

SUPPORTING INFORMATION (SI)

Appendix A: Sampling Frame, Data and Measures

Forced migrants typically include two types of individuals, refugees who cross an international border and internally displaced persons (IDPs) who flee their homes but do not cross the border into another country. Given the general difficulties worldwide of obtaining accurate data on forced migration caused by conflict, a detailed account of the issue in this case and a description of the problems one faces in enumerating conflict-induced displacement in Nepal is provided below.

Nepal went through a decade of civil war beginning in 1996 in which it was estimated that over 50,000 people were displaced from their homes and another 13,000 killed. The conflict started when the Communist Party of Nepal –Maoist (CPN-M) launched a ‘people’s war’ on February 13, 1996 with a stated objective of abolishing the monarchy and establishing a people’s republic. It ended in 2006 with the signing of a comprehensive peace accord between the rebels and the government. The United Nations High Commissioner for Refugees (UNHCR) estimates that only around 2,600 Nepalis had sought official refugee status by the end of 2006 (UNHCR 2007). This figure does not include individuals who went to India for reasons discussed below. Although the factors forcing refugees and IDPs from their homes are arguably the same, I do not have data on these refugees, so due to data limitations, the present study is limited to analyzing IDPs, who comprise the vast majority of displaced persons from the Nepali civil war.

The more difficult issue for this study is the flight of individuals to India and the accuracy of the estimates of IDPs. Individuals displaced to India are not considered refugees. An open border, established in a 1950 Treaty of Peace and Friendship between Nepal and India, makes it

easy for Nepalis to cross the border into India and difficult for them to be tracked. Citizens of the two countries are treated on par with each other, no travel documents are needed to cross the border, and citizens crossing the border are well tolerated. Although many Nepalis are reported to have fled to India, no refugee camps were established for displaced Nepalis and no systematic documentation was carried out by either side.

Most researchers on forced migration acknowledge the fact that acquiring accurate data on forced migration is challenging, if not impossible; Nepal is no exception to this. Figures on displacement during the Nepali civil war vary dramatically. For example, according to a study by the Norwegian Refugee Council (NRC 2005), between 100,000 and 200,000 people had been displaced in Nepal by the end of 2004. NRC derived this estimate from figures generated by different governmental and non-governmental organizations operating in the country. Sources cited by the NRC include: the Indian Embassy in Kathmandu, which estimated that some 120,000 Nepalis crossed the border in January 2003 alone; the Asian Development Bank, which estimated the IDP figure to be between 300,000 and 400,000; the Finance Ministry of the Government of Nepal issued estimates between 300,000 and 600,000; the United Nations Development Program (UNDP) which cited a figure of 80,000, and so on. The politics behind displacement estimates as well as the 1700 kilometer-long open border between Nepal and India further confound the issue.

Only one organization, namely the Informal Sector Service Center (INSEC), a national human rights organization operating throughout Nepal since 1988, made a concerted effort to document and verify displacement figures (see Informal Sector Service Center, various Issues, *Nepal Human Rights Year Book*. Kathmandu). Their work was conducted on a sub-national, district by district basis. According to INSEC, 50,356 people were displaced from across the 75

districts from the start of the war to the end of 2004. There is strong reason to believe that the data collected by INSEC is the most reliable and accurate representation of relative displacement across the 75 districts of Nepal for a number of reasons. Because INSEC operates in all 75 districts of Nepal, the data collected by their district offices are more reliable than the national estimates cited above. INSEC was the only organization to collect data on displacements at the level of the village development committee (VDC), the smallest administrative unit in Nepal. The number of VDCs per district ranges from 13 to 115 with an average of 52 villages. In most cases, INSEC recorded not just a count of displaced but also the names of the persons affected by the conflict — displaced, killed or abducted. In addition, INSEC made a concerted effort to track individuals displaced within the country and later assist them in their return.

