Journal of Data Science **8**(2010), 289-306

Multilevel Models and Inequality in Viet Nam

Dominique Haughton[1,2] and Phong Nguyen[3]
[1]*Bentley University,* [2]*Toulouse School of Economics*
and [3]*General Statistics Office*

Abstract: This paper proposes to investigate inequality in Viet Nam from the point of view of a study of the urban/rural gap by means of a multilevel model. Using data from the Viet Nam Household Living Standards Survey of 2002, the paper constructs a multilevel model, yielding random effects in the urban/rural gap which can be seen as location-specific random contributions to the urban/rural gap above and beyond the effects of known location characteristics, such as the level of education of the population, etc. The paper also demonstrates how the multilevel model can be used to obtain small area estimates at the commune level.

Key words: Inequality, multilevel models, small area estimation, viet nam household living standards surveys.

1. Introduction: Summary of the Literature on Inequality from the Point of View of Living Standards Data

In this article, we propose to examine the issue of inequality in Viet Nam from the point of view of living standards data, and more specifically with the help of a multilevel model, to be discussed below. The matter of inequality in Viet Nam, among other issues whether inequality is increasing or not with time, and how inequality might vary geographically across the country, is of very current interest, in part because of past experience with China where very fast growth seems to have been accompanied by a fast increase in inequality. This section briefly summarizes some relevant literature among the large body of work about inequality in China, and describes literature that pertains to the issue of inequality in Viet Nam.

Literature on inequality in China

The literature on inequality in China is abundant, and to a large extent motivates similar research on Viet Nam. We briefly summarize the issues raised in recent research below.

Ravallion and Chen (2007) point to China's (uneven) progress against poverty, and find that provinces starting with relatively high inequality saw slower progress against poverty. Two articles investigate regional and provincial differences in access to education (Hannum and Wang, 2006) and foreign direct investment (Ma, 2006).

Literature on inequality in general, particularly in the case of China, includes an active discussion on the relationship between inequality and growth. Answers are mitigated, with recent papers by Zhang and Wan (2006) and Wan, Lu and Chen (2006) pointing to an enduring negative relationship between inequality and growth. Other researchers sometimes identify negative relationships in the short term, but non-significant or even positive relationships in the longer term (see the discussion in Wan, Lu and Chen, 2006). We do not address this complicated issue in this paper, but find that this discussion is worth mentioning — albeit very briefly — since it helps underline the importance of a study of inequality.

There is also a belief by some authors (see Démurger, Fournier and Li, 2006) that measures of inequality, increase in inequality as well as regional differences in these measures in China are overstated because of the use of inappropriate price deflators.

We finally mention two papers which address the matter of inequality and health: Li and Zhu (2006) and Zhao (2006).

Literature on inequality in Viet Nam

The literature on inequality in Viet Nam is not as voluminous as that in China, and many interesting questions are still open. We refer below to papers we feel are among the most relevant pieces of work.

Recently Nguyen Binh *et al.* (2007), in an important paper, investigated urban-rural inequality in Viet Nam via quantile regressions, using living standards surveys from 1993 and 1998. A current belief among economists and others is that the relatively modest increase in inequality in Viet Nam (measured by the Gini index, for example) may be due to increases in the urban-rural gap in living standards.

The gender gap in wages, and its evolution, are addressed by two authors (Liu, 2006, using data from 1992-98, and Pham and Reilly, 2006, using data from 1993-2002) with a tentative conclusion that the gap has been reduced in recent years. However that matter is very much complicated by the fact that wages are earned only by those working in jobs with salaries, excluding those that are self-employed, for instance, a serious limitation in Viet Nam, particularly when looking at a gender gap. Liu does find that female participation in the labor force increased for younger women but decreased for women aged 35-44. In this paper, we refrain from investigating this aspect of inequality, among other reasons because of these problematic issues, although we recognize that it is a very interesting matter.

Using the panel of households common to the living standards surveys of 1993 and 1998, Haughton *et al.* (2001) aimed to identify "shooting stars" and "sinking stones", that is, households with a sharp increase (two quintiles at least) or decrease (two quintiles at least) in living standards. Regional differences were found, as well as interactions effects between regions and other variables.

Finally we mention work by Huong *et al.* (2006) which related socio-economic status inequality and major causes of death in adults.

