Development without Representation: Replication

Benjamin Hoffner-Brodsky

5/15/2020

Abstract

Jensenius (2015) demonstrates that, between 1971 and 2001, Indian assembly constituencies legally mandated to elect candidates belonging to a Scheduled Caste did not exhibit significantly different development indicators compared with general constituencies. I successfully replicated Jensenius' results. I propose four alternate models for matching Scheduled Caste (SC) constituencies with general constituencies using a multivariate matching algorithm to mitigate selection bias attributed to treatment assignment and selection bias occurring from unmatched treatment cases. Under two of the four alternate models, I find at least one development indicator with a significant difference across SC and general constituencies. While Jensenius' hypothesis is largely intact, the variance in results dependent on the selection of model illustrates the need for researchers to optimize and justify choices of matching model parameters and mechanisms.

Contents

0.1 Introduction

Since 1950, the Indian Parliament and India's state assemblies have guaranteed a minimum number of seats to Scheduled Castes (SCs). Ensuring ascriptive representation for the 16% of Indian citizens who belong to SCs was intended, in part, as a means to equitably allocate resources along caste lines (Jensenius 2015). To implement SC quotas, the Government of India (GoI) non-randomly selected constituencies in which only SC members can run for office, though all members of the constituency are allowed to vote. As mandated in the Delimitation Act of 1972, the quotas were generally assigned to the constituency with the largest proportion of SCs in an eligible district and prevented multiple contiguous constituencies from receiving quotas.

Jensenius forms pairs of reserved and non-reserved constituencies that fall within the same state, district, and parliamentary constituency. As there are 2,409 general constituencies and only 483 reserved constituencies in the dataset, Jensenius pairs treatment cases with the non-treatment case that is most similar in proportion of SCs, without replacement. In the first matching model, all pairs satisfying geographic requirements are preserved, while in the second model a caliper is applied, such that only pairs for which the difference between the proportion of SCs in the treatment and non-treatment case is less than or equal to 0.5 standard deviations. Jensenius uses a constituency-level dataset of 3,134 state assembly constituencies from the 15 largest Indian states with development data from 1971 and 2001. She finds a null constituency-level effect on overall development, redistribution to SCs, literacy rates, SC employment patterns, and village amenities.

I successfully replicated Jensenius' results and failed to reject the null hypothesis for a difference in the means of overall development, redistribution to SCs, literacy rates, SC employment patterns, and village amenities. I was able to replicate all tables and figures in Jensenius (2015) without discrepancy. I accessed the replication materials from Jensenius' personal website, which includes all data required for replication, a codebook for the dataset, and the R code necessary to replicate the majority of Jensenius' results.[1] I used R for replication and analysis (R Core Team 2019).

I examine the robustness Leamer and Leonard (1983) of Jensenius' results relative to the construct of the matching model by proposing two changes to the matching process. First, I match constituencies with replacement, allowing for multiple treatment cases to be paired to the same non-treatment case if doing so would decrease the difference in proportion SCs across the two constituencies. Matching with replacement makes pairs more similar across pre-treatment covariates, eliminates the significance of constituency ordering, and, critically, decreases the number of unmatched treatment cases. Leaving all other aspects of Jensenius' analysis unchanged, allowing for replacement in the matching method decreases the number of unmatched treatment cases by 20% in the closest match model and by 6.9% in the caliper model. In two of the models matched with replacement, I find a difference in mean development indicators across the two groups that is significant at a 5% level.

Second, I apply the recommendation coming out of Wang et al. (2013)'s Monte Carlo simulations to optimally select a caliper value of 1.62. In-line with the recommendation of Rosenbaum and Rubin (1985) a larger caliper value ameliorates the selection bias created by the systematic differences between cases which are able to be matched and those which are not. Holding all other parameters unchanged, increasing the caliper value from 0.5 to 1.62 decreases the number of unmatched treatment cases by 66% while increasing the median p-values for the difference between matched groups, indicating that the matched groups were also statistically more similar.

0.2 Literature Review

SCs have faced ritual discrimination and social exclusion in South Asia since well before the existence of India as a nation (Galanter 1984). Traditionally viewed as 'untouchable', SCs occupy the lowest rung on India's social leader and have been systematically excluded from positions of power and social influence through the caste system (Thorat and Dubey 2012). Following India's independence from Great Britain in 1947, the founding members of the GoI wrote SC protections into India's constitution. SC electoral quotas followed

[1] All data and replication materials were kindly made public by Francesca R. Jensenius, Professor of Political Science at the University of Oslo and Senior Research Fellow at the Norwegian Institute of International Affairs.

many years of debate and negotiation about the appropriate mechanism to grant constitutional rights to representation to SC communities while attempting to bridge, rather than polarize, the divide.