INSEC's primary focus was to document internal displacement. Because of the open border with India, it became very easy for individuals, especially those living in the districts along the border, to flee across the border for short periods of time to temporarily escape impending violence. The much larger figures on displacement cited by various organizations during the war likely included these temporary displacements across the border. While these figures are important to acknowledge, INSEC's data is likely the most accurate in terms of capturing the lasting impact of the war on displacement and the relative impact across districts.

The sampling frame used in this study is INSEC's list of people displaced from each district. INSEC made every attempt to include and document all individuals that were displaced from each district between 1996 and 2006 due to conflict. While INSEC was most interested in documenting incidents of human rights violations and focused on tracking displaced persons with the intention of protecting their rights and assisting them in their return, ambiguity remains as to the destination of some individuals listed by INSEC.

Given the long open border with India, some individuals crossed the border into India, temporarily or permanently, and some may have gone on to a third country. While discrepancies exist as to the precise number of people displaced by the war in Nepal, and politics likely played a role in the estimates produced especially by the government, a number of national and international organizations credited INSEC with having had the most extensive monitoring network and the most systematic means of data collection and documentation. Furthermore, as a non-governmental human rights organization, INSEC had far better access to the villages throughout Nepal since government officials were openly targeted by the Maoists. Indeed, according to INSEC officials, the government often turned to INSEC when it needed data on human rights violations, including conflict-induced displacements. In sum, the individual-level data collected on displacement by INSEC are believed to be the most accurate and the means of data collection used by INSEC was consistent across all 75 districts of the country.

Data for this study come from field research conducted during the summer and fall of 2008. The full data set used in this study consists of a national sample of 1804 respondent households from 56 village development committees (VDCs) drawn from 11 districts of Nepal, plus the capital of Kathmandu. Households were selected from 226 sampling units, called wards, from across these 11 districts. The sample represents all the five development regions (east, central, west, mid-west and far-west), three topographical zones (mountains, hills and plains), and both rural and urban parts of the country. The survey was also administered in Kathmandu where many of the internally displaced persons fled. Nepal is divided into 75 districts. The districts are further divided into 3,914 Village Development Committees (VDCs), each with 9 wards for a total of 35,226 wards. A weighted multi-stage cluster sampling technique was used to go from region, to district, to VDC, to ward level and then two samples were randomly drawn

— one of individual households at the ward level and another of displaced persons originating from those wards. Use of wards as the sampling units has the advantage of offering a paired design of individuals who decided to stay and those who decided to leave within the same contextual environment. Besides being a mountainous country, Nepal is very poorly served by road networks. It can take several hours simply to walk between wards. Cluster sampling down to the ward level and then randomly sampling within wards also made the project feasible given time and financial constraints.

The sampling method involved multiple stages because it sought to ensure that the sample represents (a) districts that were hard-hit during the conflict, (b) all three topographical regions, (c) all five development regions, and (d) both rural and urban parts of the country. In addition, given resource and logistic constraints the method aimed at ensuring that samples would be drawn from areas that produced displaced persons as a result of the conflict. In the first stage, all districts that had recorded at least 500 casualties or 500 displacements during the conflict were selected. The selection criteria were based on secondary data provided by INSEC on the number of people killed and displaced from each district. A total of 19 districts met this threshold. Four of the five economic development regions contained exactly two districts that met my threshold, and varied topographically, so these eight were chosen. The remaining eleven districts were in the mid-western region where the fighting originated; three out of these eleven districts were chosen, one in each topographical region. This resulted in a total of 11 districts, plus Kathmandu, being retained for sampling. Given resource constraints, the total number of interviewees was set at 1500 for the 11 districts, with a target of 1000 displaced and 500 non-displaced, with the remaining 304 interviewees coming from the capital. The number of displaced was further divided into two groups: 500 interviewees still displaced and 500

interviewees that had returned home. In each of the 11 districts the target number of interviewees was determined by the proportion of displaced identified by INSEC in each district. This captures the dynamics of conflict as well as the economic and geographical variance in the country.