2. Data Used in this Paper

The model used in this paper relies on data from the Viet Nam Household Living Standards Survey (VHLSS) of 2002, and we refer to past work which relies on the Viet Nam Living Standards Surveys (VLSS) of 1993 and 1998. In this section we describe a few relevant features of the surveys, and the context in which the VLSS, and then the VHLSS program were established under the auspices of the General Statistics Office (GSO 1999) in Viet Nam. Further details are available from Nguyen Phong and Haughton (2006).

The VLSSs were implemented in 1993 and 1998 in Viet Nam with financial support from the United Nations Development Program (UNDP) and the Swedish International Development Agency (SIDA), and with technical support from the World Bank. The survey methodology follows the World Bank's Living Standards Measurement Study (LSMS), listed in the bibliography, covering the following areas displayed in Table 1.

Areas covered by the VLSS surveys (1993 and 1998) are as listed below:

1. Income
2. Expenditure
3. Education
4. Health (including height, weight and arm circumference of all household members)
5. Employment
6. Agricultural activities
7. Non-farm business activities
8. Housing

9. Migration

10. Fertility

11. Savings and credit

The sample size was of 4,800 households in 1993 and 6,000 households in 1998, including 4,300 1993 households which were re-interviewed. The sample was divided into 10 parts and each month one tenth of the sample was covered by the VLSS, in an attempt to avoid seasonal effects.

The questionnaire wrote out the exact questions to be used by the interviewers, and data entry was performed in the field. The survey involved a very high rate of supervision, with one supervisor for every two interviewers.

The VLSS data are widely considered to be of very high quality; however some limitations include the fact that no direct estimates for provincial level were possible because of the relatively small sample size, that a long period of time elapsed between the two VLSSs, and finally that the cost of the survey was high, at $163 per household interviewed. The VHLSS program was established to try to address some of these limitations.

The VHLSS program

During 2000-2010 the plan is for the GSO to conduct a Viet Nam Household Living Standards Survey every two years: 2002, 2004, 2006, 2008 and 2010. So far, VHLSS 2002, 2004, 2006 and 2008 have been collected. Each survey year, a core module is conducted. Every four years or more, additional modules are conducted. Topics for core and additional modules are displayed below:

1. Basic demographic information on all household members (age, sex, relationship to head)

2. Household expenditures (food, education, health, etc.)

3. Household income (wage and salary, farm production, non-farm production, remittances, etc.)

4. Employment and labor force participation

5. Education: a small number of questions (literacy, highest diploma, fee exemption)

6. Health: a small number of questions (use of health services, health insurance)

7. Housing: a small number of questions (type of housing, electricity, water source, toilet, etc.)

8. Assets and durable goods

9. Participation in poverty programs

10. A commune questionnaire with information on local infrastructure

There are also additional modules in the VHLSS program:

1. Detailed information on agricultural activities and non-agricultural household businesses, borrowing and lending.

2. Detailed information on health and education of household members. Questionnaires for commune health centers and local schools.

3. Infrastructure, environment, local institutions and governance.

VHLSS 2002

The data used in this paper originate from VHLSS 2002. We summarize below the main features of this survey and its questionnaire design:

- VHLSS 2002 covered only the core module.

- The 2002 VHLSS questionnaire is similar in many respects to the 1997-98 VLSS questionnaire.

- Six of the 9 sections in the questionnaire are very similar to the 97-98 VLSS: Household Roster, Education, Employment; Income; Housing; Food Expenditures and Non-Food Expenditures.

- The Health Section is similar to that in the 1997-98 VLSS, but also incorporates ideas from the 2001-2002 Viet Nam Health Survey.

The sample size in 2002 is of 75,000 households (of which 30,000 are expenditure households, which implies that questions were asked both about income and expenditures of these households).

Questionnaire design for VHLSS 2002 (and later years)

- The exact questions asked of households are printed out in the questionnaire
- Questions were designed to ensure comparability with past surveys, especially for expenditure and income data
- The data entry was performed at the provincial level
- The field work was conducted as follows:
  — VHLSS 2002: Four rounds (four quarters)
  — VHLSS 2004 and 2006: Two rounds (May and September)
- Personnel and Training Issues: — Field workers are GSO staff members
  — Training for trainers was held in the North and South of the country
  — Training for interviewers was held in each province

The description of variables from VHLSS 2002 which are used in our model are listed below:

