LSMS data quality

in Maoist influenced areas of Nepal

by

Magnus Hatlebakk[*]

CMI, Bergen, Norway

magnus.hatlebakk@cmi.no

June 14, 2007

Abstract: We investigate the hypothesis that 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 due to the Maoists.

[*] The research is funded by the Social Inclusion Research Fund, which in turn is funded by the Norwegian Embassy, Nepal.

1. Introduction

In this paper we investigate the data quality of the second round of the Nepal Living Standard Survey (NLSS2). Throughout the civil war the author conducted fieldwork in Maoist influenced areas, and from our own experiences, 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 degree of control varies geographically, we shall expect the data quality to vary as well. Now, the NLSS2 team was not able to conduct any interviews in some few villages that were originally in the random sample. However, these villages are mostly located in the Far-Western region. This adds to our skepticism regarding the data quality, as the Maoists have their main strongholds, not in the Far-Western, but in the Mid-Western region, in particular Rolpa and Rukum districts. Based on what field-assistants have told us, and our own experience from encounters with local Maoist leaders, they are not very willing to allow household interviews. To cite one 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 people"*. The concern is that in districts like Rolpa and Rukum, the Maoists have either participated in filling the forms, asked the enumerators to fill the forms without conducting interviews, or they have 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 there are only 12 households in each 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 them 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 longer, the teams probably spent at least one extra day in each PSU, and in some cases had to return days later, when they got the permission from the Maoists.

The second hypothesis is that interviews conducted under the supervision of the Maoists will have more homogenous answers. That is, we expect the standard deviation of the measured variables to be smaller in these villages, as compared to comparable villages from the first round of the Nepal Living Standard Survey (NLSS1), and as compared to villages from NLSS2 that are not controlled by the Maoists. And, we expect some numbers to appear more frequently. For example, we expect ages of 5, 10 and 15 years to be more frequently reported. Now, we have to carefully select the variables when we test this hypothesis. The Maoist presence may have had real effects, in particular on the land-distribution and the wage level. We will still report these variables, as they will probably be applied in future analyses of the NLSS2 data, but we also include less essential variables where we expect to find data errors, but no real Maoist influence. If there are effects of Moist control also on less important variables, such as the age of children, then we may also expect the landholdings, and wages, to be incorrect.

If forms are filled without any real interviews taking place, we may also expect that more boring issues, or questions at the end of the questionnaire, would not be filled at all, or in a very standardized manner. The third hypothesis is thus that we shall expect low response rates on some questions. This was probably the case also for NLSS1, but we expect even lower response rates, and standard deviations in NLSS2 if there is a Maoist influence. Examples of variables can be consumption, and purchases, of various food items, as there were 5 pages with detailed questions on approximately 50 different food-items in the NLSS2-questionnaire.

As said, 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, which have been using 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, which explains why we apply the cross-sectional samples. Section 4 presents the findings on any data errors that might be due to Maoists influence. 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 for 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 the killings conducted by the state and by the Maoists, while we add up 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 cutoff for 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 was killed during the 10 years of conflict.

This is approximately equal to the national average, where about 12 000 people were 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 was 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 also 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 Maoists strongholds in Rolpa and Rukum. But this is not the normal situation. Most attacks, with the majority being initiated by the Maoists, take place in the surrounding 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 that the Maoists have their base-areas in the hill districts, but attack army barracks and also civilians in the contested areas in the terai (the plains along the boarder to India). We find a similar trend within 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 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 for the strengths of the insurgents, which we considered, is the voter-turnout in elections. The Maoists have boycotted and interrupted the last elections, and have been more successful in doing so within their

base-areas. Now, there have been no election in Nepal since 1999, and the Maoists have increased their presence, for example in the eastern hills, since then. 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, where we may set the cutoff 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 the displacement "due to the state" is unevenly distributed between the districts, with 67% from a single district, Kapilbastu. 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 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 in 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 indicate 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 are more recently controlled by the Maoists, such as Terathum in the eastern hills. However, the list does not include Bardiya, which appears to us to be a Maoists 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. In the analysis conducted in this paper we thus apply this indicator, combining 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 about here

As we may expect, killings and displacements are correlated, and the government classification is correlated with the Maoist's own classification. 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 included in three columns are also included in the government indicator, and all districts that are included 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 on any of the other indicators. 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 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 there are only two terai districts that 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 cover any Maoist controlled district.