The resulting quota system implemented through the Delimitation Act of 1972 signified a compromise, in which electoral slots were reserved for SC members while still necessitating that SC politicians be elected by non-SC members in their constituency. Dr. Bhim Rao Ambedkar, the renowned SC leader and quota lobbyist, notoriously viewed the Delimitation Act of 1972 as a failure. He has commented that "the result is that the legislator of the minority elected to the reserved seat instead of being a champion of the minority is really a slave of the majority" (Jensenius 2017).

Other researchers have probed the effects of other forms of SC quotas. Dunning and Nilekani (2013) analyzed village-level SC electoral quotas and found no evidence of SC politicians reallocating resources to SC individuals in their constituency. Pande (2003) explored the state assembly quotas and found no changes in spending habits at a state level using a regression discontinuity design. Yet no other researchers prior to or after Jensenius have delved into the effects of state assembly quotas on constituency-level development indicators or resource allocation.

0.3 Replication

Jensenius (2015) stands alone both in its treatment of the constituency-level effects of state assembly quotas and its application of multivariate pre-treatment covariate matching to estimate the average treatment effect on the treated (ATT). To best mitigate treatment selection bias, Jensenius matches treated and non-treated cases only on the pre-treatment variables that government officials in 1972 would have considered. After matching cases, she then clusters standard errors for the difference in development indicator means at the state, district, and parliamentary constituency levels. Under all three calculations, she fails to find a significant difference at the 5% level in development indicators across the treated and untreated groups, indicating a null ATT.

To operationalize the metric of development, Jensenius considers individual-level indicators with readily available data that a motivated politician could reasonably impact. Consequently, she focuses on literacy rates, employment, and agricultural labor.

Jensenius also considers village-level development indicators, taking advantage of the overlap between assembly constituencies and villages. To measure village development, she evaluates the availability of electricity, schools, medical facilities, and communication channels in villages, with increases corresponding to a higher level of village development. In all of these indicators, Jensenius finds no significant ATT.

0.4 Extension

Leamer and Leonard (1983) cautions that the flexibility economics researchers have in selecting econometric models, combined with the tendency of economics journals to select for certain categories of outcomes, can result in an equilibrium in which an apparently robust result should be expected to be a result of model selection rather than underlying statistical truth. They recommend that "researchers be given the task of identifying interesting families of alternative models and be expected to summarize the range of inferences which are implied by each of the families". It is in the spirit of expanding the families of alternative models that I propose two alterations to Jensenius (2015)'s, the combinations of which result in four alternate models to consider. Results from these alternate models consistent with Jensenius' results would render her conclusions more robust to small changes in econometric model choices, while divergent results would necessitate further exploration to determine the most appropriate choice of models.

The first modification I propose to Jensenius (2015)'s matching methodology is matching with replacement. Three advantages of matching with replacement over matching without replacement motivate this amendment.

First, sampling without order is sensitive to changes in the order of both treated and non-treated cases. When sampling without replacement, there may be multiple non-treated cases that could be assigned to the same treated case. In Sekhon and Grieve (2012), the R package Jensenius uses for creating pairs and assessing covariate balance, the potential for multiple non-treatment units to be assigned to the same treatment unit is handled by pairing whichever non-treatment unit is ordinally first. Given that there is no 'correct' ordering of observations, and that differences in order will correspond to differences in regression coefficients and significance tests, matching without replacement renders the analysis vulnerable both to randomness and to Leamer and Leonard (1983)'s concerns regarding researcher discretion. Sampling with replacement comprehensively resolves this issue. When sampling with replacement, multiple non-treatment units can be assigned to the same treatment unit, negating the role of order on groupings, minimizing variance, and avoiding researcher selection bias.

Second, matching with replacement results in closer matches on inexact pretreatment covariates. Under sampling without replacement, a non-treatment unit may be assigned to the 2nd, 3rd, or 4th closest treatment unit, rather than the 1st closest, because the closest was already paired. Smaller differences in pretreatment covariates means that treated and non-treated units become more directly comparable and the effect of selection bias is minimized.