- Y Dependent variable: Logarithm of real per capita expenditure (in '000 VND) in VHLSS 2002
- Independent variables
  1. X_1 Marital status of household head (1 if married, 0 if not)
  2. X_2 Age of household head, in years
  3. X_3 Gender of household head (1 for male, 0 for female)
  4. X_4 Ethnicity of household head (1 if Kinh, 0 if not)
  5. X_5 Household size
  6. X_6 Urban (1 if urban household, 0 if not)
  7. X_7 Number of years of education of household head
  8. X_8 Leadership job of household head (1 if yes, 0 if not)
- Skill level of job of household head Reference: Non-skilled farm job of household head
  1. X_9 High-skilled job of household head (1 if yes, 0 if not)
  2. X_{10} Medium-skilled job of household head (1 if yes, 0 if not)
  3. X_{11} Low-skilled job of household head (1 if yes, 0 if not)
  4. X_{12} Non-skilled non-farm job of household head (1 if yes, 0 if not)

5. X_{13} Squared age of household head

6. X_{14} Interaction of urban and number of year of education of household head

3. Descriptive Analysis of Inequality on the Basis of VLSS and VHLSS Data

Attention has focused on the urban/rural gap in Viet Nam because it is felt by a number of researchers (see notably Nguyen Binh *et al.* 2007) that any increase - albeit modest - in inequality is probably due to a possible increase in that gap. Table 1 supports this point of view; the estimated urban-rural gap in the mean logarithm of real per capita expenditure has indeed increased, fairly sharply between 1993 and 1998, and slightly between 1998 and 2002.

Table 1: Urban-rural gap and Gini coefficients

Year	Est. urban-rural gap in mean log real per capita exp.	Gini coefficient
1993	0.56	Whole country 0.330 Urban 0.340 Rural 0.278
1998	0.74	Whole country 0.354 Urban 0.348 Rural 0.275
2002	0.80	Whole country 0.380 Urban 0.356 Rural 0.290

Source: Nguyen Binh *et al.* 2007, Haughton *et al.* 2001 for 1993 and 1998, authors' computations for 2002.

Moreover, the Gini coefficient (computed for real per capita expenditure over the sample of households for each year, taking sampling weights into account) for the whole country has increased moderately, while increases are quite a bit smaller when considering Gini coefficients for rural and urban areas separately.

It is also interesting to examine the distribution of the logarithm of real per capita expenditures, displayed in Figure 1 as a plot of the kernel density estimate, computed over rural and urban areas separately.

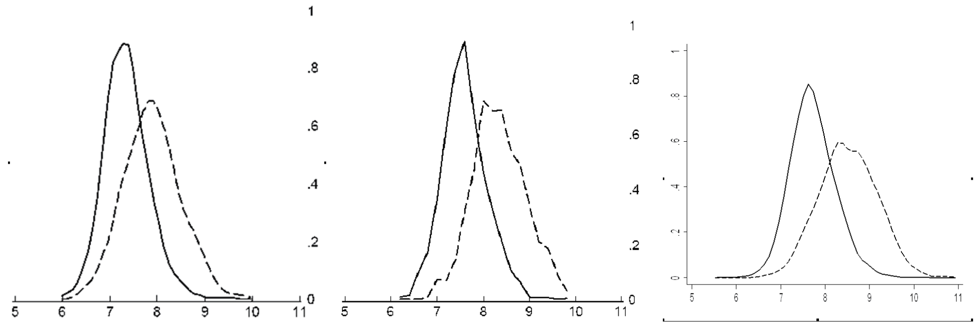


Figure 1: Kernel Density of log of real per capita expenditure in 1993 and 1998 (weighted), reproduced from Figure 1 of Nguyen Binh *et al.* and in 2002, from authors' computations.

Figure 1 suggests an increase in the horizontal shift between the urban and rural curves over the three years 1993, 1998 and 2002; moreover, it would seem (this is visible on the 2002 graph) that the urban-rural shift tends to increase with the level of wealth (as one moves from left to right on the graph), a tendency that had already been identified in Nguyen Binh *et al.* (2007).

A natural question arises of where in the country the urban-rural gap might be most pronounced, whether this gap can be attributed to known attributes of a particular part of the country, or whether in fact there might be effects to the gap due to a particular location but not attributable to known characteristics. To tease out these issues, we propose to use a multilevel regression model, to be described below.

4. Multilevel Model Approach to the Study of Inequality

The multilevel model methodology, also referred to as mixed models methodology, essentially involves generalized linear models, most commonly linear regression models, but with the introduction of suitable random effects both in intercepts of the models and in the coefficients, to allow for the fact that variables are measured at various levels; for example in our case a household is a member of a commune, itself a member of a district, itself a member of a province.