3. Selection bias in the panel data

NLSS2 also includes a panel, approximately 36% of the households from NLSS1 were included in a random sample of 1232 households. However, only 78% of these households were identified, and the missing 22% is 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%) 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 panel, 992 are from wards that were rural in NLSS1[1]. Among them 792 was 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, this percentage is now 7%, while the remaining 128 (13%) were not found by the enumerators. In NLSS1 there was 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

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

well as the three potentially non-random samples of size 791, 72, and 128, respectively.

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 final 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 the average of 124 000 among the households the team did not find. 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 (biased) panel and the random sample is not significant, we may expect unbiased estimates in regression analyses that apply the panel data. However, the significant 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. As said, in the present paper we do not use the panel, and only report on the cross-section samples.

Table 2 about here

We would like to make an additional note on the selection of PSUs for the NLSS2 cross-section. There appears to be 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 second round. But, we must keep in mind that if one, for example, applies data from only one district, then the selected wards (PSUs) are not the same in the two cross-sections. With the small number of observations from each district, the confidence-intervals for variables measured at the district level are quite high, and one should be careful in comparing the cross sections 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 has 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 four main hypotheses: 1) The survey teams spend additional days in each village due to bargaining with the Maoists. 2) The response rate is lower on some questions. 3) The responses are more standardized, with, for example, numbers of 5 and 10 being overrepresented. 4) The standard deviation of the answers will be relatively smaller.

We check these hypotheses by use of different sections of the NLSS2 data. Not all hypotheses are relevant for each section, so the presentation is organized by type of data, and not according to the sequence of hypotheses. We compare the responses for the Maoist controlled districts in NLSS2, with the remaining districts, and with the same districts from NLSS1. We only use the rural sub-samples, as the Maoists do not control urban areas. 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 test for 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 only use the cross-section data, which consists of two independent random samples. We intend to use the cross-sections also in adjoining papers, for the same reasons. In case other researchers use the panel data they have to check for selection biases.

We start investigating the first hypothesis. In NLSS1 2655 households in 215 rural wards report on the interview date, while NLSS2 has 2748 households in 229 rural wards. Due to the small population the districts Rasuwa and Mustang were not represented in the random samples. It is somewhat surprising that this happened in both rounds, as Manang has a much lower population, that is 1/5 of Rasuwa, but still is represented in both samples. In addition, the districts of Dolpa in the first round, and Achham in the second round, were not included due to problems with data collection. So in both rounds there are 72 districts, and roughly 32 are classified as Maoist controlled districts depending on the indicator, as discussed in section 2.

We have corrected dates that appear to be obvious errors. For NLSS2 we did this for 8 wards. The 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. In case this 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 right at the end of the month, for example, interviews on day 29 to 31, and day 2 to 4 in the following month, is 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.

The number of days in each ward in NLSS1 is tabulated in Table 3 for the 192 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. The median, and the most common, was 4 days of interviews, while 92% of the villages took 6 days or less. Normally it was three enumerators in each ward, meaning that they did 4 interviews each on average, during a median number of four days. So, one interview per day was the normal case.

Table 3 about here

In NLSS2 there are 12 households in all wards. Two wards have missing information on dates, so we are left with 227 wards. The number of days in NLSS2 is also tabulated in Table 3. As for NLSS1, the median, and the most common, was 4 days. And, also for NLSS2 92% of the villages took 6 days or less. And, as for NLSS1 each enumerator interviewed normally 4 households during the median of four days, so, again, the normal case was one interview per day.

There are some differences in the distribution of days, NLSS2 is concentrated on 4 and 5 days, while NLSS1 is more diverse with concentration from 3 to 6 days. If we take into account the separation into Maoist and non-Maoist districts, using the government classification from Table 1, then we find that the pre-conflict NLSS1 data

has a concentration on 4-6 days in Maoist districts, while it is 3-4 days in the non-Maoist districts. In NLSS2 we do not find such differences, see Table 4. There thus appears to be a decline in number of days in Maoist districts, and an increase in non-Maoist districts. This is the opposite of what we expected. However, the decline is not significant. Still, it may reflect that the Maoists demanded effective interviews, but, since there is basically no difference between Maoist and non-Maoist districts in NLSS2, we rather believe that the districts have become more similar. In 1995/96 the logistics of the more remote wards in Maoist controlled districts probably slowed down the data-collection. So, with 4-5 days of data collection in Maoist, as well as 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 in Rolpa and Rukum, where the interviews took respectively 4, 4, 5, and 6 days.