Third, matching with replacement decreases the number of unmatched units. When matching without replacement, there will be some units that have an appropriate match in the dataset but are dropped because the only appropriate match/es were already paired. Under matching with replacement, those same units need not be dropped. Dropped units have an outsized negative effect on bias because they are excluded as a function of pre-treatment covariate values. As Rosenbaum and Rubin (1985) states: "Discarding treated [cases] in this way can lead to serious biases, since the unmatched treated [cases] differ systematically from the matched treated children". Holding all else constant, matching with replacement decreases the number of unmatched treatment cases by 20% in the closest match model and by 6.9% in the caliper model.

Matching without replacement is not without drawbacks. Principally, replacing units decreases the effective sample size (ESS) because fewer observations are responsible for explaining a large proportion of the data. Yet this limitation is mitigated by the strict regional matching, with prevents a small number of units from being matched with too large a proportion of the data.

The second modification I propose is the modification of the caliper value used as the maximum standardized distance in proportion of SCs between matched units. Jensenius (2015) employs a 0.5 standard deviation caliper, a critical parameter choice which is not justified in the paper. There is no obviously correct choice of caliper values, as a larger caliper value preserves a large proportion of cases, while a smaller caliper value results in treatment and control groups with more similar pre-treatment covariates, both of which are desirable elements to minimize bias.

To assess the tradeoff, I apply Wang et al. (2013)'s methodology to select a caliper value of 1.62 standard deviations. The increased caliper value decreases the number of unmatched treatment cases by 66% while increasing the median p-values for the difference between matched groups, indicating that the matched groups were statistically more similar, likely because of the decrease in dropped units and corresponding increase in ESS.

Table 1 shows the number of dropped unmatched units under each matching model. A larger number indicates a greater danger of selection bias, as units are systematically dropped based on pre-treatment covariates. Predictably, matching with replacement and without a caliper yields the lowest number of dropped units, 28. The most severe results comes from Jensenius (2015)'s model of matching without replacement and with a 0.5 standard deviation caliper which drops 159 units, or 32.9% of all treated units. Though insufficient evidence to be conclusive, the lower number of dropped units under the proposed alternate models could indicate an appropriate fit.

Table 1: Number of Unmatched Cases

	Without Replacement	With Replacement
No caliper	35	28
0.5 caliper	159	148
0.2 * sd caliper	54	42

Joining my proposed alterations with Jensenius (2015)'s models results in two parameter options for matching, each with three choices, for a total of six possible combinations of replacement and caliper. Matching can either be done with or without replacement and a caliper can be 0.5, 1.62, or not applied.

For easier reading, the six different models have been assigned the following numbering:

Model 1: Matching without replacement | No caliper Model 2: Matching with replacement | No caliper
Model 3: Matching without replacement | 0.5 caliper Model 4: Matching with replacement | 0.5 caliper
Model 5: Matching without replacement | 1.62 caliper Model 6: Matching with replacement | 1.62 caliper

Table 2 shows the resulting covariate balance across treated and non-treated groups for each of the six matching models, compared against the pre-matching balance. For each model, two significance tests are displayed: a paired t-test and a bootstrapped Kolmogorov-Smirnov (KS) test—the same tests used by Jensenius (2015)—both of which test the hypothesis that the two groups have different means. For all six models, both significance tests, and all eight covariates, there is no evidence of a significant difference between means. For some models, however, this conclusion is more robust than others. Model 2—matching with replacement and no caliper—reports the highest aggregate mean and median of t-test p-values, while Model 5—matching without replacement and a 1.62 standard deviation caliper—reports the highest aggregate of KS p-values. A higher p-value is one indication of a more statistically robust match along pre-treatment covariates.

