CMIWORKINGPAPER

LSMS Data Quality in Maoist Influenced Areas of Nepal

Magnus Hatlebakk

WP 2007: 6

CMI CHR. MICHELSEN INSTITUTE

LSMS Data Quality in Maoist Influenced Areas of Nepal

Magnus Hatlebakk

WP 2007: 6

CMI CHR. MICHELSEN INSTITUTE

CMI Working Papers
This series can be ordered from:
Chr. Michelsen Institute
P.O. Box 6033 Postterminalen,
N-5892 Bergen, Norway
Tel: + 47 55 57 40 00
Fax: + 47 55 57 41 66
E-mail: cmi@cmi.no
www.cmi.no

Price: NOK 50

ISSN 0804-3639
ISBN 978-82-8062-207-5
This report is also available at:
www.cmi.no/publications

Indexing terms
Household survey
Civil war
JEL classification: C810, O120

Project title
Social exclusion-Nepal

Project number
25072

Contents

1. INTRODUCTION **1**
2. MEASURES OF MAOIST CONTROL **2**
3. SELECTION BIAS IN THE PANEL DATA **5**
4. MAOIST INFLUENCE ON DATA QUALITY **6**
5. CONCLUSIONS **13**
REFERENCES **14**

1. Introduction*

The first round of the Nepal Living Standards Survey (NLSS1) is a high quality data source that has been applied for a number of empirical studies, including Bardhan et al. (2003), Jacoby (2000), and Koolwal (2007). Data were collected in 1995/96, prior to the civil war in Nepal[1]. The Maoist party of Nepal started the so called People's war in 1996, and gradually extended their influence in the rural areas. Eight years into the war, in 2003/04, the Central Bureau of Statistics, in collaboration with the World Bank, conducted the second round of the Nepal Living Standards Survey (NLSS2). We investigate to what extent the quality of the NLSS2 data has been affected by the Maoists.

Throughout the civil war the author conducted fieldwork in Maoist influenced areas of Nepal. From our own experiences, and conversations with field assistants, human rights activists, government officials and villagers, we know that the Maoists did not always allow data collection in the areas where they had some control. As their control varies geographically, we would expect the data quality to vary as well. Now, the NLSS2 reports that they were not able to conduct interviews in some few villages that were originally in the random sample. These villages are mostly located in the Far-Western region. As the Maoists have their main strongholds not in the Far-Western but in the Mid-Western region, in particular Rolpa and Rukum districts, this adds to our skepticism regarding the quality of data from the Maoist strongholds. If the Maoists allowed data collection in their core areas of the Mid-Western hills, but not in districts where they presumably have less control, then they were probably able to have some control over the data collection in the core areas.

To cite a field assistant, who in turn cited a local Maoist: *"You know the situation here, just fill the form, you do not have to talk to the people"*. Our concern is that in districts like Rolpa and Rukum the Maoists either participated in filling the forms, asked the enumerators to fill the forms without conducting interviews, or directly participated in the interviews. Note that this would be manageable for them, they are well organized at the village level in their core areas, and only 12 households were interviewed in each ward, or Primary Sampling Unit (PSU).

When a survey team arrives in a village in the most affected districts, they will normally be stopped by the Maoists, and they will have to inform the Maoists about everything concerning the survey. This leads to our *first hypothesis*, we expect to find that the survey teams spent more time in Maoist-controlled villages than in other villages. Filling the forms probably took less time. But we expect that discussions and bargaining with the Maoists took additional time, the teams probably spent at least one extra day in each PSU, and in some cases had to return days later, after they got permission from the Maoists.

We also expect that interviews conducted under Maoist supervision will have more standardized answers, which leads to our second and third hypotheses. The *second hypothesis* is that some numbers will appear more frequently in the data. In particular, we expect ages of 5 and 10 to be more frequently reported in Maoist-controlled villages. The *third hypothesis* is that the standard deviation of the measured variables will be smaller in these villages than in comparable villages from NLSS1, and as compared to villages from NLSS2 that are not controlled by the Maoists. We will report on the standard deviation for landholdings and agricultural wages.

* The research is funded by the Social Inclusion Research Fund, which in turn is funded by the Norwegian Embassy, Nepal. Thanks go to Ivar Kolstad for useful comments.

[1] For details on the Nepal Living Standards Survey, see www.worldbank.org/lsms.

To minimize the influence of factors other than data errors, we report difference-in-difference measures, that is, we compare the change in Maoist-controlled areas to the change in non-Maoist areas. And we restrict our analysis to the hill and mountain districts of Nepal, as the Maoists control only a few districts in the plains (terai) of Nepal. Obviously, a larger change in Maoist-controlled hill/mountain districts may be due to a real Maoist influence, or other factors that vary between districts. Thus, in analyzing the data we will have to discuss alternative explanations for the empirical findings.

The *fourth hypothesis* is that we expect a low response rate on some questions. If forms are filled without real interviews taking place, we may expect that more boring issues, or questions at the end of the questionnaire, would not have been filled in at all. This was probably the case also for NLSS1, but we expect even lower response rates if there is a Maoist influence. We report on the response rate for landholdings and purchases of food items. The latter is, in particular, a relevant test, as there were as many as 5 pages of detailed questions on approximately 50 different food items in the NLSS2 questionnaire. Finally, for agricultural wages we also test a *fifth hypothesis*, that is, the Maoists may have influenced the reported wages in support of their minimum wage policy.