Table 4 about here

Next, we analyze the age data. In Table 5 we report the ages of all household members reported in the rural NLSS1 and NLSS2 samples. As the ages 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 5-years 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 about here

Overrepresentation of the fives is most prominent in non-Maoists districts in NLSS1, where the factor is 3 or larger for all ages from 35 to 75. That is, the fives are reported three times as much as the nearest ages. We do not see the same tendency in the Maoist districts at that time, and we do not see such a strong tendency in any districts in NLSS2. So, from NLSS1 to NLSS2 there is only a slight increase in the overrepresentation in Maoist districts for the older ages of 55, 60, 70 and 75., while there is a decrease in the non-Maoist districts. We interpret the NLSS2 findings as a normalization, where the non-Maoist districts still have the highest overrepresentation of the fives. This interpretation is supported by the fact that the lower fives are not overrepresented. So, elderly people tend not to remember their exact age wherever they live.

We still want to explain the overrepresentation of the fives in non-Maoist districts in NLSS1. Looking into the disaggregated data we find that it is due to the fact that the Maoist districts are mainly located in the hills. If we construct the same fractions only for the hill districts, then we find no particular pattern, Maoist and non-Maoist districts have practically the same degree of overrepresentation of the fives, see Table 6. For the hills there is, rather, an increase in the overrepresentation in non-Maoist districts. But again no support for the hypothesis of standardized answers in Maoist controlled districts. We conclude that there are data-errors everywhere for high ages, but not more so in Maoist controlled districts.

Table 6 about here

This far, it appears that real interviews have been conducted. Still, there is a suspicion that the teams selected some households, and not necessarily the correct

ones, in the district head-quarter to avoid visiting the villages. If that was the case, then one may expect that these households have less information on the details of agricultural activities. For example, if the team has been able to locate a family member of a particular household that lives in the district headquarter, then this family member may know the total land holdings of the household, but not necessarily all the different plots. We thus now report on the number of plots, in addition to other measures of land holdings.

Now, we shall expect that some landlords have sold land in the Maoist areas during the conflict period, and thus may own fewer plots, and also on average the land sales may lead to consolidation of plots in these villages. Still, if interviews are conducted in the district head-quarter, or under Maoist supervision, we shall, on average, expect that some of the plots are not reported, which means that the total amount of land reported within the district should have declined. However, there is an underlying increase in agricultural land, as well as in households that own land. As the number of households increases faster than the land, the average landholding is declining, in particular in terai. Subject to this underlying trend, our hypothesis will be that the reported average land holding declines faster in Maoist controlled districts than elsewhere. We report the findings on the full sample in Table 7 and the hill sample in Table 8. We see that the mean land holding is declining, but more so in the non-Maoist terai districts. In the hill sample, the average is apparently declining in the Maoist districts only, but the decline is not significant. So, there is some, but insignificant, support for the hypothesis of underreporting of land.

Table 7 about here

When it comes to the average number of plots, it is significantly reduced in the Maoist district, which indicates a consolidation of plots, which, in turn, is probably explained by real land-sales in the Maoist districts. This is supported by an apparent decrease in the larger holdings. To conclude, there is a reported, but insignificant, decrease in the mean land holding. And, there is a significant consolidation of plots, where people sell of plots to reduce their land holdings in the Maoist controlled districts. We cannot exclude the possibility of under-reporting of plots, but we find consolidation to be a more likely explanation than the under-reporting hypothesis.

Table 8 about here

We recall that another hypothesis was to expect a smaller standard deviation in the Maoist controlled districts of NLSS2. There is a decline from NLSS1 in the Maoist controlled districts, even when compared to the mean value. But, again this can be due to the sales of land at larger farms.

We now go on to the next hypothesis. One may imagine that the Maoists allowed proper interviews to take place, but that they participated and made sure that some important answers corresponded with their policy. In particular, one may imagine that they would 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 policies from misrepresentation of the wage. However, we may argue that the policy is working 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 in Baglung and Gulmi. When we report on the wage data below, we find that the class B (according to level of Maoist control) districts are lagging behind the class A and C districts, which is also the case if we only use data from the hills. The wage level in

the B districts has increased less, and is lower that in other districts according to NLSS2. If the Maoists have been able to affect the wages in districts of type A and C, then we would expect them to be able to affect the reported wages in the intermediate category as well. The more likely explanation is that the lower wages in category B districts is due to real economic differences between districts.