Table 2: Difference in Means for Control and Treatment Groups

Covariate	Before matching		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	t *p*-value	KS *p*-value	*t*	KS	*t*	KS	*t*	KS	*t*	KS	*t*	KS	*t*	KS
Population size	0.11	0.01	0.16	0.93	0.13	0.58	0.41	0.75	0.46	0.72	0.1	0.88	0.12	0.62
Percentage of STs	0.52	0.14	0.62	0.97	0.53	0.68	0.53	0.9	0.5	0.86	0.37	0.94	0.4	0.84
Literacy rate (non-SCs)	0	0.01	0.44	0.96	0.93	0.87	0.93	0.91	0.97	0.68	0.58	0.96	0.91	0.79
Literacy rate (SCs)	0	0	0.87	0.93	0.97	0.87	0.61	0.99	0.44	1	0.85	0.95	0.91	0.81
Employment (non-SCs)	0.19	0.23	0.39	0.78	0.35	0.6	0.96	0.81	0.98	0.72	0.53	0.97	0.51	0.8
Employment (SCs)	0.31	0.05	0.67	0.64	0.9	0.74	0.85	0.92	0.98	0.95	0.64	0.63	0.65	0.68
Agricultural laborers (non-SCs)	0.01	0.01	0.71	0.99	0.79	0.89	0.5	1	0.6	0.98	0.92	0.96	0.98	0.92
Agricultural laborers (SCs)	0.54	0.87	0.76	0.99	0.49	0.98	0.39	0.98	0.41	0.95	0.9	0.98	0.52	0.97
Mean	0.21	0.17	0.58	0.9	0.64	0.78	0.65	0.91	0.67	0.86	0.61	0.91	0.62	0.8
Median	0.15	0.03	0.64	0.95	0.66	0.81	0.57	0.91	0.55	0.91	0.61	0.96	0.58	0.8

Having verified that the covariate balance was successful, I turn to comparing the regression results. Tables 3 and 4 show the results of regressing each outcome variable of interest on treatment status after matching. For each covariate, both the direct difference and the difference in difference effect, denoted by the term "gap" is shown. These tables mirror Figure 3 from Jensenius (2015), reproduced in the appendix, to allow for an easy comparison. Most noticeably, unlike in Models 1 and 3, under which Jensenius finds no differences significant at a 5% level, here Model 4 has a 0.05 p-value for the difference coefficient of literacy rate and Model 2 has a 0.05 p-value for the difference coefficient of having a medical facility in a village. Though both of these point towards a non-null ATT, contrary to Jensenius (2015)'s findings, the results should be considered with skepticism. Both barely cross the arbitrary 0.05 threshold and, given the sharp uptick in the number of regressions due to tripling the number of models being considered, it wouldn't be surprising to find seemingly significant effects even if the ATT is null. These results should be interpreted as opening the possibility that alternate models could yield a non-null ATT, rather than independently establishing a non-null ATT.

Table 3: Models 2 and 4 Treatment Regression

Covariate	Before Matching		Model 2		Model 4	
	Difference	P-value	Difference	P-value	Difference	P-value
Literacy rate	-0.02	0.95	0.19	0.61	0.65	0.05
Employment rate	-0.08	0.70	-0.26	0.31	0.09	0.79
Agricultural laborers	-0.02	0.91	-0.28	0.22	-0.31	0.34
Electricity in village	-1.41	0.05	-0.93	0.14	-0.16	0.78
School in village	-0.45	0.20	-0.49	0.22	-0.18	0.71
Medical facility in village	-0.42	0.50	-1.09	0.05	-0.41	0.47
Comm. channel in village	-0.27	0.72	-0.22	0.80	0.96	0.27
Literacy gap	-0.08	0.64	-0.04	0.87	0.30	0.23
Employment gap	0.05	0.62	-0.03	0.83	-0.15	0.31
Agricultural laborers gap	-0.01	0.93	-0.13	0.41	-0.22	0.34
Electricity in village gap	0.09	0.41	0.22	0.24	0.24	0.15
School in village gap	-0.10	0.40	-0.07	0.61	-0.10	0.44
Medical facility in village gap	-0.12	0.75	0.36	0.27	0.45	0.17
Comm. channel in village gap	-0.35	0.12	-0.26	0.39	-0.14	0.65