As indicated, we expect the data quality to vary with the degree of Maoist control. However, this variable itself is of particular concern. Previous studies of the Nepal conflict that have used the NLSS data tend to apply killings as an indicator of Maoist influence. This is not a well founded indicator, as we will discuss in section 2. Section 3 demonstrates a selection bias in the panel sub-sample of NLSS2, which explains why we apply the cross-sectional samples. Section 4 presents the findings on potential data errors, while section 5 concludes.

2. Measures of Maoist control

In conducting empirical analysis of the effects of the Maoist insurgency in Nepal, one has to be clear on what is actually measured, and how the variables relate to the stated hypothesis. Our interest is in the degree of Maoist control, while Murshed and Gates (2005), in particular, study the level of conflict. They measure the level of conflict by the total number of people killed, referring to Gautam (2001). As we will argue in this section, this is not the best indicator of Maoist control. But for comparison we also report on killings, in addition to other indicators. For killings we use data from INSEC (2006a), which is probably the best available data source.

INSEC reports separately killings by the state and by the Maoists, while we add the two and report on total killings during the civil war from 1996 to 2006. In Table 1 we report a dummy representing a large number of total killings, relative to the district population in the 2001 census. To make the indicator comparable to our preferred indicator, which is a government classification of Maoist influence, we have selected the cut-off that defines the dummy such that we get the same number of Maoist-controlled districts as in the government classification. At the cutoff, 0.067% of the district population were killed during the 10 years of conflict. This is approximately equal to the national average of about 12 000 people killed out of a population that increased from approximately 20 million in 1996 to approximately 25 million in 2006. The district of Rukum, which is a stronghold of the Maoists, was hardest hit, with 0.5% of the population being killed during the conflict, that is, 733 people were killed by the state and 169 by the Maoists, which is on average 2 people every week, in a district of 190 000 people.

Although Rukum and Rolpa, where the killings are high, are obviously controlled by the Maoists, the number of people killed will not necessarily reflect the strength of the Maoists. At times, for example in April 2005, the army attacked the Maoist strongholds in Rolpa and Rukum. But this was

not the normal situation. Most attacks, with the majority being initiated by the Maoists, took place in the neighboring districts. According to INSEC (2005) the Maoists killed three people and the state killed six in Rolpa district in 2004, while in the neighboring Dang district, the Maoists killed 42 and the state killed 11. This reflects the fact that the Maoists have their base areas in the hill districts, but attack army barracks and also civilians in contested areas in the terai (the plains along the border to India). We find a similar trend within the terai. The Maoists have a stronghold in Bardiya district (where 32 people were killed by the Maoists and 30 by the state), but attack the neighboring districts of Kailali (47 killed by the Maoists and 70 by the state) and Banke (51 killed by the Maoists and 54 by the state). Within the base areas it appears that there is a tacit (or explicit) truce between the Maoists and the local security forces. Heavy attacks on the Maoists in these areas are conducted by security forces from Kathmandu, as in Rukum and Rolpa in April 2005, and in Bardiya in February 2005.

One alternative quantitative indicator of the strength of the insurgents, which we considered, is voter turnout in elections. The Maoists have boycotted and interrupted recent elections, and have been more successful in doing so within their base areas. Now, there have been no elections in Nepal since 1999, and since then the Maoists have increased their presence, for example in the eastern hills. Still, this is a potentially useful indicator that to some extent reflects the strength of the insurgents. For the indicator to be useful, we would like to find a discontinuity in ranked-ordered participation rates at which we may set the cut-off for an indicator of Maoist control. However, the data shows a gradually decreasing participation rate, with a drop only for the most obvious Maoist districts of Rukum, Rolpa, and Salyan. We are thus not able to identify a useful cutoff for this indicator, and it will not be reported here.

Another quantitative indicator, which we do report in Table 1, is the number of displaced people. The data is reported by INSEC (2006b), apparently for the period 2002-2004. We apply the numbers of people displaced "due to the Maoists", or "due to terror". That is, we do not count those who are displaced "due to the state". This is because displacement "due to the state" is unevenly distributed between the districts, with 67% from a single district, Kapilbastu. Next, we divide by the district population from the 2001 census, and make a dummy if the share is at least 0.12%, which gives the same proportion of Maoist-controlled districts as the preferred government ranking. Information is missing for five districts, Bhaktapur, Kathmandu, Mustang, Manang, and Parsa. The first four districts are probably not recorded with displaced people. The same may be the case for Parsa, as people there probably move into the major city of Birgunj, and thus still live within the same district.

In addition, we report subjective rankings based on Sharma (2003). The Home Ministry has categorized districts into *sensitive classes A, B and C*, where six adjacent hill districts in the Mid-Western region are in the *A* class, including Rukum and Rolpa. Now, the *C* class even includes Lalitpur, which is part of the Kathmandu metropolitan area, which in turn may suggest that this ranking is not a good indicator of Maoist control. An alternative is to use the Maoists' own "ranking", as they have announced "People's government" in 21 districts according to Sharma (2003). This latter indicator includes districts that came under Maoist control more recently, such as Terathum in the eastern hills. However, the list does not include Bardiya, which appears to us to be a Maoist-controlled district.

All the indicators in Table 1 have some merit, but from our regular visits to different parts of Nepal during the civil war, and regular reading of the daily news from Nepal, it is our impression that the government classification is the most reliable. As any indicator will be subjective, including the subjective choice of a numeric indicator, we will apply the government classification, from which we combine all three sensitivity classes *A, B* and *C*. Table 1 lists the Maoist-controlled districts according to the different indicators, where our preferred indicator is marked.