Table 9 about here

We report the wages for NLSS1 in Table 9, and for NLSS2 in Table 10, where in both cases we separate the Maoist controlled districts into class A and B. Then we compile the Maoist districts into a single category, and report both NLSS1 and NLSS2 data for the hill districts in Table 11. 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. 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 thus 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 this kind of bonus transactions can be of many different kinds, including transactions that are typically not recorded in this kind of interviews. 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 10 about here

We find that in NLSS1 the wages are higher in all Maoist districts, as compared to the non-Maoist districts. Next, the wages increased less in Maoist sensitivity class B districts, than in all other districts. So in NLSS2 the mean wages are lower there than in other Maoist districts. Also the wages at the 25% centiles are lower than in the Maoist core areas. We conclude that the Maoists have been able to increase the wages in their core, class A, districts, but not to the same extent in class B districts, and they allowed this to be reported to NLSS. We conclude that also the wage data appears to be reliable within the Maoist controlled areas, as the Maoists allowed low wages to be reported in class B districts. Another indicator of reliable data is that we do not see more standardized answers as measured by smaller standard errors. We find some indication of the opposite, standard errors have increased relatively more than the mean in Maoist controlled districts. This can be explained by the reported difference between class B and the two other classes of Maoist districts.

Table 11 about here

Next, we look at 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 12. In case the Maoists, or the interviewers, due to the Maoists, wanted the interview to be quickly done, we should expect more households to report zero apple purchases in the Maoist controlled districts in NLSS2. However, the increase in the share of households that purchases apples is the same here as in the non-Maoist districts. If we only look at the hills, then there is some

difference in the growth rate. That is, most Maoist districts are in the hills, so the growth rate for the hills is approximately the same as reported for all districts in Table 12, but the non-Maoist hill districts have a larger growth rate. So, there is some support for the hypothesis that apple purchases are underreported in the Maoist controlled districts, although the lower growth rate can also be a real effect.

Table 12 about here

As there has been a relatively large 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 13. 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. In the hills the increase in non-Maoist areas is not even significant. So, for lentils there is no support for the underreporting hypothesis, that is, no support for the second hypothesis.

Table 13 about here

To summarize, it appears that the lower growth in apple purchases in Maoist districts as compared to other hill districts is the only support we have for the hypotheses listed at the start of this section. However, the difference in the share of households reporting purchases of apples in NLSS2 is not large, the share is 30.4% in Maoist districts and 37.9% in non-Maoist districts. So even this evidence is not strong, and may, of course, be a real effect of the Maoist insurgency, as the Maoists have regularly interrupted transportation of goods. The main conclusion is thus that it

appears that the Maoists have had (at best) only marginal effects on the data-quality of NLSS2.

The finding is supported by interviews we have conducted with an enumerator as well as 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 he told 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 by 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 that a member of the sampled households were, in stead, called to the district head-quarter for interviews.

4. 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 the Maoist controlled area. 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 they in most cases 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 Maoists 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 to 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

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.

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.

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

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

Table 3. Days spent in each ward, NLSS

	NLSS1			NLSS2		
Days in the ward	No. of wards	Percent	Cumulative	No. of wards	Percent	Cumulative
2	4	2%	2%	2	1%	1%
3	33	17%	19%	18	8%	9%
4	70	36%	**56%**	97	43%	**52%**
5	32	17%	72%	67	30%	81%
6	38	20%	92%	25	11%	92%
7	8	4%	96%	14	6%	98%
8	2	1%	97%			
9	3	2%	99%			
10	2	1%	100%	3	1%	100%
11				1	0%	100%
	N = 192	100%		N = 227	100%	
Mean (weighted)			4.65 days			4.69 days

Table 4. Days spent in each ward, NLSS2, Maoist controlled, and not

	Maoist controlled			Non-Maoist districts		
Days in the ward	No. of wards	Percent	Cumulative	No. of wards	Percent	Cumulative
2	1	1%	1%	1	1%	1%
3	5	6%	7%	13	9%	10%
4	37	43%	**50%**	60	43%	**52%**
5	27	31%	81%	40	28%	81%
6	10	12%	93%	15	11%	91%
7	3	3%	97%	11	8%	99%
10	3	3%	100%	0		
11	0			1	1%	100%
	N = 86	100%		N = 141	100%	
Mean, NLSS2 (weighted)			4.72 days			4.68 days
Conf. int.			(4.44 – 5.01)			(4.48 – 4.88)
Mean, NLSS1 (weighted)			5.03 days			4.40 days
Conf. int.			(4.73 - 5.32)			(4.12 – 4.68)

