

**How crucial is the role of education and social group for  
Indian female-headed household's well-being? A study  
using nonparametric multivariate first order stochastic  
dominance**

Juan Tellez and Jaya Krishnakumar  
University of Geneva, Switzerland  
presented by  
Jaya Krishnakumar

Workshop on  
*Advances in stochastic dominance for welfare analysis*  
18-19 September 2014  
FERDI, Clermont-Ferrand

- Basic premise: Well being is multidimensional, and needs to be grasped through multiple indicators.

- Basic premise: Well being is multidimensional, and needs to be grasped through multiple indicators.
- Many approaches advocate such a multidimensional vision: quality of life approach, capability approach, living conditions approach, basic needs approach etc.

- Basic premise: Well being is multidimensional, and needs to be grasped through multiple indicators.
- Many approaches advocate such a multidimensional vision: quality of life approach, capability approach, living conditions approach, basic needs approach etc.
- If we go by this premise, then the question arises as to how to combine various indicators for comparing well-being across individuals or over time?

- Basic premise: Well being is multidimensional, and needs to be grasped through multiple indicators.
- Many approaches advocate such a multidimensional vision: quality of life approach, capability approach, living conditions approach, basic needs approach etc.
- If we go by this premise, then the question arises as to how to combine various indicators for comparing well-being across individuals or over time?
- Possible solutions
  - Composite indices (from simple averages to generalised means)
  - Parametric model based aggregates : latent variable scores of well-being (FA, MIMIC, SEM, generalised SEM etc.)
  - Non-parametric comparisons of multivariate distributions and testing for dominance

- In this paper, we apply the third approach to compare welfare distributions across different sub-populations

- In this paper, we apply the third approach to compare welfare distributions across different sub-populations
- Although the use of this methodology to compare univariate distributions is common, extension of the non-parametric dominance technique to the multivariate case is recent (cf. Maasoumi and Racine 2013)

- In this paper, we apply the third approach to compare welfare distributions across different sub-populations
- Although the use of this methodology to compare univariate distributions is common, extension of the non-parametric dominance technique to the multivariate case is recent (cf. Maasoumi and Racine 2013)
- This technique is particularly suitable in our context given the multidimensional definition of well-being that we adopt and allows us to make welfare comparisons without either aggregating over dimensions or over households.



- Our research question : To what extent do social group and education levels contribute to enhancing or reducing welfare of Indian households, and does education offset the effect of social discrimination.

- Our research question : To what extent do social group and education levels contribute to enhancing or reducing welfare of Indian households, and does education offset the effect of social discrimination.
- As there is a variety of socio-economic characteristics that define an Indian household, and as literature has pointed out large differences in behaviour between households headed by men and those headed by women, we decided to only focus on one of the two for a certain homogeneity among the households.

- Our research question : To what extent do social group and education levels contribute to enhancing or reducing welfare of Indian households, and does education offset the effect of social discrimination.
- As there is a variety of socio-economic characteristics that define an Indian household, and as literature has pointed out large differences in behaviour between households headed by men and those headed by women, we decided to only focus on one of the two for a certain homogeneity among the households.
- The reason for selecting female-headed households is that there are only a few studies on them as noted by Gangopadhyay and Wadhwa (2004) and we wanted to add to the scarce literature in this domain.

- 1 Motivation
- 2 Nonparametric Functions
- 3 Testing Stochastic Dominance
- 4 Data
- 5 Results
- 6 Conclusions

- Nonparametric techniques **do not require** a researcher to specify a **functional form** for an object being estimated. Let us note the density function of a continuous random variable  $X_j^c$  as  $f(x_j^c)$ .

$$\hat{f}(x_j^c) = \frac{1}{nh_j} \sum_{i=1}^n k\left(\frac{x_j^c - X_{ij}^c}{h_j}\right),$$

- The following assumptions are made for  $k(\cdot)$ :
  - 1  $k(u) = k(-u)$
  - 2  $\int k(u)du = 1$
  - 3  $\int uk(u)du = 0$
  - 4  $\int u^2k(u)du = v_2 < \infty$