Each district is divided into VDCs, with each VDC further subdivided into nine wards. Only VDCs with ten or more displaced persons were used in the sampling of respondents. From each district, 5 VDCs meeting this threshold were randomly selected, and the targeted number of respondent was determined by the proportion of displaced in each of the VDCs. Next, the targeted number of respondents from each of the 5 VDCs was randomly sampled from the wards in proportion to the number of displaced in each ward. Displaced respondents were randomly selected from a list of all displaced persons originating from the wards. This list was maintained by INSEC at the ward level throughout the conflict.

The 500 non-displaced respondents were randomly selected from the same districts/VDCs/wards in which the displaced originated. Target numbers of non-displaced from each ward were based on the same proportions used for sampling the displaced. The table below lists the 11 districts identifying the economic development region and topographic zone where each is located, and the target number (and actual number) of displaced respondents based on the proportion of displaced originating in each of the districts out of the total number of displaced persons identified in the eleven districts. So, for example, Rolpa had 1,817 displaced out of the total 17,386 displacements in the 11 districts, resulting in a target number of 105 displaced interviewees, and 52 non-displaced. Rolpa is further broken down into the five randomly selected VDCs. Based on the proportion of actual displacements in each of the five VDCs, a target number of interviewees is given, along with the actual number of displaced persons

interviewed and the number of non-displaced interviewed. The target and actual number of interviewees differs somewhat for each VDC because INSEC's and the Nepali government's identification and documentation of displaced persons as well as people injured, killed and disappeared was still on-going at the time the interviews were conducted, so the identification of conflict-induced displacement was still somewhat in flux. In addition, the monsoons were ongoing during part of the interview period. Farmers had begun to work by the time the first phase of the survey was conducted and many displaced persons were working in the fields to earn a living, making it more difficult to locate them.

Eleven Districts Selected for Sampling with Target (and Actual) Number of Respondents Interviewed

Economic Development Regions						
Topographic Zones		Far West	Mid-West	Western	Central	East
	Mountains	Bajura: 84 (70)	Kalikot: 203(218)			Taplejung: 44(50)
	Hills		Rolpa: 105(96)	Lamjung: 49(47)	Ramechhap: 73(88)	
			<i>Thawang</i>			
			<i>Kureli</i>			
			<i>Uwa</i>			
<i>Mirul</i>						
<i>Bhawang</i>						
Plains	Kailali: 118(124)	Bardiya: 94(108)	Kapilbastu: 152(151)	Chitwan: 48(43)	Jhapa: 30(17)	

Rolpa as an Example of the Sampling Process

VDCs	Proportion of actual displacement in five randomly selected VDCs	Target Number of Interviewees	Actual Number interviewed	Actual Non-displaced interviewed
Thawang	0.27	28	19	28
Kureli	0.26	27	37	12
Uwa	0.23	24	20	11
Mirul	0.20	21	15	7
Bhawang	0.03	3	5	2
Total	100%	105	96	60

While some of the randomly selected individuals for the sample refused to be interviewed, others halted in the middle of the survey, and still others could not be found. The response rate of 86.3%, with only .44% of the responses coming from proxy respondents, resulted in an overall sample of 1804. The number of observations used in the analysis varies due to random missing data.

Appendix B: Measures, Descriptive Statistics, Additional Analysis and Illustrations

The variable income measures respondents' estimated annual household income. During the survey, respondents were asked to identify their annual household income from the following 6 categories: (1) < 25,000, (2) 25,000 to 34,999, (3) 35,000 to 49,999, (4) 50,000 to 74,999, (5) 75,000 to 99,999, and (6) >100,000 Nepali Rupees.