Table 5. Ages reported in NLSS1 and NLSS2

Age	NLSS1	NLSS2	NLSS1, mao	NLSS1, non-mao	NLSS2, mao	NLSS2, non-mao
5	1.08	1.03	0.98	1.14	1.13	0.99
10	1.10	1.10	1.06	1.12	1.13	1.07
15	0.99	1.19	1.03	0.96	1.31	1.11
20	1.18	1.05	0.96	1.30	1.08	1.04
25	**1.86**	1.47	1.52	***2.07***	1.25	*1.60*
30	**2.28**	1.77	1.94	***2.49***	1.81	1.75
35	**2.78**	2.21	**2.19**	***3.12***	1.89	*2.41*
40	**2.90**	2.39	**2.70**	***3.03***	2.15	*2.55*
45	**3.60**	2.42	**2.90**	***4.02***	2.16	*2.58*
50	**3.11**	1.84	**2.36**	***3.66***	1.58	*1.98*
55	**3.13**	2.64	1.93	***4.20***	**2.41**	*2.77*
60	3.63	**3.96**	3.31	*3.82*	**3.91**	4.00
65	**3.52**	3.04	2.06	***4.79***	2.18	*3.68*
70	**3.49**	2.57	1.80	***5.55***	**2.31**	*2.77*
75	3.33	3.12	2.00	***4.24***	**2.84**	*3.33*
N	15095	14679	5462	9633	5342	9337

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

Table 6. Ages reported in NLSS1 and NLSS2, only hills

Age	NLSS1, mao	NLSS1, nonmao	NLSS2, mao	NLSS2, nonmao
5	0.99	1.19	1.13	1.10
10	1.02	0.90	1.14	0.97
15	1.01	0.99	1.46	0.86
20	1.00	1.12	0.98	1.00
25	1.46	0.91	1.26	1.16
30	***2.16***	1.44	1.71	1.64
35	1.84	1.66	*1.88*	1.43
40	*2.34*	1.57	2.14	**2.20**
45	***2.70***	1.81	2.02	***2.86***
50	1.86	2.11	1.52	*2.22*
55	*2.17*	1.67	*1.90*	1.54
60	*3.26*	2.00	3.24	**3.18**
65	*1.96*	1.43	1.82	***2.43***
70	2.04	***3.50***	1.84	*2.88*
75	0.93	*2.00*	***2.86***	**2.43**
N	4010	2091	3948	1807

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

Table 7. Reported land-holdings (weighted estimates)

	NLSS1, maoist	NLSS1, non-mao	NLSS2, maoist	NLSS2, non-mao
Landless	5.0%	20.0%	5.6%	22.2%
Median holding*	0.70 bigha	0.58 bigha	0.71 bigha	0.50 bigha
Mean holding	1.06 bigha	1.25 bigha	0.98 bigha	0.98 bigha
St. dev.	1.32 bigha	2.03 bigha	1.17 bigha	1.64 bigha
5 bigha+	2%	5%	1%	2%
Median no. plots	3.0 plots	2.0 plots	3.0 plots	2.0 plots
Mean no. plots	3.9 plots	2.9 plots	3.1 plots	2.4 plots
Median value Rs	70 000	60 000	129 000	130 000
Mean value Rs	164 000	224 000	261 000	344 000
300 000 Rs +	15%	14%	26%	33%
	N = 996	N = 1660	N = 1032	N = 1716

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

Table 8. Reported land-holdings (weighted estimates), only hills

	NLSS1, maoist	NLSS1, non-mao	NLSS2, maoist	NLSS2, non-mao
Landless	4.3%	11.0%	5.8%	11.3%
Median holding*	0.69 bigha	0.60 bigha	0.71 bigha	0.50 bigha
Mean holding	0.98 bigha	0.95 bigha	0.91 bigha	0.96 bigha
St. dev.	1.14 bigha	1.09 bigha	0.90 bigha	1.08 bigha
5 bigha+	1%	1%	1%	1%
Median no. plots	3.0 plots	3.0 plots	3.0 plots	2.0 plots
Mean no. plots	3.8 plots	3.5 plots	3.1 plots	3.1 plots
Median value Rs	70 000	73 000	120 000	125 000
Mean value Rs	169 000	387 000	237 000	371 000
300 000 Rs +	15%	15%	25%	32%
	N = 744	N = 392	N = 780	N = 372