Conditional density functions and cumulative distribution functions are expressed as:

$$\hat{f}(y|x) = \frac{\hat{f}(y, x)}{\hat{f}(x)}$$

$$\hat{F}(y|x) = \frac{\int^y \hat{f}(s, x) ds}{\hat{f}(x)}$$

Conditional density functions and cumulative distribution functions are expressed as:

$$\hat{f}(y|x) = \frac{\hat{f}(y, x)}{\hat{f}(x)}$$

$$\hat{F}(y|x) = \frac{\int^y \hat{f}(s, x) ds}{\hat{f}(x)}$$

where

$$\hat{f}(y, x) = n^{-1} \sum_{i=1}^n K_{\gamma_y}(Y_i, y) K_{\gamma_x}(X_i, x),$$

$$\hat{f}(x) = n^{-1} \sum_{i=1}^n K_{\gamma_x}(X_i, x),$$

and

$$\int^y \hat{f}(s, x) ds = n^{-1} \sum_{i=1}^n G_{\gamma_y}(Y_i, y) K_{\gamma_x}(X_i, x),$$

- Let us note two distributions  $A$  and  $B$ , with respective CDFs  $F_A$  and  $F_B$ . We will say that  $B$  **dominates**  $A$  **stochastically at first order** if for any  $r$ :

$$F_A(r) \geq F_B(r)$$



- Let us note two distributions  $A$  and  $B$ , with respective CDFs  $F_A$  and  $F_B$ . We will say that  $B$  **dominates**  $A$  **stochastically at first order** if for any  $r$ :

$$F_A(r) \geq F_B(r)$$

- In order to test the multivariate stochastic dominance we will consider the following functions:

$$F_A = F(y|x = a)$$

$$F_B = F(y|x = b)$$

- Let us note two distributions  $A$  and  $B$ , with respective CDFs  $F_A$  and  $F_B$ . We will say that  $B$  **dominates  $A$  stochastically at first order** if for any  $r$ :

$$F_A(r) \geq F_B(r)$$

- In order to test the multivariate stochastic dominance we will consider the following functions:

$$F_A = F(y|x = a)$$

$$F_B = F(y|x = b)$$

- We will use the following **Kolmogorov-Smirnov (KS)** statistic as:

$$D = \min[\max(F_A - F_B), \max(F_B - F_A)]$$

- Let us note two distributions  $A$  and  $B$ , with respective CDFs  $F_A$  and  $F_B$ . We will say that  $B$  **dominates  $A$  stochastically at first order** if for any  $r$ :

$$F_A(r) \geq F_B(r)$$

- In order to test the multivariate stochastic dominance we will consider the following functions:

$$F_A = F(y|x = a)$$

$$F_B = F(y|x = b)$$

- We will use the following **Kolmogorov-Smirnov (KS)** statistic as:

$$D = \min[\max(F_A - F_B), \max(F_B - F_A)]$$

- Assumptions:

$$H_0 = D > 0$$

$$H_a = D \leq 0$$

- Our data come from the third **National Family Health Survey (NFHS-3)**, which was conducted in two phases, the first from December 2005 to April 2006, and the second from April 2006 to August 2006 by the Ministry of Health and Family Welfare with the collaboration of eighteen research organizations giving a representative sample for India. We have 6'086 woman-headed households in our sample.

- Our data come from the third **National Family Health Survey (NFHS-3)**, which was conducted in two phases, the first from December 2005 to April 2006, and the second from April 2006 to August 2006 by the Ministry of Health and Family Welfare with the collaboration of eighteen research organizations giving a representative sample for India. We have 6'086 woman-headed households in our sample.
- We take two dimensions of well-being : an economic indicator (*wealth index*), and a health indicator (*level of hemoglobin*).