Table 1. Maoist-controlled districts according to different indicators

People's government	Government classification**	Displacement high	Killings high
Achham	Achham		Achham
	Arghakhanchi	Arghakhanchi	Arghakhanchi
	Baglung		
		Baitadi	
Bajura		Bajura	Bajura
		Banke	Banke
	Bardiya*	Bardiya*	Bardiya*
			Bhojpur
		Dadheldhura	Dadheldhura
Dailekha	Dailekha	Dailekha	Dailekha
	Dang*		Dang*
Dhading	Dhading		
Dolakha	Dolakha		Dolakha
	Dolpa	Dolpa	Dolpa
		Doti	Doti
Gorkha	Gorkha	Gorkha	Gorkha
Gulmi	Gulmi		
		Humla	Humla
Jajarkot	Jajarkot	Jajarkot	Jajarkot
Jumla	Jumla	Jumla	Jumla
		Kailali*	
Kalikot	Kalikot	Kalikot	Kalikot
		Kapilbastu*	
	Kavrepalanchoc	Kavrepalanchoc	Kavrepalanchoc
	Khotang		
	Lalitpur		
Lamjung	Lamjung	Lamjung	Lamjung
	Makwanpur		
		Mugu	Mugu
			Myagdi
Nuwakot	Nuwakot		
	Okhaldhunga	Okhaldhunga	Okhaldhunga
Palpa			
		Panchtar	
Parbat	Parbat		
	Pyuthan		
Ramechhap	Ramechhap	Ramechhap	Ramechhap
Rasuwa			
Rolpa	Rolpa	Rolpa	Rolpa
Rukum	Rukum	Rukum	Rukum
Salyan	Salyan	Salyan	Salyan
Shankuwasabha			Shankuwasabha
Sindhuli	Sindhuli	Sindhuli	Sindhuli
Sindhupalchok	Sindhupalchok		Sindhupalchok
		Solukhumbu	Solukhumbu
	Surkhet	Surkhet	Surkhet
Tanahu	Tanahu		
		Taplejung	Taplejung
Tehratum		Tehratum	
	Udayapur		

*Terai districts. **Our preferred indicator.

As we may expect, killings and displacements are correlated, and the government classification is correlated with the Maoists' own classification[2]. Note that the Maoist classification does not include Bardiya and Surkhet districts, which is the main reason why we prefer the government classification. Note that all districts that are mentioned in as many as three columns are also included in the government indicator, and all districts that are mentioned in all four columns are classified in sensitivity class *A* by the government, with one exception, Pyuthan. Pyuthan is classified in *A* by the government, but does not appear in any of the other columns. Still, the government classification makes sense, because the district is located between other Maoist-controlled districts in the Mid-Western hills.

Also note that very few of the Maoist-controlled districts are in the terai. So, for analysis of the terai sub-sample the data problems discussed in this paper will not be serious. And it is only the displacement indicator that includes terai districts from non-Mid-Western districts. According to the other indicators, only two terai districts are controlled by the Maoists, that is, Bardiya and Dang in the Mid-Western region. Thus, if we use the government classification, then any analysis of the terai sub-sample from Kapilbastu in the west to Jhapa in the east, which has a sample size of 996, and constitutes 81% of the terai NLSS2 sample, will not include any Maoist-controlled district.

3. Selection bias in the panel data

NLSS2 also includes a panel, as approximately 36% of the households from NLSS1 constitute a random sub-sample of 1232 households. However, only 78% of these households were identified, and the missing 22% are not a random sample as we shall demonstrate by use of the reported land values. Now, even for the NLSS2 cross-section some observations are missing, not because the enumerators did not find them, but because they did not conduct interviews in those wards. To sum up, out of the 270 missing households, 16 are from a PSU that does not exist anymore, and 56 are from 4 PSUs that the enumerators did not visit due to the Maoists, in total 72 households (6%) were from PSUs that were not visited. The remaining 198 households (16%) were not found by the enumerators. This leaves 962 households that were interviewed in both rounds. Among the 1232 in the original panel, 992 are from wards that were rural in NLSS1[3]. Among them, 792 were identified, and one of these households did not report land value in NLSS1, so we have information on land value for 791 households. As all the 72 households from non-visited PSUs are in the rural areas, the percentage not visited out of the *rural* sample is 7%, while the remaining 128 (13%) were not found by the enumerators. In NLSS1 there were in total 2657 rural households, leaving 1665 as a randomly selected comparison group. In Table 2 we report the land values for the comparison group, as well as the three potentially non-random samples of respectively 791, 72, and 128 households.

Table 2. Land values (weighted estimates) NLSS1

	NLSS1, panel	NLSS1, no access	NLSS1, not found	NLSS1, rest
Mean value Rs	245 000	91 000	124 000	192 000
95% conf. interval	182 – 308 000	63 - 119 000	87 – 160 000	165 – 218 000
Median value Rs	80 000	61 000	49 000	60 000
300 000+	16%	6%	11%	15%
N = 2656	N = 791	N = 72	N = 128	N = 1665

[2] 25 out of 30 districts (83%) with many displaced people also had many killings, and 19 out of 24 districts (79%) with People's government were also on the government list, while only 20 out of 32 districts (63%) with many killings were also on the government list.

[3] Taulihawa VDC in Kapilbastu was rural in NLSS1, and urban in NLSS2.