The technique has been known for some time, particularly when referred to as the method of mixed models, but it has become popular in large part because of important work related to educational studies in the U.K. which recognized the importance of properly taking into account the various levels in the data (pupils as members of classes as members of schools, for example) often ignored in practice when building regression models. This gave rise to a number of useful software packages such MLWIN (Multi Level for Windows), HLM (Hierarchical Linear Models) or GLLAMM (Generalized Linear Latent and Mixed Models, a suite of commands for Stata), and the methodology has now become widespread in a number of areas, including management, where employees are seen as members of departments, in turn members of firms, in turn member of industries etc.

In the area of living standards, we mention recent work (2007) by Arpino who investigated models of expenditure per capita from VLSS 1993 and 1998 and used the panel which exists between these two surveys to build a model of exit from poverty; Arpino obtained commune level random effects, and argued, as we do here, that these effects can be seen as location-specific (positive or negative) contributions to wealth, when other variables are controlled for. Recent work by Bono *et al.* (2007) used a multilevel approach to study regional inequalities in consumption patterns in Italy.

Our contributions in this paper are two-fold. First we propose a multilevel model to study geographical variations in the urban/rural gap when other variables are controlled for, by proposing random effects in the coefficient of the dummy variable (1/0) on the urban location of the household. Secondly, following work by Moura (1994, 1999), we demonstrate briefly how our model can be used to perform small area estimation at the commune level, where direct estimation would be too imprecise because of insufficient sample sizes. Further development of our models for the purpose of small area estimation is the object of another paper, but we thought it worthwhile to mention this application of our model, at least briefly, in this paper.

Among the possible tools available to build multilevel models, we have elected to use MLWIN, because of its Bayesian capabilities which will be necessary when our models are extended to models involving for instance more than one year of data at the same time (such future perspectives are mentioned in the conclusion to this paper). The algorithm we used to build the model is a fairly standard algorithm, the Iterated Generalized Least Squares (IGLS) algorithm, the details of which are described in a number of standard references on multilevel modeling, available for example from the Center for Multilevel Modeling at the University of Bristol[1].

Our multilevel model for (the logarithm of) real expenditure per capita (lrpc-exp) can then be formulated in general form as follows in equations (4.1), (4.2), (4.3)and (4.4):

$$Y_{ijkl} = \beta_{0jkl} + \sum_{p} \beta_{pjkl} X_{pijkl} + \epsilon_{ijkl} \tag{4.1}$$

$$\beta_{pijkl} = \gamma_{00} + f_{0l} + \nu_{0kl} + \nu_{0jkl} \tag{4.2}$$

$$\beta_{pijkl} = \gamma_{p0} \quad \text{for } p \neq 5 \tag{4.3}$$

$$\beta_{6jkl} = \gamma_{60} + f_{6l} + \nu_{6kl} \tag{4.4}$$

[1]See http://www.cmm.bristol.ac.uk

where the Y_{ijkl} represent the values of the dependent variable at the first level (household level). This is in our case the logarithm of the real per capita expenditure of the ith household ($i = 1, \ldots, n_j$), level 1) in the jth commune ($j = 1, \ldots, m_k$), level 2) of the kth district ($k = 1, \ldots, r_l$), level 3) of the lth province ($l = 1, \ldots, 61$, level 4); X_{pijkl} represent the value of the pth explanatory variable measured at the first level (household level) of the ith household in the jth commune of the kth district of the lth province; here $p = 1, \ldots, 14$ corresponding to the 14 variables listed in Table 4.

Note that in VHLSS 2002, the value of n_j, in principle equal to 25 for all communes, in fact varies: 759 communes had more than 5 households (17-25) in the sample, and 2142 communes had 3-5 households in the sample).

All random residuals are assumed to be independent and normally distributed with mean zero and a variance which is constant within a level.

The estimated coefficients are displayed with standard errors in Table 2, as well as the estimated variances and their standard errors of all random effects; all coefficients and variances are significant. For example, for the variable "married" the estimated value of the coefficient is 0.056 with its standard error of 0.007, which yields a significant coefficient.