- Our data come from the third **National Family Health Survey (NFHS-3)**, which was conducted in two phases, the first from December 2005 to April 2006, and the second from April 2006 to August 2006 by the Ministry of Health and Family Welfare with the collaboration of eighteen research organizations giving a representative sample for India. We have 6'086 woman-headed households in our sample.
- We take two dimensions of well-being : an economic indicator (*wealth index*), and a health indicator (*level of hemoglobin*).
- We compare the bivariate distribution (*wealth, health*) among different sections characterised by different *levels of education* and *social groups* in order to understand the role played by these factors in the determination of well-being.

- Our data come from the third **National Family Health Survey (NFHS-3)**, which was conducted in two phases, the first from December 2005 to April 2006, and the second from April 2006 to August 2006 by the Ministry of Health and Family Welfare with the collaboration of eighteen research organizations giving a representative sample for India. We have 6'086 woman-headed households in our sample.
- We take two dimensions of well-being : an economic indicator (*wealth index*), and a health indicator (*level of hemoglobin*).
- We compare the bivariate distribution (*wealth, health*) among different sections characterised by different *levels of education* and *social groups* in order to understand the role played by these factors in the determination of well-being.

## Three dominance tests:

1

$F(\text{wealth, hemoglobin} | \text{education} = 2, \text{group} = j)$

$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = j)$



## Three dominance tests:

1

$$F(\text{wealth, hemoglobin} | \text{education} = 2, \text{group} = j)$$

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = j)$$

2

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = j)$$

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = i)$$

## Three dominance tests:

1

$$F(\text{wealth, hemoglobin} | \text{education} = 2, \text{group} = j)$$

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = j)$$

2

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = j)$$

$$F(\text{wealth, hemoglobin} | \text{education} = 10, \text{group} = i)$$

3

$$F(\text{wealth, hemoglobin} | \text{education} = 2, \text{group} = j)$$

$$F(\text{wealth, hemoglobin} | \text{education} = 2, \text{group} = i)$$

Figure: Nonparametric estimated CDF difference of Wealth and Health

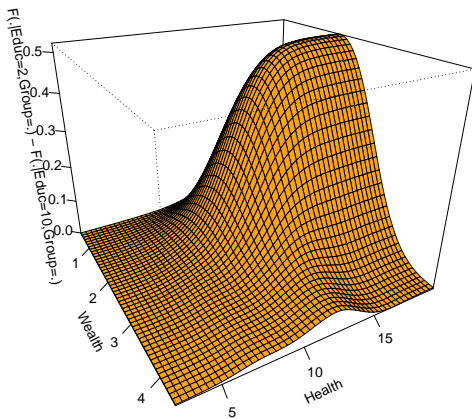


Figure: Contourlines of Wealth and Health for high and low education levels

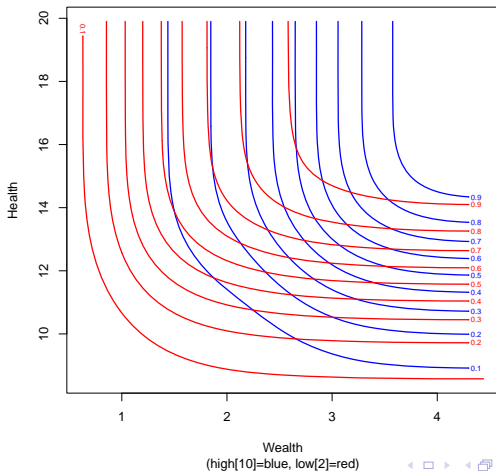


Table: Dominance results for each social group

Group	Range	D	P-value	q(0.01)	q(0.05)	q(0.1)
<i>Scheduled Castes</i>						
Smoothed	1	-0.0036	0.0375	-0.0069	-0.0012	0.0017
	0.9	-0.0236	0.0025	-0.0165	-0.0086	-0.0048
Empirical	1	0.0481	0.0576	0.0182	0.0464	0.0653
	0.9	0.0324	0.0751	-0.0292	0.0236	0.0436
Frequency	1	0.0109	0.1303	-0.0149	-0.0040	0.0056