Threat of violence is a composite index, computed from information about the gravity of threat as perceived by each respondent on a scale of 1 (not important at all) to 4 (very important) for the following six components: (1) physical threat intended to physically harm the respondent either by beating or some kind of physical assault, (2) threat of political coercion such as forced conversion of political ideology or related acts intended at curtailing one's political freedom, (3) threat of forced recruitment into either the rebel or state army, (4) threat of murder of a family member, (5) threat of physical or mental torture such as amputation, harassment, etc., and (6) threat of sexual humiliation. Given that X_{ij} , $\min(X_{ij})$, and $\max(X_{ij})$ represent the actual, minimum and maximum gravity (on a scale of 1 through 4) of the i^{th} respondent ($i = 1$ to 1804) and the j^{th} component of threat ($j = 1$ to 6), then the *Threat of Violence* _{i} is defined as

$$\text{Threat of Violence}_i = \sum_{j=1}^6 \frac{X_{ij} - \min(X_{ij})}{\max(X_{ij}) - \min(X_{ij})}.$$

The threat of violence index represents individual respondents' perceptions about the level of physical and mental threat that they faced during the conflict. Equal weights have been accorded to all 6 sources of threat mentioned above in computing the threat of violence index. Thus, the variable *THREAT OF VIOLENCE* in the empirical analysis represents the gravity of threat to an individual with a possible range of 0 to 6, accounting for 6 different sources of threat. It gives a

broad measure of the perceived threat of violence to an individual, taking into account the variance in violence across the villages in Nepal.

Table B.1 below provides the descriptive statistics for the variable *THREAT OF VIOLENCE*, along with all of the other variables used in the analysis. In Table B.2, I report results from difference of means tests for the independent variables before and after matching displaced and non-displaced samples. Table B.3 shows the matching diagnostic output from the “pstest” command in stata 11 and Table B.4 reports results from a probit analysis for the unmatched data.

Table B.1: Descriptive Statistics

Variables	N*	Mean	Stand Deviation	Minimum	Maximum
<i>DISPLACED</i>	1424	0.78	0.41	0	1
<i>ACTUAL VIOLENCE</i>	1424	0.86	0.30	0	1
<i>THREAT OF VIOLENCE</i>	1424	2.20	1.25	0	6
<i>INDUSTRY PRESENT</i>	1424	0.06	0.24	0	1
<i>INCOME</i>	1424	2.23	1.45	0	6
<i>LAND (Logged)</i>	1424	7.67	2.76	0	12.91
<i>CROP/ANIMAL-LOSS</i>	1424	0.74	0.86	0	2
<i>LAND-LOSS</i>	1424	0.31	0.46	0	1
<i>HOME-DESTROYED</i>	1424	0.18	0.39	0	1
<i>INDUSTRY DESTROYED</i>	1424	0.07	0.26	0	1
<i>SOCIAL NETWORK</i>	1424	0.66	0.47	0	1
<i>MOTORABLE ROAD</i>	1424	0.32	0.47	0	1
<i>CPN(M)</i>	1424	0.11	0.32	0	1
<i>CHILDREN</i>	1424	2.01	2.95	0	26
<i>EDUCATION</i>	1424	1.30	1.40	0	6
<i>MALE</i>	1424	0.60	0.49	0	1
<i>AGE</i>	1424	41.28	13.19	17	85
<i>AGE SQUARE</i>	1424	1877.80	1193.02	289	7225

*= The *N* corresponds to the matched sample used in the probit models in the body of the article.