Table 2: Multilevel model: Estimated Coefficients

Fixed Coefficients			Random Coefficients		
	Estimated values	Standard errors		Estimated vlaues	Standard errors
γ_{00}	7.170	0.036	Variance of f_{0l}	0.033	0.007
β_1	0.056	0.007	Variance of ν_{0kl}	0.020	0.002
β_2	0.022	0.001	Variance of ν_{0kjl}	0.017	0.001
β_3	-0.039	0.007	Variance of f_{6l}	0.012	0.004
β_4	0.164	0.012	Variance of ν_{6kl}	0.017	0.005
β_5	-0.081	0.001			
γ_{00}	0.230	0.022			
β_7	0.033	0.001			
β_8	0.254	0.015			
β_9	0.266	0.019			
β_{10}	0.181	0.015			
β_{11}	0.137	0.007			
β_{12}	0.048	0.007			
β_{13}	-0.0000154	0.000009			
β_{13}	0.004	0.001			

The intercept includes both fixed and random effects, the latter at province, district and commune levels. The household level corresponds to a household-specific random error ϵ_{ijkl}.

The coefficient of the urban variable also includes both fixed and random effects, the latter at province and district levels (communes are either fully urban or fully rural, so it is not possible to include a commune level effect in the slope of the urban dummy variable).

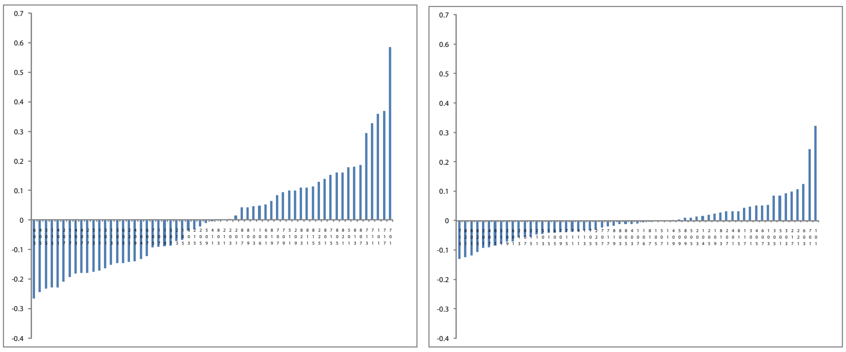
It follows from equation (4.4) that the province-level random effect f_{6l} in the urban coefficient, which is estimated for each province (all provinces are represented in the survey) provides a measure of the contribution to the urban/rural gap due to a location in this particular province, when other characteristics in the model (such as education, etc) are controlled for. In the same way the district-level random effect ν_{6kl} in the urban coefficient represents the contribution to the urban/rural gap due to a location in this particular district, and not due to characteristics such as education etc. In other words, these random effects provide a location specific random contribution to the urban/rural gap, which is likely to be accounting for variables which are not included in the model and may not be available.

By adding the two random effects in the coefficient of urban, that due to the province and that due to the district, one obtains for each district a total random effect, higher levels of which indicate districts with higher location-specific urban/rural gaps.

The same analysis applies to random effects in the intercept; by adding all three random effects, due to the province, the district and the commune, we obtain a location-specific contribution to the overall level of wealth in the commune which is not due to characteristics such as education, etc.

The province-level random effects in the intercept are displayed in increasing order in Figure 2a and the province-level random effects in the coefficient of the urban variable are similarly displayed in Figure 2b, with the province codes on the horizontal axis in both figures (a list of provinces and corresponding codes is provided in the Appendix). One can see (Figure 2a) that Hanoi (code 101) and Ho Chi Minh City (code 701), which are considered as provinces administratively, as well as Ba Ria Vung Tau (code 717) display the highest location-specific random contribution to overall wealth as one might expect (the latter, Ba Ria Vung Tau, because of the presence of off-shore oil and tourism). It appears in Figure 2b that Hanoi and Ho Chi Minh City also display the highest location-specific random contributions to the urban/rural gap, which makes intuitive sense when one remembers the presence of a number of very poor rural communes in the Hanoi area, for example.

It should be noted that our model was built for at least two reasons, one to investigate the urban/rural gap, and the other to investigate the possibility of obtaining small area estimates of the mean (logarithm of) expenditure per capita at the commune level. This latter goal follows a methodology proposed by Moura in his 1994 thesis and subsequent publications. While a full discussion of small area estimation is outside the scope of this paper, one can mention that our model can be used to obtain small area estimates at commune level with attractive properties, provided that the independent variables used in the model are available for the whole population, as for example from a census.