	0.9	-0.0106	0.1704	-0.0537	-0.0359	-0.0236
<i>Scheduled Tribes</i>						
Smoothed	1	-0.0431	0.0000	-0.0133	-0.0067	-0.0035
	0.9	-0.0566	0.0025	-0.0236	-0.0151	-0.0094
Empirical	1	0.1316	0.5313	-0.0741	-0.0147	0.0349
	0.9	0.0239	0.1278	-0.0682	-0.0123	0.0140
Frequency	1	0.0018	0.1779	-0.0721	-0.0367	-0.0172
	0.9	-0.0142	0.2330	-0.1083	-0.0647	-0.0452
<i>Other backward classes</i>						
Smoothed	1	-0.0156	0.0000	-0.0056	-0.0016	0.0016
	0.9	-0.0268	0.0000	-0.0149	-0.0069	-0.0035
Empirical	1	-0.0252	0.0075	-0.0146	0.0234	0.0371
	0.9	-0.0150	0.0075	-0.0113	0.0104	0.0230
Frequency	1	-0.0127	0.0350	-0.0246	-0.0073	-0.0023
	0.9	-0.0320	0.0350	-0.0483	-0.0277	-0.0201
<i>None of above</i>						
Smoothed	1	-0.0135	0.0000	-0.0059	-0.0007	0.0019
	0.9	-0.0427	0.0000	-0.0132	-0.0079	-0.0050
Empirical	1	-0.0359	0.0000	0.0145	0.0397	0.0501
	0.9	-0.0658	0.0000	-0.0249	0.0019	0.0237
Frequency	1	-0.0112	0.0325	-0.0221	-0.0063	0.0013
	0.9	-0.0682	0.0000	-0.0468	-0.0282	-0.0187

Table: Dominance Results for high education between social groups

Group	Range	D	P-value	q(0.01)	q(0.05)	q(0.1)
<i>Scheduled Castes vs Scheduled Tribes</i>						
Smoothed	1	0.0008	0.0375	-0.0019	0.0019	0.0051
	0.9	-0.0053	0.0300	-0.0086	-0.0047	-0.0020
Empirical	1	0.0654	0.1253	0.0167	0.0435	0.0609
	0.9	0.0414	0.1929	0.0046	0.0154	0.0290
Frequency	1	0.0037	0.0827	-0.0094	0.0011	0.0076
	0.9	-0.0178	0.0350	-0.0276	-0.0136	-0.0069
<i>Scheduled Castes vs Other backward classes</i>						
Smoothed	1	-0.0002	0.0275	-0.0019	0.0020	0.0048
	0.9	-0.0001	0.1228	-0.0066	-0.0030	-0.0009
Empirical	1	0.0598	0.2155	0.0222	0.0356	0.0453
	0.9	0.0547	0.3984	0.0030	0.0199	0.0299
Frequency	1	0.0064	0.1077	-0.0060	0.0008	0.0059
	0.9	-0.0068	0.0426	-0.0155	-0.0064	-0.0013
<i>Scheduled Castes vs None of above</i>						
Smoothed	1	-0.0025	0.0175	-0.0034	0.0002	0.0027
	0.9	-0.0115	0.0025	-0.0072	-0.0044	-0.0019
Empirical	1	0.0170	0.0025	0.0238	0.0350	0.0460
	0.9	0.0140	0.0501	-0.0089	0.0149	0.0239
Frequency	1	-0.0025	0.0275	-0.0059	0.0016	0.0063
	0.9	-0.0027	0.1378	-0.0137	-0.0092	-0.0056
<i>Scheduled Tribes vs Other backward classes</i>						
Smoothed	1	0.0786	1.0000	-0.0041	0.0006	0.0033
	0.9	0.0232	0.9323	-0.0089	-0.0043	-0.0026
Empirical	1	0.1310	0.8621	0.0268	0.0442	0.0566
	0.9	0.0365	0.1478	-0.0098	0.0209	0.0300
Frequency	1	0.1084	1.0000	-0.0078	0.0016	0.0060
	0.9	0.0062	0.3283	-0.0233	-0.0115	-0.0070
<i>Scheduled Tribes vs None of above</i>						
Smoothed	1	0.0601	0.9949	-0.0056	0.0000	0.0027
	0.9	0.0063	0.4511	-0.0109	-0.0049	-0.0024
Empirical	1	0.1794	0.9874	0.0127	0.0403	0.0501
	0.9	0.0306	0.1654	-0.0101	0.0056	0.0202
Frequency	1	0.0988	0.9949	-0.0111	-0.0006	0.0058
	0.9	0.0142	0.5087	-0.0244	-0.0144	-0.0090
<i>Other backward classes vs None of above</i>						
Smoothed	1	0.0011	0.0501	-0.0029	0.0013	0.0036
	0.9	-0.0001	0.1604	-0.0065	-0.0045	-0.0018
Empirical	1	0.0199	0.0125	0.0189	0.0312	0.0396
	0.9	0.0199	0.0927	0.0007	0.0126	0.0220
Frequency	1	0.0134	0.2581	-0.0038	0.0016	0.0049
	0.9	0.0119	0.5588	-0.0118	-0.0062	-0.0033