Table B.2: Difference of Means Test for Independent Variables Before and After Matching Displaced and Non-Displaced Samples (Paired Differences)

Variables	Mean Difference (Before Matching)	Mean Difference (After Matching)
<i>ACTUAL VIOLENCE</i>	-0.23 (.02)**	-0.24 (.02)**
<i>THREAT OF VIOLENCE</i>	1.28 (.06)**	-1.37 (.03)**
<i>INDUSTRY PRESENT</i>	.01 (.01)	.02 (.01)
<i>INCOME</i>	.08 (.08)	.08 (.09)
<i>LAND (logged)</i>	-0.30 (.15)*	-0.27 (.17)
<i>CROP/ANIMAL-LOSS</i>	-0.61 (.04)**	-0.65 (.05)**
<i>LAND-LOSS</i>	-0.33 (.02)**	-0.35 (.03)**
<i>HOME-DESTROYED</i>	-0.20 (.02)**	-0.20 (.02)**
<i>INDUSTRY DESTROYED</i>	-0.04 (.01)**	-0.03 (.02)**
<i>SOCIAL NETWORK</i>	.04 (.02)*	.03 (.03)
<i>MOTORABLE ROAD</i>	-0.01 (.02)	.004 (.03)
<i>CPN(M)</i>	.17 (.02)**	.18 (.02)
<i>CHILDREN</i>	0.22 (.11)**	.19 (.12)
<i>EDUCATION</i>	-0.25 (.07)**	-0.16 (.09)*
<i>MALE</i>	-0.09 (.03)**	-0.04 (.03)
<i>AGE</i>	-1.13 (.71)	-1.36 (.82)
<i>AGE SQUARED</i>	-107.31 (64.97)	-120.38 (74.40)
<i>N</i>	1739-1802	1481-1528

Note: Figures in (parenthesis) are standard errors, ** =Significant at the .05 level or better, * =significant at the .10 level.

Table B.3: Matching Diagnostic Output from the “pctest” command in stata 11

Variables	Sample	Mean		Bias		t-test	
		Treated	Control	% bias	% reduced bias	t-value	p-value
<i>EDUCATION</i>	Unmatched	1.3173	1.1536	12	52.1	1.91	0.056
	Matched	1.3173	1.2388	5.7		1.4	0.163
<i>MALE</i>	Unmatched	0.61688	0.57971	7.6	75	1.24	0.213
	Matched	0.61688	0.62616	-1.9		-0.47	0.641
<i>CHILDREN</i>	Unmatched	1.9502	2.1391	-9.9	54	-1.57	0.116
	Matched	1.9502	1.8633	4.5		1.14	0.256
<i>AGE</i>	Unmatched	41.57	40.214	10.2	98.6	1.65	0.098
	Matched	41.57	41.551	0.1		0.03	0.972
<i>AGE SQUARE</i>	Unmatched	1909.4	1789	10	76.1	1.62	0.106
	Matched	1909.4	1880.6	2.4		0.6	0.551