(a) (b)

Figure 2: (a) Intercept province random effects, in increasing order; (b) Urban/rural province random effects, in increasing order.

This is the reason why the independent variables in our model are restricted to those available in the Viet Nam 1999 census. Given this availability, one simply needs to apply our model equation (estimated equation (4.1)) and replace the independent variables by their commune population means to get small area estimates for each commune covered in the survey. The ten to fifteen years-old methodology referred to as poverty mapping, still currently widely used, follows the same idea but without the inclusion of random effects (see for example Baulch and Minot (2002) for an example of such an application, using a subset (3%) of the 1999 Census, and Swinkels and Turk 2007). Moura's work, among others', indicates that the use of random effects improves the properties of small area estimators, probably because it does make it possible to capture the effect of variables necessarily excluded from the model by the fact that one can only use variables available for the whole population as independent variables.

We present in Figure 3 box plots by province of the small area estimates at the commune level of the mean (logarithm of) real per capita expenditure obtained from this procedure, arranged in increasing order of the median in each box plot. Note that the identification of outlying communes (relative to the level of wealth in their province) can be quite useful for purposes of targeting.

Figures 4 and 5 display box plots by province of district random effects in the intercept (Figure 4) and in the urban/rural gap (Figure 5), in other words in the coefficient for the urban dummy variable. Finally Figure 6 displays box plots by province of the commune random effects which are part of the intercept. In Figures 4-6, as in Figure 3, the box plots are arranged in increasing order of the median in each boxplot.

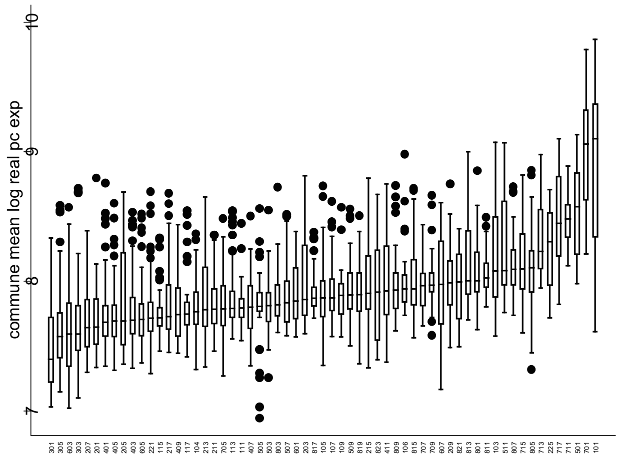


Figure 3: Small area estimates for commune level mean (log of) real per capita expenditure