# Table: Dominance Results for low education between social groups

Group	Range	D	P-value	q(0.01)	q(0.05)	q(0.1)
<i>Scheduled Castes vs Scheduled Tribes</i>						
Smoothed	1	0.0095	0.5538	-0.0029	0.0000	0.0015
	0.9	0.0092	0.7192	-0.0041	-0.0022	-0.0002
Empirical	1	0.2569	0.8796	0.0168	0.0524	0.0799
	0.9	0.2314	0.8120	0.0416	0.06944	0.0820
Frequency	1	0.1357	0.9122	-0.0148	-0.0033	0.0044
	0.9	0.1236	0.9448	-0.0226	-0.0101	-0.0022
<i>Scheduled Castes vs Other backward classes</i>						
Smoothed	1	0.0123	0.7619	-0.0014	0.0003	0.0018
	0.9	0.0123	0.8771	-0.0031	-0.0018	-0.0005
Empirical	1	0.1616	0.6441	0.0312	0.0552	0.0738
	0.9	0.1743	0.6867	0.0232	0.0544	0.0726
Frequency	1	0.0525	0.5764	-0.0060	0.0014	0.0085
	0.9	0.0525	0.6616	-0.0172	-0.0070	-0.0018
<i>Scheduled Castes vs None of above</i>						
Smoothed	1	-0.0008	0.0275	-0.0014	0.0001	0.0015
	0.9	-0.0054	0.0000	-0.0040	-0.0013	0.0002
Empirical	1	0.1324	0.4010	0.0211	0.0571	0.0749
	0.9	0.1564	0.5789	0.0232	0.0511	0.0723
Frequency	1	0.0783	0.7919	-0.0121	0.0000	0.0063
	0.9	0.0670	0.8070	-0.0253	-0.0104	-0.0046
<i>Scheduled Tribes vs Other backward classes</i>						
Smoothed	1	0.0044	0.2155	-0.0011	0.0009	0.0022
	0.9	0.0041	0.3734	-0.0032	-0.0014	-0.0004
Empirical	1	0.2759	0.9498	0.0286	0.0618	0.0800
	0.9	0.2515	0.9548	0.0039	0.0463	0.0731
Frequency	1	0.1152	0.8922	-0.0186	-0.0004	0.0094
	0.9	0.1148	0.9649	-0.0238	-0.0137	-0.0063
<i>Scheduled Tribes vs None of above</i>						
Smoothed	1	0.0037	0.2230	-0.0027	-0.0001	0.0010
	0.9	0.0028	0.2731	-0.0060	-0.0028	-0.0006
Empirical	1	0.2625	0.9473	0.0120	0.0502	0.0794
	0.9	0.3250	0.9924	0.0137	0.0556	0.0760
Frequency	1	0.1471	0.9674	-0.0166	-0.0071	0.0007
	0.9	0.1429	0.9924	-0.0365	-0.0185	-0.0100
<i>Other backward classes vs None of above</i>						
Smoothed	1	-0.0003	0.0526	-0.0018	-0.0003	0.0011
	0.9	-0.0011	0.0551	-0.0037	-0.0012	-0.0001
Empirical	1	0.1365	0.6441	0.0195	0.0444	0.0657
	0.9	0.1597	0.7969	0.0115	0.0500	0.0626
Frequency	1	0.0443	0.5739	-0.0137	-0.0034	0.0036
	0.9	0.0324	0.5388	-0.0220	-0.0090	-0.0031