Figure B.1: Matching Diagnostic Output from the “psgraph” command in stata 11

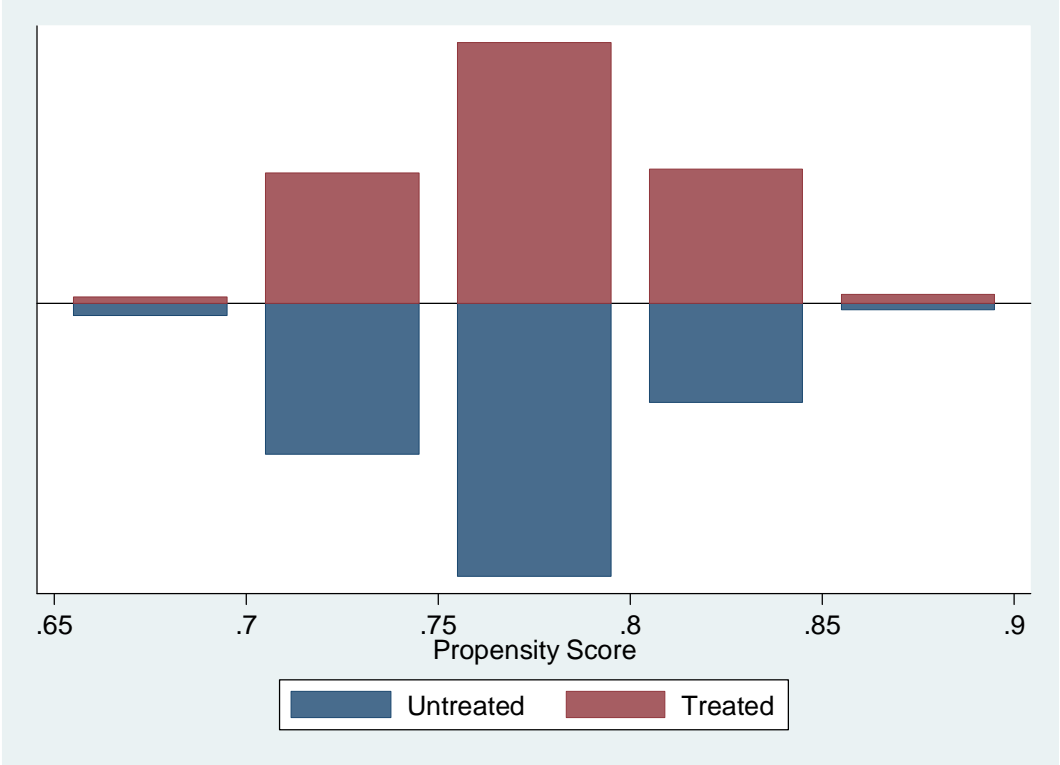


Table B.4: Probit Analysis of Internal Displacement with Unmatched Data

Variables	Model 1	Marginal Effects	Model 2	Marginal Effects
<i>ACTUAL VIOLENCE</i>	1.17(.15)**	.39**	1.15(.16)**	.38**
<i>THREAT OF VIOLENCE</i>	.43(.06)**	.11**	.42(.06)**	.10**
<i>INDUSTRY PRESENT</i>	-.64(.21)**	-.16**	-.62(.22)**	-.15**
<i>INCOME</i>	-.06(.03)*	-.01*	-.08(.04)*	-.02*
<i>LAND (Logged)</i>	.03(.01)**	.01**	.03(.02)*	.01*
<i>CROP/ANIMAL-LOSS</i>	.25(.08)**	.06**	.29(.09)**	.07**
<i>LAND-LOSS</i>	1.07(.17)**	.21**	1.10(.18)**	.22**
<i>HOME-DESTROYED</i>	.48(.23)**	.10**	.51(.27)*	.11**
<i>INDUSTRY DESTROYED</i>	.77(.27)**	.13**	.80(.29)**	.14**
<i>SOCIAL NETWORKS</i>	-.10(.12)	-.02	-.15(.12)	-.03
<i>MOTORABLE ROAD</i>	.28(.10)**	.07**	.24(.10)**	.06**
<i>CPN(M)</i>	-.58(.15)**	-.17**	-.56(.17)**	-.16**
<i>CHILDREN</i>	-	-	-.04(.02)*	-.01*
<i>EDUCATION</i>	-	-	.08(.04)**	.02**
<i>MALE</i>	-	-	.19(.08)**	.04**
<i>AGE</i>	-	-	0.01(.02)	.003
<i>AGE SQUARED</i>	-	-	-0.0001 (.0002)	-.00003
<i>CONSTANT</i>	-1.66 (.21)**	-	-1.88(.50)**	-
<i>N</i>	1672	-	1559	-
<i>Wald χ^2</i>	270.06**	-	259.62**	-
<i>% positive</i>	72.55	-	71.84	-
<i>% correctly predicted</i>	83.01	-	82.87	-
Notes: <i>N</i> is less than 1804 due to random missing data; Figures in (parenthesis) are robust standard errors clustering at the ward level. ** = Significant at the .05 level or better; * = Significant at the .10 level.				