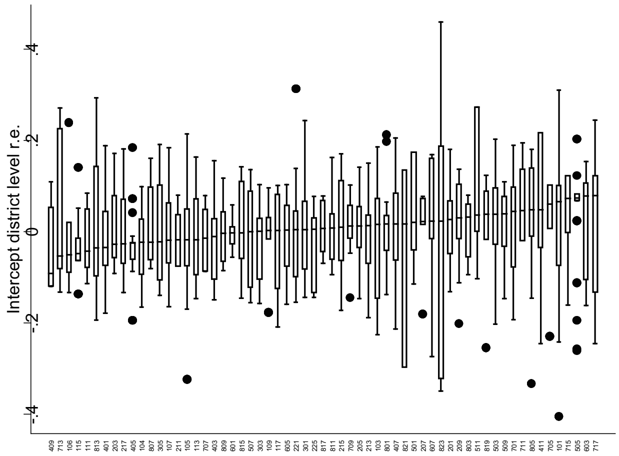


Figure 4: Box plots by province of intercept district-level random effects

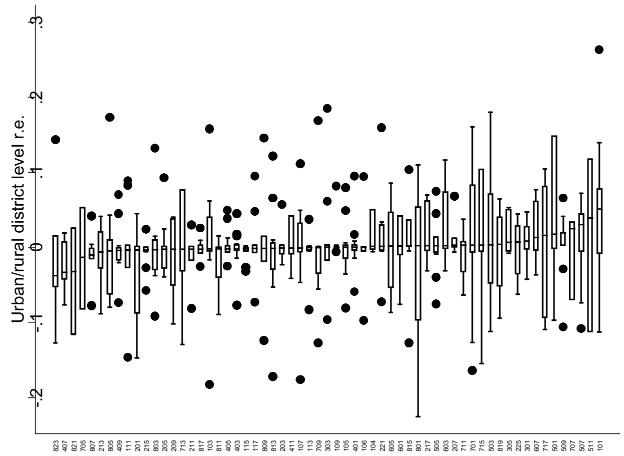


Figure 5: Box plots by province of urban/rural district-level random effects

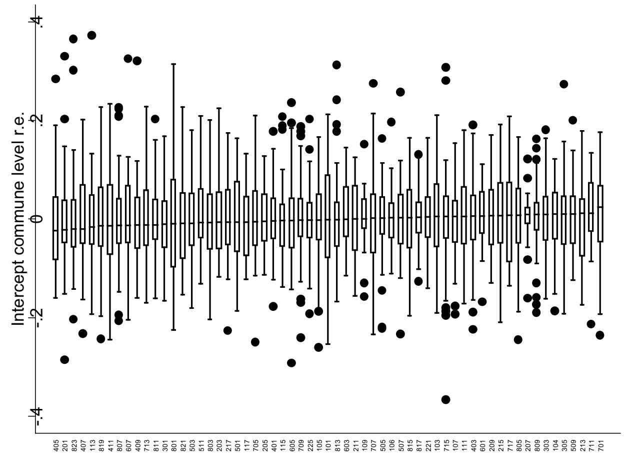


Figure 6: Box plots by province of intercept commune-level random effects

5. Conclusions and Future Perspectives

This paper has proposed to study inequality in Viet Nam by focusing on the urban/rural gap, and using a multilevel model to help identify location-specific random contributions to this gap, above and beyond any effects of known characteristics of these locations (such as education levels of the population, the structure of the labor force, ethnic composition etc). At the same time, we have demonstrated that our model can be used to obtain small area estimates at the commune level.

The small area estimation thread in our work can and will be expanded. For instance, we expect to be building separate models for urban and rural areas, so that we can use commune-level data available in the population only for rural areas. We also expect to use our models to obtain small area estimates even for those communes not included in the survey, by averaging random effects in near-by communes which were surveyed.

Another interesting future perspective involves multilevel models over several years of data, using 2002, 2004 and 2006 VHLSSs. Because of the presence of partial panels across these years, the opportunity arises to create multilevel model with cross-classification and multiple membership (see Browne *et al.* 2001), affording the possibility of comparing time variations with geographical variations. These models will also provide further understanding of the evolution of the urban/rural gap over time, and across geographical areas.

Finally, in models such as we have considered, it is quite likely that spatial effects play a role; this implies that living standards in a commune (for example) are associated with characteristics of that commune and random effects arising from that commune, but also with effects arising from neighboring communes. Such effects can be investigated with multiple membership multilevel models and will be the object of future work.

Appendix: List of Provinces and Codes

101 TP. Ha Noi	211 Tuyen Quang	501 TP.Da Nang	713 Dong Nai
103 TP. Hai Phong	213 Yen Bai	503 Quang Nam	715 Binh Thuan
104 Vinh Phuc	215 Thai Nguyen	505 Quang Ngai	717 Ba Ria Vung Tau
105 Ha Tay	217 Phu Tho	507 Binh Dinh	801 Long An
106 Bac Ninh	221 Bac Giang	509 Phu Yen	803 Dong Thap
107 Hai Duong	225 Quang Ninh	511 Khanh Hoa	805 An Giang
109 Hung Yen	301 Lai Chau	601 Kon Tum	807 Tien Giang
111 Ha Nam	303 Son La	603 Gia Lai	809 Vinh Long
113 Nam Dinh	305 Hoa Binh	605 Dak Lak	811 Ben Tre
115 Thai Binh	401 Thanh Hoa	607 Lam Dong	813 Kien Giang
117 Ninh Binh	403 Nghe An	701 TP. Ho Chi Minh	815 TP. Can Tho
201 Ha Giang	405 Ha Tinh	705 Ninh Thuan	817 Tra Vinh
203 Cao Bang	407 Quang Binh	707 Binh Phuoc	819 Soc Trang
205 Lao Cai	409 Quang Tri	709 Tay Ninh	821 Bac Lieu
207 Bac Kan	411 Thua Thien – Hue	711 Binh Duong	823 Ca Mau
209 Lang Son			

References

Arpino, B. (2007). Dynamic multi-level analysis of households' living standards and poverty: evidence from Vietnam. ISER Working Paper 2007-10, Institute for Social and Economic Analysis, University of Essex, Essex, United Kingdom.