The four sub-samples will have the same expected average land value if there is no selection bias in the panel data. However, we find that the panel is biased, the average land value is 245 000 rupees, which is significantly larger than the average of 91 000 for the villages where NLSS did not get access, and than the average of 124 000 among the households the team did not find within the accessed villages. But the panel is not significantly different from the random NLSS1 reference group (the t-value for the difference between the means is 1.53), probably because the samples of non-interviewed households are so small. Due to the fact that the difference between the panel and the control group is not significant, we may expect unbiased estimates in regression analyses that apply the panel data. However, the significantly lower land values for the non-interviewed households indicate that one should test for selection bias in each and every regression analysis that applies the panel data. As ownership of land tends to keep people from moving, it is not surprising that the missing households have less land. So, one should in particular test for selection bias when one expect differences in behavior between landed and landless households. In the present paper we do not use the panel, and report on the cross-sectional samples only to avoid the selection bias.

As we, in this section, discuss sample issues related to the two surveys, we want to make an additional note on the selection of PSUs for the NLSS2 cross-section. There is no overlap in PSUs between the two cross-sections, that is, it appears that the PSUs from NLSS1 were not included in the random sampling for NLSS2. Since both samples are random, this will not affect the randomness of the second sample, and it is a correct sampling technique if one believes that the first round data collection may affect the quality of the second round data. But one must keep in mind that if one uses data from, for example, only one district, then the selected wards (PSUs) are not the same in the two cross-sections. This will add to the standard deviation for any reported difference between the two cross sections, and one should be careful in comparing the two data sets for such a small geographical area. However, for larger sub-samples it is possible, and useful, to compare the cross sections, for example for the two eastern terai regions, which have a combined sample size of 816 rural households in the NLSS2.

4. Maoist influence on data quality

As discussed in the introduction, we test a series of hypotheses regarding the data quality in Maoist-controlled districts. We have five main hypotheses: 1) The survey teams spend additional days in each village due to bargaining with the Maoists. 2) The responses are more standardized, with, for example, numbers of 5 and 10 being overrepresented. 3) The standard deviation of the answers will be relatively smaller. 4) The response rate is lower on some questions. 5) Answers will support Maoist (minimum wage) policies.

As explained in the introduction, we check the hypotheses by comparing the reported change in a number of variables in Maoist and non-Maoist districts. Note that we use the government indicator of Maoist control as described in section 2. We only use the rural sub-samples, as the Maoists do not control urban areas, and only data from the hills and mountains, as the Maoists control few terai districts. For variables that we believe will be applied in future analyses of the NLSS data, we will report weighted estimates using population weights that adjust for the sampling procedure. The sampling weights are provided by NLSS. For variables that we construct only to control the data quality we do not use sampling weights. As discussed in the previous section there are selection problems with panel data in a war-ridden country. In this paper we thus use only the cross-sectional data, which consists of two independent random samples. We intend to use the cross-sections also in adjoining papers, for the same reason. If other researchers use the panel data they should check for selection biases.

We start investigating the *first hypothesis* by comparing the number of days spent in each PSU. We have corrected dates that appear to be obvious errors. For NLSS2 we did this for eight wards. These observations might have been outliers, as most corrections were for one household only, a household that was interviewed, let us say, one week before the others. If the date is not correct, then we would get a completely wrong picture of the data collection process, and we decided to correct such outliers. For example, if most interviews took place from day 25 to 29, and there is one observation on day 5, then we corrected that date to 25. We also had to make sure that we counted correctly at the end of the month, for example, interviews on days 29 to 31 and days 2 to 4 in the following month are recorded as a total of 7 days. As we applied the same rule for all months, in some few cases we lost one day at the end of the month, that is, if no interviews were conducted on, say, day 31. The assumed outliers might actually have been real interviews conducted before the team was approached by the Maoists. However, for NLSS2, which might have been conducted under Maoist influence, only 5 of the outliers were reported prior to the others. So, we decided that the problem of including a few biased outliers, which may change the means significantly, was a larger problem than correcting these few dates. For NLSS1 we only use wards that had 12 households in each ward. In the Far-Western region NLSS interviewed 16 households, and we do not include these wards here. And for NLSS2 two wards had missing information on dates.

The number of days in each ward is tabulated in Table 3 for the Maoist-controlled districts, and in Table 4 for the other districts. For the Maoist areas there are only small changes in the time taken to conduct the 12 interviews. The median number of days is 5 in both rounds. The mean has declined but the decline is not significant. Six days spent in the ward is more common in NLSS1 than in NLSS2, which explains the decline. But the general conclusion is no noticeable change in the time spent in the Maoist-controlled villages.

Table 3. Days spent in each ward, Maoist-controlled hill/mountain districts

	NLSS1			NLSS2		
Days in the ward	No. of wards	Percent	Cumulative	No. of wards	Percent	Cumulative
2	1	1%	1%	1	1%	1%
3	6	9%	10%	5	6%	8%
4	22	31%	41%	31	39%	47%
5	16	23%	**64%**	26	33%	**80%**
6	17	24%	89%	10	13%	92%
7	6	9%	97%	3	4%	96%
8	1	1%	99%			
9	1	1%	100%			
10				3	4%	100%
11						
	N = 70	100%		N = 79	100%	
Mean (weighted)			5.03 days			4.78 days
Conf. int.			(4.71 - 5.35)			(4.47 - 5.09)

Median in bold.