In addition to the results reported here and in the body of the article, I performed a number of additional checks on the analysis. First, because of the multicollinearity between existence of a motorable road and elevation, the impact of the elevation of the countryside was assessed by replacing motorable roads with a variable that measures whether the respondent comes from a region in the mountains (3), hills (2), or plains (1). The statistically significant results suggest that people residing in the mountainous region of the country are around 6% less likely to flee, holding all other variables constant at their mean or mode.

Because the survey was conducted after the end of the conflict, most respondents were readily forthcoming in revealing their party affiliation; only 5% of the sample declined to identify with a specific party. I also test for the impact of membership in NC, RPP, and UML on displacement separately, leaving out Maoists as the base of comparison. The results support the hypothesis on party affiliation. All three coefficients are positively signed, while the coefficients for NC and RPP are statistically significant at 5% level throughout the analyses, implying that individuals affiliated with the NC or RPP in particular were more likely to flee than members of the Maoist party.

Third, I tested the effect of age by including two dummy variables in the models—one for young adults between the age of 18 and 40 and one for middle-aged individuals between 41 and 65 with elderly as the comparison group—to control for the possibility that there is a non-linear relationship between age and forced migration. It is possible that those who are below 40 lacked the autonomy and financial resources to flee while those in the middle were more able, and the aged once again less able to move. Both coefficients were positive, but not significant.

Finally, some of the empirical results presented in the body of the article are illustrated graphically in Figures B.2, B.3 and B.4 below. Figure B.2 indicates that the probability of flight

is higher on average when people experience actual violence, and the probability increases with an increase in the threat of violence. Also, presence of a motorable road increases the probability that they will flee as compared to those that have no access to a road. Figure B.3 shows the impact of violence and threat of violence while controlling for elevation. The figure shows that people living in the plain region of the country have a higher probability of fleeing than those in the hill or mountain regions of the country. However, these differences diminish as threat of violence increases, and the probabilities converge in the presence of actual violence; people who experience violence and are under a higher perceived threat of violence are more likely to flee no matter where they live.

Figure B.4 illustrates the impact of social networks on individuals' decisions to flee. With and without actual violence, the presence of social networks somewhat dampens the probability of flight. These results provide some support for my theoretical proposition that social networks provide an important mechanism for individuals to cope with their circumstances. However, for those who experience human rights abuses, as the threat of violence intensifies, the probabilities converge.