Table 4: Models 6 and 5 Treatment Regression

Covariate	Before Matching		Model 6		Model 5	
	Difference	P-value	Difference	P-value	Difference	P-value
Literacy rate	-0.05	0.90	-0.13	0.65	0.63	0.06
Employment rate	-0.01	0.96	-0.05	0.76	0.66	0.09
Agricultural laborers	-0.01	0.97	0.02	0.93	-0.06	0.90
Electricity in village	-1.00	0.11	-0.78	0.29	0.43	0.46
School in village	-0.44	0.19	-0.42	0.20	-0.00	1.00
Medical facility in village	-0.43	0.46	-0.47	0.27	0.14	0.89
Comm. channel in village	-0.30	0.69	-0.45	0.57	1.34	0.09
Literacy gap	-0.14	0.42	-0.12	0.51	0.45	0.10
Employment gap	0.04	0.71	-0.00	0.95	-0.25	0.09
Agricultural laborers gap	0.02	0.88	0.05	0.71	-0.15	0.72
Electricity in village gap	0.24	0.15	0.18	0.31	0.22	0.19
School in village gap	-0.09	0.44	-0.07	0.55	-0.09	0.34
Medical facility in village gap	-0.00	1.00	-0.14	0.61	0.06	0.84
Comm. channel in village gap	-0.36	0.10	-0.32	0.14	-0.02	0.91

0.5 Conclusion

The GoI's non-random assignment of constituencies mandated to have SC state assembly representatives complicates the process of assessing the causal effect of SC quotas on constituency-level development indicators. Since the Delimitation Act of 1972 required reserved constituencies to be geographically spread out and located in constituencies with high proportions of SCs, Jensenius (2015) augments the value of this observational data by matching treated and non-treated cases on geographic location and proportion of SCs.

To address the asymmetry in the number of general and reserved constituencies, Jensenius pairs treatment cases with non-treatment cases without replacement. In her first model, all pairs satisfying geographic requirements are preserved, while in her second model, a maximum distance caliper is applied, such that only pairs which are a maximum of 0.5 standard deviations of proportion of SCs apart from each other are preserved. Under both models, Jensenius fails to reject the null hypothesis that there was no constituency-level ATT on overall development, redistribution to SCs, literacy rates, SC employment patterns, or village amenities.

I successfully replicated Jensenius' results and failed to reject the null hypothesis using both of her models. I then proposed two changes to the matching model, resulting in four new models to assess. First, I match constituencies with replacement, so that multiple treatment cases can be paired with the same non-treatment case. Leaving all else constant, I find that matching with replacement lowers the count of unmatched treatment cases by 20% in the closest match model and by 6.9% in the caliper model. I also find a statistically significant ATT under two models at the 5% level.

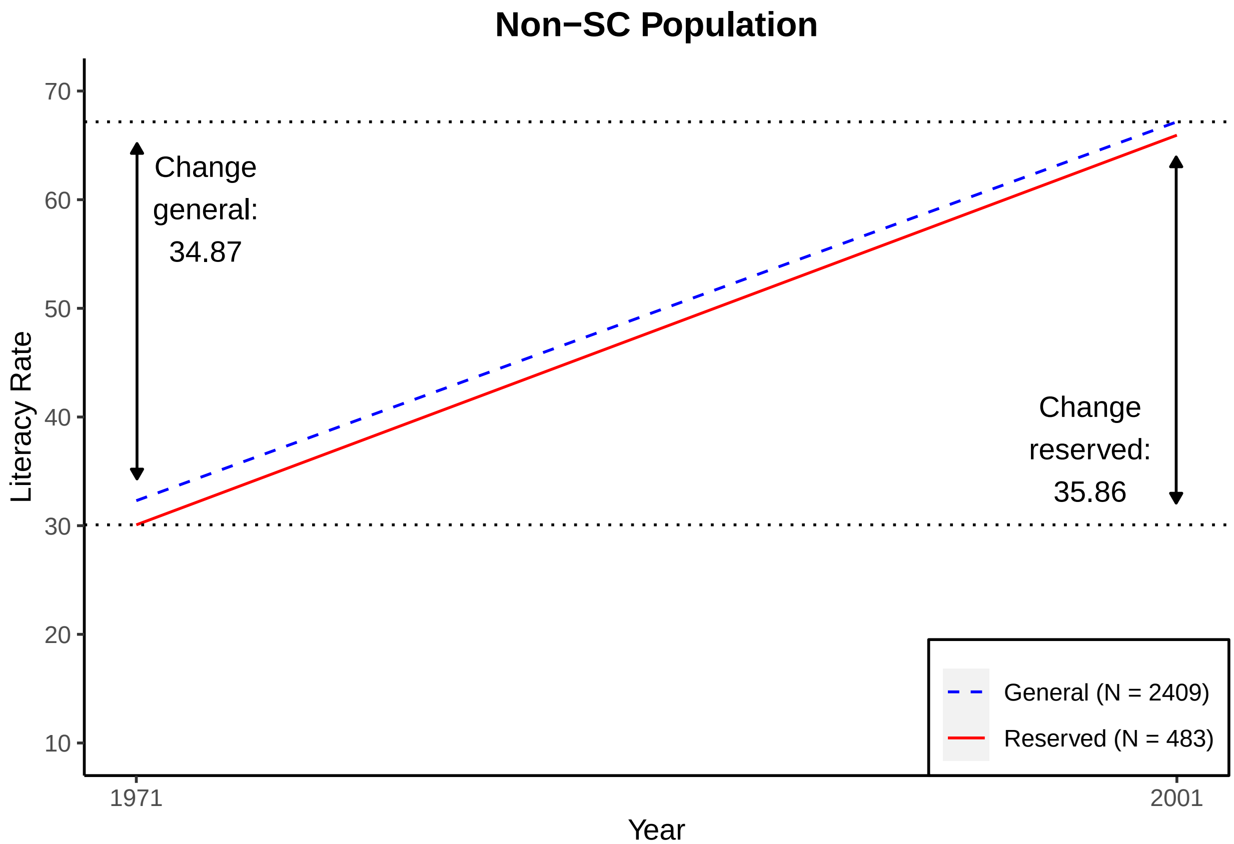
Second, I employ Wang et al. (2013)'s advice to optimally select a caliper value of 1.62. Holding all else constant, widening the caliper from 0.5 standard deviations to 1.62 standard deviations decreases the number of unmatched treatment cases by 66%.