If we go on to the non-Maoist areas, see Table 4, then there is a significant decrease in the mean number of days. In particular, there used to be more villages where it took more than 5 days. The decrease may be due to improved transportation in these hilly districts, and the decrease can be seen as a normalization, as the mean number of days has declined to the level of the Maoist areas, that is, 5 days. We would expect more than 5 days in the Maoist areas if there were serious problems with the Maoists. Now, we have, later, got some indication that the enumerators, rather, hurried in the Maoist-controlled areas. So, the lack of a decline in Maoist areas may hide an increase in some districts and a decline in other districts. However, we find no support for this, there is, rather, also a normalization within the Maoist areas, with 4 and 5 days becoming the most common. So, with 4-5

days of data collection in Maoist as well as in non-Maoist districts in NLSS2, it appears that the Maoists did not obstruct the data collection, and we conclude that we do not find support for the first hypothesis. At the micro-level this is illustrated by the 4 wards (PSUs) in Rolpa and Rukum, where the interviews took respectively 4, 4, 5, and 6 days.

Table 4. Days spent in each ward, non-Maoist-controlled hill/mountain districts

	NLSS1			NLSS2		
Days in the ward	No. of wards	Percent	Cumulative	No. of wards	Percent	Cumulative
2				1	2%	2%
3				2	4%	7%
4	11	31%	31%	14	30%	37%
5	7	19%	**50%**	17	37%	**74%**
6	13	36%	86%	8	17%	91%
7	1	3%	89%	3	7%	98%
8	1	3%	92%			
9	2	6%	97%			
10	1	3%	100%			
11				1	2%	100%
	N = 36	100%		N = 46	100%	
Mean (weighted)			5.61 days*			5.03 days
Conf. int.			(5.11 - 6.12)			(4.61 – 5.46)

Median in bold.
* Significantly different from NLSS2 at the 10%-level.

We continue with the *second hypothesis*, and analyze the age data. In Table 5 we report the ages of all household members reported in the rural hill and mountain NLSS1 and NLSS2 samples. As ages in rural Nepal have increased slightly from a weighted average of 23.5 years in NLSS1 to 24.3 years in NLSS2 (the median is constant at 18 years), and we are comparing different cohorts, we do not compare frequencies directly. We construct a fraction at each five-year gap, that is, the frequency at that particular age compared to the average of the two ages on each side. For example, we report the average number of people of age 3, 4, 6, and 7, compared to the number of people being 5 years old.

Table 5. Ages reported in NLSS1 and NLSS2, hill/mountain districts

Age	NLSS1, mao	NLSS2, mao	NLSS1, nonmao	NLSS2, nonmao
5	0.97	1.12	1.12	0.99
10	1.04	1.15	1.00	1.07
15	0.99	1.34	0.99	0.91
20	0.95	1.06	0.94	1.04
25	*1.55*	1.28	1.15	1.22
30	1.94	*1.86*	**1.97**	1.56
35	2.04	2.05	**2.21**	1.77
40	***2.68***	2.22	1.95	2.13
45	**2.68**	2.03	2.55	*2.81*
50	**2.26**	1.55	2.17	*2.05*
55	2.00	2.11	1.97	1.83
60	*3.11*	**3.61**	1.97	**3.41**
65	1.86	2.07	2.05	***2.56***
70	1.79	1.96	***3.67***	*2.38*
75	1.45	***2.81***	***3.69***	2.07
N	4865	4779	3475	2866

Bold for between, and italic for within, survey comparison indicates a fraction of more than 1.5 and a difference of at least 0.3.

In Maoist districts the overrepresentation of the fives is more than hundred percent (which gives a factor larger than 2) for the ages, 35, 40, 45, 50, 55 and 60 in NLSS1, and for the ages 35, 40, 45, 55, 60, 65, and 75 in NLSS2. There is a decline in overrepresentation for ages 40, 45, and 50, and an increase for ages 60 and 75. The general impression is that the pattern of overrepresentation in the Maoist districts is relatively constant. In non-Maoist districts the overrepresentation is more than hundred percent for the ages 35, 45, 50, 65, 70, and 75 in NLSS1, and for the ages 40, 45, 50, 60, 65, 70, and 75 in NLSS2. There is a decline in overrepresentation for ages 30, 35, 70, and 75, and an increase for ages 60 and 65. Again the pattern of overrepresentation is relatively constant. The general conclusion is that overrepresentation of the fives happens at high ages in all districts, and we conclude that we find no support for the hypothesis of more standardized answers in the Maoist-controlled districts.

This far, it appears that real interviews have been conducted. Still, there is a suspicion that the teams interviewed some households, and not necessarily the correct ones, in the district headquarter to avoid visiting the villages. For example, a team may have been able to locate a family member of a particular household who lives in the district headquarter, or the team may have called for a family member to come to the district headquarter for the interview. In those cases, one may expect that the respondents have less information, or are less willing to spend time on all the detailed questions on agricultural activities. For example, they may not mention all the plots owned by the household. We thus now report on the number of plots, to test the *fourth hypothesis* of a low response rate, and we report on the standard deviations for landholdings, and thus report on the *third hypothesis*.

Now, we may expect that some landlords have sold land in the Maoist areas during the conflict period, and thus own fewer plots, which, in turn, may imply a consolidation of plots in these villages. Still, the average landholding will be unaffected as the plots are transferred to someone else. However, if interviews are conducted in the district headquarter, or under Maoist supervision, we may expect that some of the original plots will no longer be reported, which means that the total amount of land reported in the district will decline. At the same time, there is an underlying increase in agricultural land, as well as in households that own land, and we do not know a priori what tendency will dominate. To some extent we take this into account by reporting difference-in-difference measures. That is, we check whether the reported average landholding declines faster in Maoist-controlled districts. The findings are reported in Table 6.