- For all four social groups - **scheduled castes, scheduled tribes, other backward castes, none of the above**, woman-headed households with high level of education stochastically dominate at first order women with low level of education.



- For **all four social groups - scheduled castes, scheduled tribes, other backward castes, none of the above**, woman-headed households with high level of education stochastically dominate at first order women with low level of education.
- For woman-headed households with **low level of education**, the group *None of the above* i.e. the 'non-backward' group stochastically dominates *Scheduled Castes* and *Other Backward Classes* at the first order.

- For **all four social groups - scheduled castes, scheduled tribes, other backward castes, none of the above**, woman-headed households with high level of education stochastically dominate at first order women with low level of education.
- For woman-headed households with **low level of education**, the group *None of the above* i.e. the 'non-backward' group stochastically dominates *Scheduled Castes* and *Other Backward Classes* at the first order.
- For woman-headed households with **high level of education**, the group *Scheduled Castes* is stochastically dominated at first order by all the other three social groups.

- Our results confirm the **strong role played by education in improving well-being**, even in presence of other unfavourable factors.

- Our results confirm the **strong role played by education in improving well-being**, even in presence of other unfavourable factors.
- Households belonging to ‘backward’ castes still face **inequality of opportunity** even if the caste system has been legally abolished since independence, as the ‘non-backward’ castes seem to systematically dominate the ‘backward’ castes, at any level of education.

- Our results confirm the **strong role played by education in improving well-being**, even in presence of other unfavourable factors.
- Households belonging to 'backward' castes still face **inequality of opportunity** even if the caste system has been legally abolished since independence, as the 'non-backward' castes seem to systematically dominate the 'backward' castes, at any level of education.
- Education **is** an important contributor to welfare improvement, but it does not remove all the negatives of social discrimination as even among people with a high level of education, the 'lowest' caste is still dominated by the other three groups. However, it does offset the effect of discrimination to a significant extent as there is no dominance among the other three groups when the education level is high whereas one still finds dominance among these three groups (in favour of the 'non-backward' group) for a low level of education.

- Our results confirm the **strong role played by education in improving well-being**, even in presence of other unfavourable factors.
- Households belonging to 'backward' castes still face **inequality of opportunity** even if the caste system has been legally abolished since independence, as the 'non-backward' castes seem to systematically dominate the 'backward' castes, at any level of education.
- Education **is** an important contributor to welfare improvement, but it does not remove all the negatives of social discrimination as even among people with a high level of education, the 'lowest' caste is still dominated by the other three groups. However, it does offset the effect of discrimination to a significant extent as there is no dominance among the other three groups when the education level is high whereas one still finds dominance among these three groups (in favour of the 'non-backward' group) for a low level of education.

- From a methodological point of view, the clear **multivariate dominance** results show that one group is less well-off than another irrespective of where the poverty line may be set in both dimensions as our method compares entire distributions rather than certain derived measures like percentiles.

- From a methodological point of view, the clear **multivariate dominance** results show that one group is less well-off than another irrespective of where the poverty line may be set in both dimensions as our method compares entire distributions rather than certain derived measures like percentiles.
- The **nonparametric approach** allowed us to draw conclusions that were different from those obtained using **standard approaches** based on the frequency and empirical distributions.