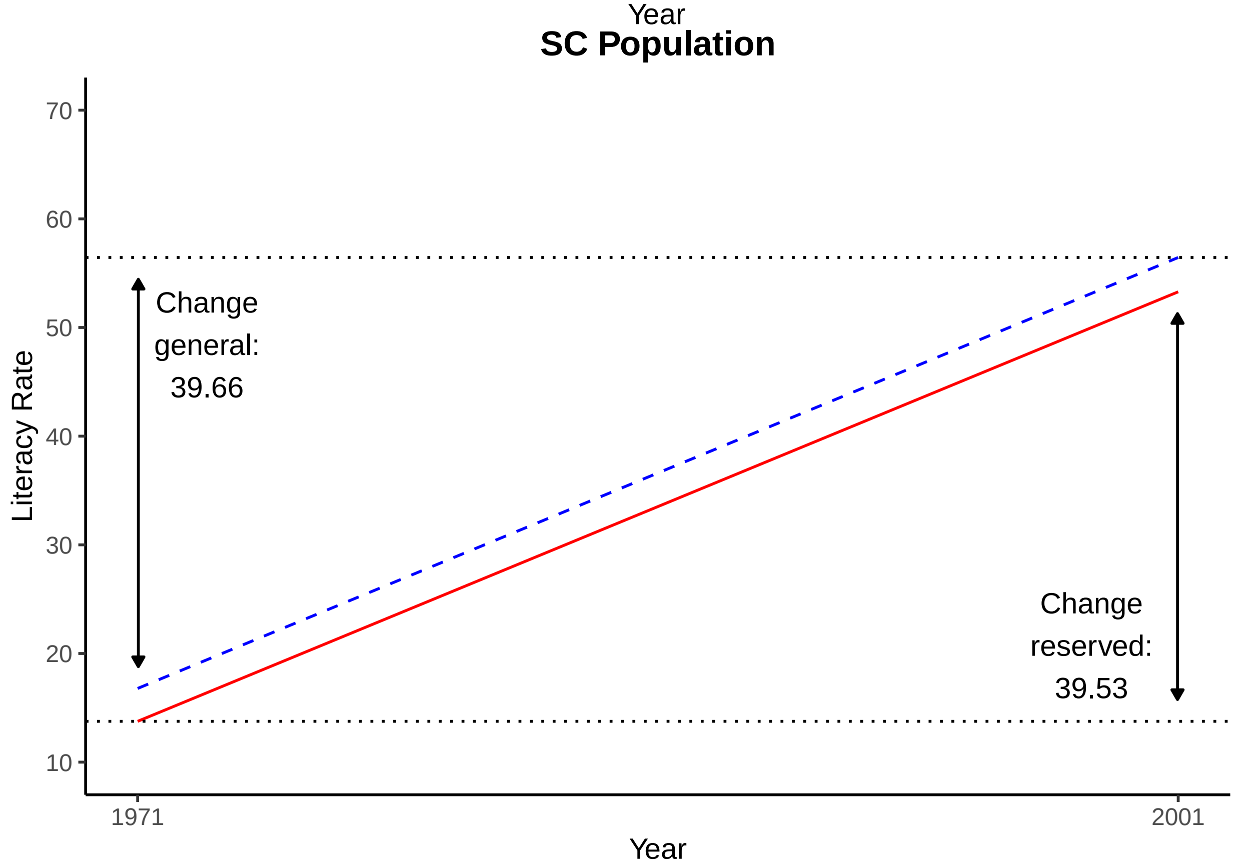
Among the proposed alternate models are evidence of modifications that could result in fewer unmatched cases, better covariate balanced groups, and potentially statistically significant treatment effects. Thus, further research into the optimal choice of matching algorithm is needed.