Figure B.2: Presence of Road and probability of being displaced

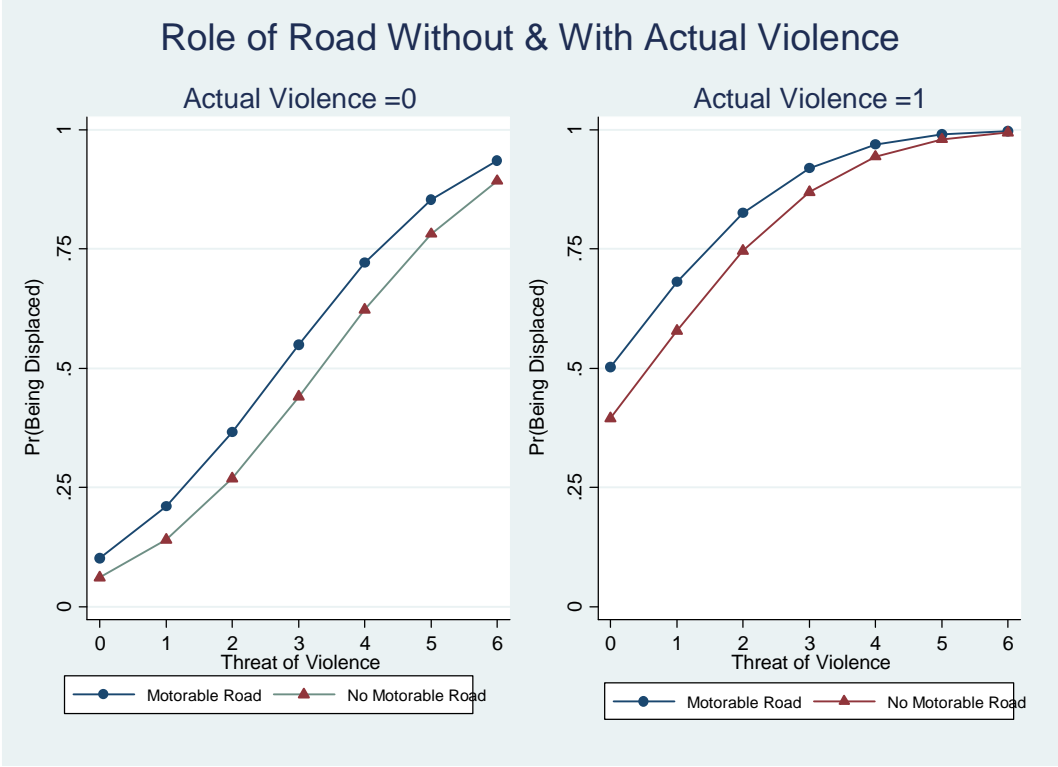


Figure B.3: Elevation and probability of being displaced

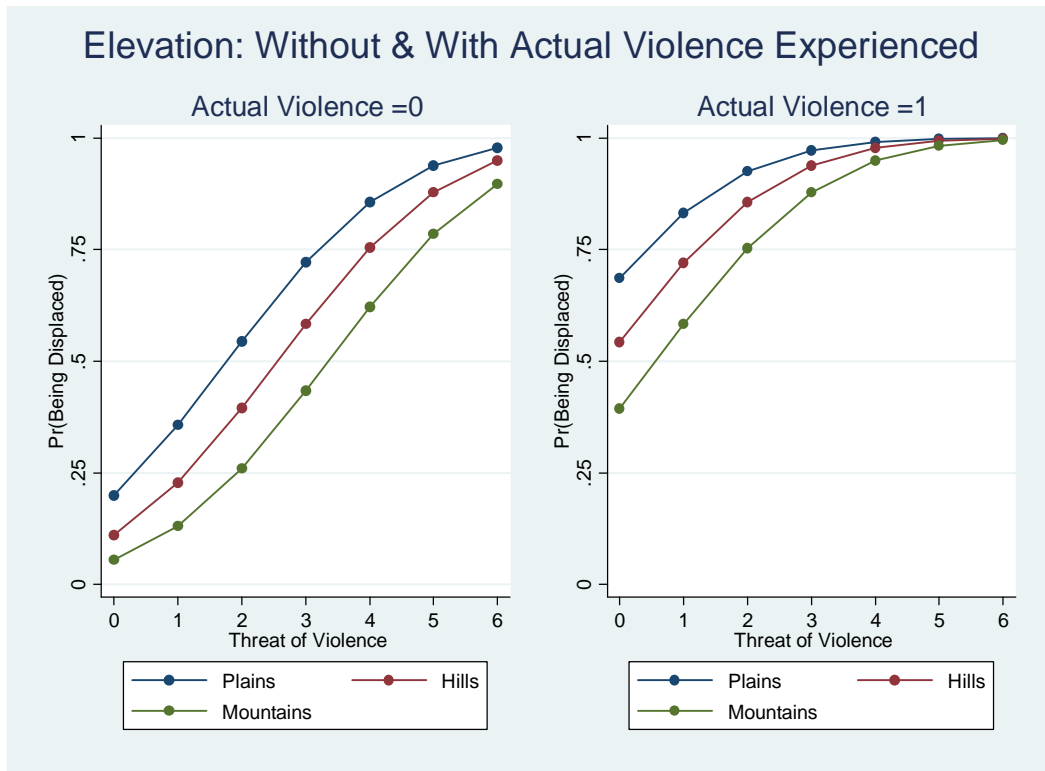


Figure B.4: Social Network and probability of being displaced