Baulch, B. and Minot, N. (2002). The spatial distribution of poverty in Vietnam and the potential for targeting. World Bank working paper 2829.

Bono, F., Cuffaro, M. and Giaimo, R. (2007). Regional inequalities in consumption Patterns: a multilevel approach to the case of Italy. *International Statistical Review* **75**, 44-57.

Browne, W. J., Goldstein, H. and Rasbash, J. (2001). Multiple membership multiple classification (MMMC) Models. *Statistical Methodology* **1**, 103-124.

Démurger, S., Fournier, M. and Li, S. (2006). Urban income inequality in China revisited (1988-2002). *Economics Letters* **93**, 254-359.

General Statistical Office. (1999). *Vietnam Living Standards Surveys.*

Hannum, E. and Wang, W. (2006). Geography and educational inequality in China. *China Economic Review* **17**, 253-265.

Haughton, D., Haughton, J. and Nguyen P., editors. (2001). *Living Standards During an Economic Boom; the Case of Vietnam.* United Nations Development Program and General Statistical Office, Statistical Publishing House, Hanoi. Available at http://www.undp.org.vn/undp/docs/2001/living/lse.pdf.

Huong, D. L., Minh, H. V., Janlert, U., Van, D. D. and Byass, P. (2006). Socio-economic status inequality and major causes of death in adults: a five-year follow-up study in Vietnam. *Public Health* **120**, 497-504.

Li, H. and Zhu, Y. (2006). Income, income inequality and health: evidence from China. *Journal of Comparative Economics* **34**, 668-693.

Liu, A. Y. C. (2006). Changing wage structure and education in Vietnam, 1992-98. *Economics of Transition* **14**, 681-706.

Ma, A. (2006). Geographical location of foreign direct investment and wage inequality in China. *The World Economy* **29**, 1031-1054.

Moura, F. A. S. (1994). Small Area Estimation Using Multilevel Models. Ph.D. Thesis, University of Southampton, U.K.

Moura, F. A. S. and Holt, D. (1999). Small area estimation using multilevel models. *Survey Methodology* **25**, 73-80.

Nguyen, B. T., Albrecht, J. W., Vroman, S. and Westbrook, M. D. (2007). A quantile regression decomposition of urban-rural inequality in Vietnam. *Journal of Development Economics* **83**, 466-490.

Nguyen, P. and Haughton, D. (2006). An overview of socio-economic surveys in Vietnam: methods, achievements and challenges. *Proceeding of Statistics Canada Symposium 2005, Methodological Challenges for Future Information Needs.*

Pham, T. H. and B. Reilly, B. (2006). The gender gap in Vietnam, 1993-2002: A quantile regression approach. Poverty Research Unit at Sussex Working Paper 34.

Ravallion, M. and Chen, S. (2007). China's (uneven) progress against poverty. *Journal of Development Economics* **82**, 1-42.

Swinkels, R. and Turk, C. (2007). Poverty mapping in Vietnam, chapter 14, In *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies* (Edited by Bedi, T., Coudouel, A. and Simler, K., chapter 14). The World Bank.

Vietnam: State Planning Committee and General Statistical Office (1994). *Vietnam Living Standards Survey 1992-1993.* Hanoi: Statistical Publishing House.

Vietnam: General Statistical Office (2000). *Vietnam Living Standards Survey 1997-98.* Hanoi: Statistical Publishing House.

Wan, G., Lu, M. and Chen, Z. (2006). The inequality-growth nexus in the short and long run: empirical evidence from China. *Journal of Comparative Economics* **34**, 654-667.

World Bank. Living Standards Survey Measurement Program. http://www.worldbank.org/lsms/guide/select.html.

Zhang, Y. and Wan, G. (2006). The impact of growth and inequality on rural poverty in China. *Journal of Comparative Economics* **34**, 694-712.

Zhao, Z. (2006). Income inequality, unequal health-care access, and mortality in China. *Population and Development Review* **32**, 461-483.

footnotesize

Received March 4, 2008; accepted September 5, 2008.

Dominique Haughton
Bentley College
175 Forest Street
Waltham, MA 02452, USA
dhaughton@bentley.edu

Also:
Toulouse School of Economics
Universite' Toulouse I
Manufacture des Tabacs
Alle'es de Brienne
31000 Toulouse
France

Phong Nguyen
General Statistics Office, Viet Nam
Hoang Van Thu, Hanoi, Viet Nam
nphong@gso.gov.vn