There is some, but not significant, support for the hypothesis. While the mean land-size in non-Maoist areas is constant at 1.04 bigha, there is a small, but non-significant, decline from 0.97 to 0.94 bigha in the Maoist-controlled districts. In both areas the average number of plots decreases significantly, which thus reflects a consolidation of plots. At the same time the median landholding increases and the number of large holdings decreases, which indicates that the smallholders are buying plots from larger farms. The share of landless households, on the other hand, increases only non-significantly. To conclude, we find only insignificant support for the hypothesis of underreporting of land. The smaller number of plots is, rather, due to a real consolidation of land.

We recall that another hypothesis was to expect a smaller standard deviation in the Maoist-controlled districts of NLSS2. There is a larger decline from NLSS1 to NLSS2 in the Maoist-controlled districts, even when compared to the mean value. But, again this may be due to the consolidation on medium-sized farms.

Table 6. Reported landholdings (weighted estimates), hill/mountain districts

	NLSS1, mao	NLSS2, mao	NLSS1, nonmao	NLSS2, nonmao
Landless	4.1%	5.2%	9.3%	9.7%
Median holding*	0.68 bigha	0.75 bigha	0.64 bigha	0.68 bigha
Mean holding	0.97 bigha	0.94 bigha	1.04 bigha	1.04 bigha
St. dev.	1.12 bigha	0.96 bigha	1.28 bigha	1.22 bigha
5 bigha+	1.0%	0.6%	2.0%	1.3%
Median no. plots	3.0 plots	3.0 plots	3.0 plots	3.0 plots
Mean no. plots	4.0 plots	3.2 plots	3.8 plots	3.3 plots
Median value Rs	68 000	121 000	68 000	140 000
Mean value Rs	159 000	240 000	316 000	341 000
300 000 Rs +	14%	25%	13%	31%
	N = 912	N = 948	N = 633	N = 576

* Due to non-standard units for land in NLSS1 the number of observations for bigha is smaller.

We now go on to the *fifth hypothesis*. One may imagine that the Maoists allowed interviews to take place, but that they participated and made sure that important answers corresponded with their policy. In particular, one may expect them to prefer the workers to report the minimum wage set by them, rather than the real wage. Now, it is hard to separate a real effect of the Maoist minimum wage policy from misrepresentation of the wage. However, we can argue that the policy may work in their core areas, but not to the same extent in the periphery of their areas, for example that it is working in Rolpa and Rukum, but not necessarily in Baglung and Gulmi. When we report on the wage data below, we find that the class B (medium level of Maoist control according to the government classification) districts are lagging behind the class A and C districts. The wage level in the B districts has increased less, and is lower that in other districts. If the Maoists have been able to affect wages in districts of types A and C, then we would expect them to be able to affect the reported wages in the intermediate category. The more likely explanation is that the lower wages in category B districts are due to real economic differences between districts.

Table 7. Reported agricultural daily nominal wages (weighted estimates) NLSS1, hill/mountain districts

	maoA	maoB	maoC	mao	non-mao
Median wage	50.0 Rs	40.0 Rs	40.0 Rs	40.0 Rs	35.0 Rs
Mean wage	**50.1 Rs**	43.1 Rs	**47.4 Rs**	**45.9 Rs**	36.8 Rs
St. dev. wage	15.5 Rs	18.3 Rs	21.0 Rs	19.4 Rs	17.9 Rs
t-value	3.2	2.4	2.3	2.4	2.1
95% conf. int.	44.6–55.5 Rs	40.4–45.8 Rs	43.9-51.0 Rs	43.8–47.9 Rs	35.0–38.7 Rs
Lowest 25%	40.0 Rs	33.0 Rs	35 Rs	35.0 Rs	25.0 Rs
Highest 25%	60.0 Rs	47.0 Rs	60 Rs	55.0 Rs	40.0 Rs
	N = 33	N = 178	N = 139	N = 350	N = 359

Bold means that maoA and maoC are different from maoB, and mao is different from non-mao at the 95%-significance level.

We report the wages for NLSS1 in Table 7, and for NLSS2 in Table 8, in both cases we separate the Maoist-controlled districts into classes A, B and C. The (agricultural) wage is defined as in Hatlebakk (2002). That is, if a worker report more than one agricultural activity then we calculate the average wage, with activities weighted by the number of days. The worker is the unit of observation, meaning that there can be more than one observation per household. From NLSS1 we include all workers with activity code 62, which is "agricultural and animal husbandry workers" and includes 92% of the people who report wage employment in agriculture. The second most reported activity, which we do not include, is "farmers". From NLSS2, we include workers with NSCO code

921, which is "agricultural, fishery, and related workers" and includes 99% of the people who report wage employment in agriculture. For both data sets we do not include end-of-season in-kind bonuses. This is because such bonus transactions can be of many different kinds, including transactions that are typically not recorded in household surveys. The reported bonus payments are also quite low, in particular when divided by the number of days. When it comes to the number of days, the median is 60 days, while the mean is 79 days.

