difference (for details on the calculation of the standard error, see Leuven and Siansesi [2003]). Since hypotheses one and two are clearly directional, I use one tailed tests of significance. The upper part of the table shows the expected size of the strong and weak tie portion of the personal network for treated and control groups, where treatment is defined as being a popular culture consumer, and the lower part of the table show the results for the same analysis, this time with the treatment variable being whether the respondent engages in highbrow culture consumption.

Table 1. Average Number of Strong and Weak Ties for Matched Groups of Popular and Highbrow Consumers and Non-Consumers 2002 GSS

Outcome	Groups	E(Y) Treated	E(Y) Controls	δ	S.E. (δ)
Treatment: Popular Culture Consumption					
Number of weak ties	Unmatched	27.18	21.39	5.78	2.07**
	Matched[a]	27.18	18.53	8.65	2.86**
Number of strong ties	Unmatched	8.86	7.71	1.15	.61*
	Matched[a]	8.86	8.48	.38	1.03
Treatment: Highbrow Culture Consumption					
Number of weak ties	Unmatched	27.81	20.73	7.08	2.05**
	Matched[a]	27.81	25.53	2.28	3.69
Number of strong ties	Unmatched	9.93	6.89	3.04	.60**
	Matched[a]	9.93	7.22	2.71	1.00**

[a] Treated and control groups are matched on the mean for the following variables: years of education, years of age, average parental education, and highbrow and popular taste.
* $p < .05$; ** $p < .01$ (one-tailed tests).

For popular culture consumption, the results are consistent with the log-linear and the instrumental variables analyses. For the unmatched samples, the "naïve" estimate of the casual effect of popular culture consumption on the expected number of weak ties is 5.78, meaning that respondents who belong to the popular culture consumption group are expected to have close to six more weak ties, on average, than respondents who do not engage in popular culture consumption. As shown in the table this gap between the treated (popular culture consumers) and the untreated (popular culture abstainers) remains substantial (increasing to about 8.65), *even when respondents are matched on education, family background, age and highbrow culture consumption*. This suggests that the prior association between *none* of those other variables and popular culture consumption is responsible for the effect. Thus if we had to two hypothetical respondents matched on those sociodemographic dimensions but were to suddenly render one of them a popular culture non-consumer, she would lose on average between 6 and 8 weak ties.

Notice that the same is *not* the case for strong ties. In the unmatched sample the difference between popular culture consumers and non-consumers in the number of strong ties is about 1.15. However, once respondents are matched on the relevant sociodemographic dimensions—essentially "controlling" for them—this difference disappears ($p > .10$), suggesting that popular culture consumption status has *no casual influence* on the number of strong ties, with everything else held constant.

In the case of highbrow culture consumption, a different pattern of results emerges, also consistent with the previous instrumental variable estimates. In the unmatched sample, there is a difference of about 7.08 weak ties between highbrow culture consumers and non-highbrow consumers, suggesting a potentially substantial "naïve" effect of highbrow culture consumption on weak ties. However, once individuals are matched on popular culture consumption status and the relevant sociodemographic dimensions, this effect is reduced by more than two-thirds (to 2.28) and is no longer statistically significant within conventional thresholds, suggesting that highbrow consumption has *no casual influence* on the number of weak ties once individuals are matched on the relevant sociodemographic covariates. One important difference between this result and the IV estimates is that there does not appear to be a negative effect of highbrow consumption on the number of weak ties, but simply a *null* effect.

The results when the expected value of the number of strong ties is the outcome of interest are very different, and are consistent with the idea of a *selective* effect of highbrow consumption on strong ties. For individuals matched on age, SES, educational and parental class background, a change of status from highbrow culture consumer to highbrow culture non-consumer would result in the loss of about (2.7) two strong ties on average, a quantity that is very similar to the unconditional casual estimate (3.04). This implies that while highbrow consumption has a null effect on the number of weak ties, it has a strong effect on the expected number of strong ties, even after respondents are matched on education, family background, age and popular culture consumption.

REFERENCES

Barron, David N. 1992. "The Analysis of Count Data: Overdispersion and Autocorrelation." *Sociological Methodology* 22:179–220.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–60.

Winship, Christopher and Stephen L. Morgan. 1999. "The Estimation of Causal Effects from Observational Data." *Annual Review of Sociology* 25:659–706.

Cochran, W. G. and Douglas B. Rubin. 1973. "Controlling Bias in Observational Studies" *Sankyha* 35:417–46.

Leuven, Edwin and Barbara Sianesi. 2003. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing." http://ideas.repec.org/c/boc/bocode/s432001.html.

Rosenbaum, Paul .R. and Douglas B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrika* 70:41–55.

Rosenbaum, Paul .R. and Douglas B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39:33–38.

Rubin, Douglas B. 1974. "Estimating Causal Effects of Treatments in Randomised and Non-Randomised Studies." *Journal of Educational Psychology* 66:688–701.

Table 2. Mean Differences T-tests for Differences Between the Matched and Unmatched Samples on the Matching Covariates Used in the Analysis

Variable	Sample	Mean Treated	Mean Control	Percent Bias	Percent Reduction in Bias	T-test	*p* > t
Average Parental Education	Unmatched	1.2088	.71042	53.9		9.47	.000
	Matched	1.2088	1.1235	9.2	82.9	1.34	.182
Education	Unmatched	14.317	12.95	48.6		8.14	.000
	Matched	14.317	14.204	4	91.7	.69	.493
Age	Unmatched	39.825	50.644	–66.3		–11.12	.000
	Matched	39.825	39.298	3.2	95.1	.58	.565
Popular Taste	Unmatched	.57	.32	51.60		8.9	.000
	Matched	.57	.57	–1.30	97.50	–.19	.846
Average Parental Education	Unmatched	1.1807	.71248	50.5		8.93	.000
	Matched	1.1807	1.2354	–5.9	88.3	–.84	.403
Education	Unmatched	14.701	12.624	75.0		12.96	.000
	Matched	14.701	14.639	2.3	97.0	.35	.723
Age	Unmatched	46.887	45.968	5.3		.91	.364
	Matched	46.887	45.844	6.0	–13.5	.96	.335
Highbrow Taste	Unmatched	.53906	.29453	51.2		8.90	.000
	Matched	.53906	.55664	–3.7	92.8	–.56	.572