0.6 Appendix

Jensenius (2015) Table 3: Models 1 and 3 Treatment Regression

Covariate	Before Matching		Model 1		Model 3	
	Difference	P-value	Difference	P-value	Difference	P-value
Literacy rate	-0.18	0.58	0.06	0.85	0.55	0.07
Employment rate	-0.08	0.70	-0.25	0.29	0.15	0.67
Agricultural laborers	0.09	0.69	-0.19	0.46	-0.21	0.55
Electricity in village	-1.19	0.16	-0.79	0.23	0.03	0.97
School in village	-0.32	0.35	-0.53	0.21	-0.21	0.68
Medical facility in village	-0.26	0.53	-1.03	0.10	-0.29	0.64
Comm. channel in village	-0.51	0.50	-0.34	0.73	0.96	0.30
Literacy gap	-0.03	0.88	0.01	0.96	0.38	0.12
Employment gap	0.00	0.96	-0.06	0.64	-0.18	0.20
Agricultural laborers gap	0.02	0.89	-0.04	0.80	-0.13	0.59
Electricity in village gap	0.02	0.84	0.22	0.26	0.26	0.16
School in village gap	-0.06	0.54	-0.01	0.92	-0.04	0.71
Medical facility in village gap	-0.29	0.40	0.33	0.28	0.43	0.16
Comm. channel in village gap	-0.31	0.18	-0.20	0.52	-0.07	0.83


Non−SC Population
Literacy Rate
70
60
50
40
30
20
10
Change
general:
34.87
Change
reserved:
35.86
General (N = 2409)
Reserved (N = 483)
1971
2001
Year


SC Population
Literacy Rate
70
60
50
40
30
20
10
Change
general:
39.66
Change
reserved:
39.53
1971
2001
Year

References

Dunning, Thad, and Janhavi Nilekani. 2013. "Ethnic Quotas and Political Mobilization: Caste, Parties, and Distribution in Indian Village Councils." *American Political Science Review* 107 (1). Cambridge University Press: 35–56.

Galanter, Marc. 1984. "Competing Equalities: Law and the Bacliward Classes in India."

Jensenius, Francesca R. 2017. *Social Justice Through Inclusion: The Consequences of Electoral Quotas in India*. Oxford University Press.

Jensenius, Francesca Refsum. 2015. "Development from Representation? A Study of Quotas for the Scheduled Castes in India." *American Economic Journal: Applied Economics* 7 (3): 196–220.

Leamer, Edward, and Herman Leonard. 1983. "Reporting the Fragility of Regression Estimates." *The Review of Economics and Statistics*. JSTOR, 306–17.

Pande, Rohini. 2003. "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India." *American Economic Review* 93 (4): 1132–51.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rosenbaum, Paul R, and Donald B Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39 (1). Taylor & Francis Group: 33–38.

Sekhon, Jasjeet Singh, and Richard D. Grieve. 2012. "A Matching Method for Improving Covariate Balance in Cost-Effectiveness Analyses." *Health Economics* 21 (6): 695–714.

Thorat, Sukhadeo, and Amaresh Dubey. 2012. "Has Growth Been Socially Inclusive During 1993-94-2009-10?" *Economic and Political Weekly*. JSTOR, 43–53.

Wang, Yongji, Hongwei Cai, Chanjuan Li, Zhiwei Jiang, Ling Wang, Jiugang Song, and Jielai Xia. 2013. "Optimal Caliper Width for Propensity Score Matching of Three Treatment Groups: A Monte Carlo Study." *PloS One* 8 (12). Public Library of Science.