Table 8. Reported agricultural daily nominal wages (weighted estimates) NLSS2, hill/mountain districts

	maoA	maoB	maoC	mao	non-mao
Median wage	100.0 Rs	68.3 Rs	85.0 Rs	80.0 Rs	72.0 Rs
growth	100.0%	70.8%	112.5%	100.0%	105.7%
Mean wage	**102.3 Rs**	78.8 Rs	**95.9 Rs**	**89.4 Rs**	79.5 Rs
growth	104.2%	82.8%	102.3%	94.8%	116.0%
St. dev. wage	34.7 Rs	39.1 Rs	45.4 Rs	43.3 Rs	38.5 Rs
t-value	2.9	2.0	2.1	2.1	2.1
95% conf. int.	87.3–117.3 Rs	73.6–84.0 Rs	90.1-101.8 Rs	85.5–93.3 Rs	74.9–84.1 Rs
Lowest 25%	80.0 Rs	55.0 Rs	62 Rs	60.0 Rs	50.0 Rs
growth	100.0%	66.7%	77.1%	71.4%	100.0%
Highest 25%	120.0 Rs	90.0 Rs	115 Rs	100.0 Rs	90.0 Rs
growth	100.0%	91.5%	91.7%	81.8%	125.0%
	N = 23	N = 217	N = 234	N = 474	N = 270

Bold means that maoA and maoC are different from maoB, and mao is different from non-mao at the 95%-significance level.

We find that in NLSS1 wages are higher in Maoist districts, as compared to the non-Maoist districts. And wages increased less in Maoist sensitivity class B districts than in all other districts. The low wages in class B districts are probably real, and can be explained by local market conditions[4]. That is, the wage variation appears to be explained by real differences between districts, which may include variation in Maoist control. If the variation was, rather, due to variation in Maoist influence on the interviews, then we would expect higher wages, or a higher increase in wages, in class B districts, as compared to class C districts where they have less control.

Another indicator of reliable data is that we do not see more standardized answers in Maoist districts as measured by standard deviations as formulated in the *third hypothesis*. Rather, we find some indication of the opposite, standard deviations have increased relatively more than the mean, as measured by a decline in the t-value, in Maoist-controlled districts. We thus conclude that the reported variation in agricultural wages appears to be explained by real effects, and not by Maoist influence.

Next, we return to the *fourth hypothesis* and report on the response rate to a question where we may expect a low response, that is, purchases of a non-staple and less essential food item. We have selected apples (food code 64), see Table 9. If the Maoists, or interviewers influenced by the Maoists, wanted the interviews to be done quickly, we should expect more households to report zero apple purchases in the Maoist-controlled districts in NLSS2. However, the share of households reporting zero apple purchases is the same (not significantly different) in Maoist and non-Maoist areas in both NLSS1 and NLSS2. That is, the growth rate is the same, and there is thus no relative

[4] As nearly 50% of the workers in the class B districts live in Sindhupalchok district, and the wages are lower here than in other class B districts, the explanation appears to be district specific. For some reason there are many agricultural workers in this district, which may explain a downward pressure on wages due to a high labor supply.

decline in reported purchases in Maoist areas. However, there is a slight relative increase in the amount spent on apples in non-Maoist areas.

Table 9. Apple purchases (weighted estimates), hill/mountain districts

	NLSS1, mao	NLSS2, mao	NLSS1, non-mao	NLSS2, non-mao
Purchased	15.3%	29.6%	15.5%	32.9%
growth		93.5%		112.3%
Cost (>0) per month	43.4 Rs	62.1 Rs	44.2 Rs	**73.7 Rs**
growth		43.1%		66.7%
Cost per month, all	6.7 Rs	18.4 Rs	6.9 Rs	**24.2 Rs**
growth		174.6%		250.7%
	N = 912	N = 948	N = 633	N = 576

Bold means that non-mao is different from mao at the 95%-level.

As there has been a relatively large absolute increase in apple purchases, which probably reflects economic growth, we also look into the change in purchases of a staple food, that is, lentils (Musoro, food code 22), see Table 10. Also for lentils there has been an increase in purchases, but not to the same extent as for apples, and there has been a larger increase in Maoist areas. So, for lentils there is no support for the underreporting hypothesis. For the amount spent there are no significant differences between Maoist and non-Maoist districts.

Table 10. Lentil purchases (weighted estimates), hill/mountain districts

	NLSS1, mao	NLSS2, mao	NLSS1, non-mao	NLSS2, non-mao
Purchased	30.5%	42.7%	32.9%	**36.3%**
growth		40.0%		10.3%
Cost (>0) per month	69.1 Rs	71.8 Rs	53.8 Rs	66.4 Rs
growth		3.9%		23.4%
Cost per month, all	21.1 Rs	30.7 Rs	17.7 Rs	24.2 Rs
growth		45.5%		36.7%
	N = 912	N = 948	N = 633	N = 576

Bold means that non-mao is different from mao at the 95%-level.

To summarize, there is some indication of Maoist influence on data quality, in particular the lower standard deviation for reported landholdings in Maoist-controlled districts. However, for most indicators we find no sign of low data quality. Our main conclusion is that the Maoists have had only marginal effects on the data quality of NLSS2.

The finding is supported by interviews we have conducted with an enumerator and a supervisor. The enumerator explained in detail how interviews were conducted in a village we have visited ourselves. The information given on how they tackled the security situation was so detailed, and consistent with our knowledge of the village, that we conclude that interviews were conducted in this village. The enumerator also reported that they were able to conduct interviews everywhere, and that the supervisor probably talked with the Maoists in some villages. But, in general they tried to avoid the Maoists by traveling by public transportation, or on foot, and they did not spend time on data entry in the village, but completed the interviews and left as soon as possible. Then we interviewed a supervisor responsible for districts in another part of the country. He explained in detail how he had to bargain with the Maoists, and he explained why they allowed him to work. So, in general we have the impression that the interviews were conducted as planned in most villages.

However, in a few PSUs of the Maoist core areas we have indications, from other sources, that members of the sampled households were in stead called to the district headquarter for interviews.