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

Table 9. Reported agricultural daily nominal wages (weighted estimates) NLSS1

	NLSS1, maoA	NLSS1, maoB	NLSS1, maoist	NLSS1, non-mao
Median wage	50.0 Rs	40.0 Rs	40.0 Rs	32.4 Rs
Mean wage	50.1 Rs	45.1 Rs	46.9 Rs	36.2 Rs
St. dev. wage	15.5 Rs	18.1 Rs	18.6 Rs	18.1 Rs
t-value	3.2	2.5	2.5	2
95% conf. int.	44.6 – 55.5 Rs	42.6 – 47.7 Rs	45.0 – 48.7 Rs	35.3 – 37.2 Rs
Lowest 25%	40.0 Rs	35.0 Rs	35.0 Rs	25.0 Rs
Highest 25%	60.0 Rs	50.0 Rs	60.0 Rs	45.0 Rs
	N = 33	N = 203	N = 407	N = 1410

Table 10. Reported agricultural daily nominal wages (weighted estimates) NLSS2

	NLSS2, maoA	NLSS2, maoB	NLSS2, maoist	NLSS2, non-mao
Median wage	100.0 Rs	70.0 Rs	80.0 Rs	62.0 Rs
growth	100%	75.0%	100 %	91.4%
Mean wage	102.3 Rs	83.1 Rs	90.3 Rs	69.5 Rs
growth	104.2%	84.3%	92.5%	92.0%
St. dev. wage	34.7 Rs	39.1 Rs	41.7 Rs	30.5 Rs
t-value	2.9	2.1	2.2	2.3
95% conf. int.	87.3 – 117.3 Rs	78.1 – 88.0 Rs	86.7 – 93.9 Rs	67.8 – 71.2 Rs
Lowest 25%	80.0 Rs	60.0 Rs	60.0 Rs	50.0 Rs
growth	100%	71.4%	71.4%	100%
Highest 25%	120.0 Rs	100.0 Rs	104.1 Rs	80.0 Rs
growth	100%	100%	73.5%	77.8%
	N = 23	N = 242	N = 523	N = 1273

Table 11. Reported agricultural daily nominal wages (weighted estimates), only hills

	NLSS1, maoist	NLSS1, non-mao	NLSS2, maoist	NLSS2, non-mao
Median wage	40.0 Rs	35.0 Rs	80.0 Rs	72.0 Rs
growth			100 %	105.7%
Mean wage	46.6 Rs	36.9 Rs	92.6 Rs	76.8 Rs
growth			98.7%	108.1%
St. dev. wage	19.2 Rs	18.6 Rs	43.4 Rs	40.6 Rs
t-value	2.4	2.0	2.1	1.9
95% conf. int.	44.2 – 49.0 Rs	34.5 – 39.3 Rs	88.0 – 97.2 Rs	71.0 – 82.6 Rs
Lowest 25%	35.0 Rs	25.0 Rs	65.0 Rs	50.0 Rs
growth			85.7%	100%
Highest 25%	55.0 Rs	42.0 Rs	106.1 Rs	90.0 Rs
growth			92.9%	114.3%
	N = 250	N = 234	N = 345	N = 192

Table 12. Apple purchases (weighted estimates)

	NLSS1, maoist	NLSS1, non-mao	NLSS2, maoist	NLSS2, non-mao
Purchased	15.7%	22.8%	29.5%	42.9%
growth			87.9%	88.2%
Cost (>0) per month	45.8 Rs	39.8 Rs	63.7 Rs	67.4 Rs
growth			39.1%	69.3%
Cost per month, all	7.2 Rs	9.1 Rs	18.8 Rs	28.9 Rs
growth			161.1%	217.6%
	N = 996	N = 1661	N = 1032	N = 1716

Table 13. Lentils purchases (weighted estimates)

	NLSS1, maoist	NLSS1, non-mao	NLSS2, maoist	NLSS2, non-mao
Purchased	28.9%	44.3%	41.4%	50.4%
growth			43.3%	13.8%
Cost (>0) per month	69.7 Rs	78.6 Rs	74.2 Rs	88.2 Rs
growth			6.5%	12.2%
Cost per month, all	20.2 Rs	34.8 Rs	30.7 Rs	44.5 Rs
growth			52.0%	27.9%
	N = 996	N = 1661	N = 1032	N = 1716