5. Conclusions

This paper was motivated by the author's strong impression from 10 years of fieldwork during the civil war in rural Nepal that data from NLSS2 would be of low quality in Maoist-controlled areas. However, the data analysis indicates that the data quality is as good as in other districts, and as good as in NLSS1. We still believe that the Maoists had to approve the data collection in many villages, but the data analysis indicates that in most cases they allowed it, and did not intervene significantly in the interviews.

Note that most of the Maoist-controlled districts are in the hills. So even if one still believes that data from Maoist-controlled districts are questionable, then the terai sub-sample can be applied. Only 7 out of the 86 rural wards in the terai sample of NLSS2 are in the two Maoist-controlled districts, according to the government classification. Furthermore, these two districts, Bardiya and Dang, are both in the Mid-Western region. So, if one is very critical of data from the Maoist areas, then one may still use the sub-sample for the terai districts ranging from Kapilbastu in the west to Jhapa in the east, which gives a sample size of 996, that is, 81% of the terai sample and 36% of the national sample.

References

Bardhan, P., Baland, J.M., Das, S., Mookherjee, D. and Sarkar, R. (2002) *The Environmental Impact of Poverty: Evidence from Firewood Collection in Rural Nepal*. DP-126. The Institute for Economic Development. Boston University.

Gautam, S. (2001). *Women and Children in the Periphery of the People's War*. Institute of Human Rights Communications Nepal.

Hatlebakk (2002). *Triadic Power Relations with Production, External Markets and Multiple Agents*. Working paper. 02-18. Center for Analytical Economics. Department of Economics. Cornell University. www.arts.cornell.edu/econ/cae/02.18a.pdf.

INSEC (2005). *Human Rights Yearbook 2005*. Ed. by Gautam, K, INSEC, Kathmandu.

INSEC (2006a). *No. of Victims killed by State and Maoists in connection with "People's War" (13 Feb 1996 – 8 Oct 2006)*. Downloaded from www.inseconline.org in November 2006.

INSEC (2006b). *Conflict-induced Displacement in Nepal 2002 to 2004*. Downloaded from www.inseconline.org in November 2006.

Jacoby, H.G. (2000). "Access to Markets and the benefits of Rural Roads" *Economic Journal*. **110**(465): 713-737.

Koolwal, G.B. (2007). "Son Preference and Child Labor in Nepal: The Household Impact of sending Girls to Work". World Development. **35**(5): 881-903.

Murshed, S.M. and Gates, S. (2005). "Spatial–Horizontal Inequality and the Maoist Insurgency in Nepal". *Review of Development Economics*. **9**(1): 121-134.

Sharma, S. (2003). "The Maoists Movement: An Evolutionary Perspective" in Thapa (2003).

Thapa, D. (2003) (ed.). *Understanding the Maoists Movement of Nepal*. Chautari Books Series - 10. Kathmandu.

Recent Working Papers

WP 2007: 4
TØNNESEN, Liv: Gender Citizenship in Sudan: Competing Perceptions of Women's Civil Rights within the Family Laws among Northern and Southern Elites in Khartoum. Bergen, 2007, 12 pp.

WP 2007: 3
ESPELAND, Rune Hjalmar: When Neighbours Become Killers: Ethnic Conflict and Communal Violence in Western Uganda. Bergen, 2007, 15 pp.

WP 2007: 2
KOLSTAD, Ivar: The Resource Curse: Which Institutions Matter? Bergen, 2007, 7 pp.

WP 2007: 1
KNUDSEN, Are: The Law, the Loss and the Lives of Palestinian Refugees in Lebanon. Bergen, 2007, 25 pp.

WP 2006: 21
WIMPELMANN, Torunn: The Aid Agencies and the Fragile States Agenda. Bergen, 2006, 26 pp.

WP 2006: 20
SKAAR, Elin and José Octávio Serra Van-Dúnem: Courts under Construction in Angola: What Can they do for the Poor? Bergen, 2006, 17 pp.

WP 2006: 19
KJØSTVEDT, Hilde: Palestinian Women:Is there a Unitary Conception of Rights. Bergen, 2006, 14 pp.

WP 2006: 18
MÆSTAD, Ottar et al. Distribution Matters: Expressed Value Judgements among Health Planners in Tanzania. Bergen, 2006, 17 pp.

WP 2006: 17
MÆSTAD, Ottar and Ole Frithjof Norheim: Eliciting People's Preferences for the Distribution of Health: A Procedure for a more Precise Estimation of Distributional Weights. Bergen, 2006, 23 pp.

WP 2006: 16
VILLANGER, Espen: Is Bonded Labor Voluntary? Evidence from the Liberation of the Kamaiyas in the far-Western Region of Nepal. Bergen, 2006,

WP 2006: 15
FJELDSTAD, Odd-Helge: Local Revenue Mobilization in Urban Settings in Africa. Bergen, 2006, 24 pp.

WP 2006:14
KNUDSEN, Are: Hamas and the Quest for Palestinian Statehood. Bergen, 2006, 14 pp.

CMI's publications, Annual Report and quarterly newsletters are available on CMI's homepage www.cmi.no

CMI
Chr. Michelsen Institute
Bergen, Norway
Tel: +47 55 57 40 00
Fax: +47 55 57 41 66
cmi@cmi.no
www.cmi.no

SUMMARY

We investigate whether the quality of LSMS data from Nepal is affected by the Maoists in the districts they control. We find, if any, only minor support for the hypothesis. Furthermore, the Maoists have less control in the plains (terai), where a majority of the population lives, so data from the terai sub-sample of NLSS2 is, in particular, not likely to be biased by the Maoists.

ISSN 0804-3639
ISBN 978-82-8062-207-5