Session Number: Parallel Session 4d
Time: Tuesday, August 24, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Estimating the Stability of Poverty Analysis:
Out-of-Sample Prediction in Dynamic Poverty Mapping

Melanie Grosse
Boris Branisa

For additional information please contact:

Name: Melanie Grosse
Affiliation: University of Gottingen

Email addresses: (Branisa) bbranis@uni-goettingen.de, (Grosse) mgrosse@uni-goettingen.de

Estimating the Stability of Poverty Analysis: Out-of-Sample Prediction in Dynamic Poverty Mapping

Boris Branisa and Melanie Grosse *

This version: July 12, 2010

Abstract

In Bolivia, as in many other developing countries, a sufficiently long time series of nationally representative income surveys does not exist which makes it difficult to analyze trends and determinants of poverty and inequality over a longer time period. However, in many countries, there are urban household surveys, and there are nationally representative Demographic and Health Surveys (DHS) which lack information on incomes. For the case of Bolivia, we have two urban household surveys and four nationally representative DHS available since 1989, while comparable nationally representative household income surveys only exist since 1999. In this paper, we modify a technique developed for (static) poverty mapping exercises by combining urban household income surveys with DHS data to (dynamically) extend the time series of household income data back in time until 1989 and 1994, starting from the base period (1998/9). Our technique explicitly estimates the robustness of this backward extension by repeating it for a second base periods with two sets of nationally representative household surveys and DHS (2002/3). Furthermore, we use and compare two different methods of modeling dynamics. In doing so, we are able to gain insights about the stability and reliability of dynamic poverty mapping analysis.

***JEL* Classification:** C81, D31, I31, I32, O54.

Key words: Microsimulation, survey matching, poverty, inequality, poverty mapping, Bolivia

*University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, email: bbranis@uni-goettingen.de; mgrosse@uni-goettingen.de. We would like to thank Stefan Sperlich and Stephan Klasen for detailed comments and discussion.

1 Introduction

In the mid-1980s, Bolivia experienced a dramatic macroeconomic crisis, with an annual inflation rate that peaked at 25.000 percent and great social unrest (Sachs and Morales, 1988). According to some authors, the crisis was a consequence of the government's growing fiscal deficit and of the public companies that had been financed with external debt during the 1970s (Morales, 1994). This funding was used to support a model where the government played a major role in the economy. The debt crisis in Latin America made it very difficult for Bolivia to obtain new external funding, and the government was forced to finance the fiscal deficit with monetary emission. As the government was unable to solve the crisis, President Siles Suazo resigned and called for general elections. In 1985, the newly elected government of President Paz Estenssoro started a stabilization program which included a tax reform, a sharp reduction in government spending, and liberalization of the economy. The program was successful to combat hyperinflation, and was the beginning of a reform process that continued during the 1990s and led to a more market-oriented economy in Bolivia (Morales, 2001).[1] The impact of these reforms on the levels of income poverty and inequality is, however, subject of debate. As nationally representative household income surveys have been only conducted since 1999, the trends of income poverty and inequality in Bolivia since the late 1980s are still an open empirical question.

In Bolivia, as in many other developing countries, a sufficiently long time series of nationally representative income surveys does not exist leading to difficulties analyzing the determinants and trends of poverty and inequality over a longer time period. A method used in the literature to overcome this difficulty is modifying or extending poverty mapping models. The basic idea of poverty mapping is to use one specific micro-data survey which contains all relevant information for poverty analysis, and to combine this information with that of another survey which typically contains only part of the necessary information. The classical application is to use a household survey such as a Living Standard Measurement Survey (LSMS)[2] which contains the relevant information on income and income determinants for a representative subsample of the population and to apply this information structure to a Census that does not contain all parts of the information,

[1] The market-oriented approach continued until the year 2005 when President Evo Morales was elected and initiated the return to government-led development.

[2] For simplicity, we call all kinds of household income surveys LSMS even if they belong to a different kind of income survey family.

i.e., missing information on income, but providing some other information for the whole population of a country, such as asset ownership, education, or demographics. The logic is to establish a statistical correlation structure between various covariates and income, for example with an OLS regression, in the first survey (LSMS) that can be applied to the second survey (Census) to predict incomes. Proposed by Hentschel, Lanjouw, Lanjouw, and Poggi (2000) and Elbers, Lanjouw, and Lanjouw (2003) for Ecuador, the method has been applied to many countries and different kinds of surveys. Applying this method makes it possible to calculate detailed poverty maps, e.g., at the municipality level, for example, to target public spending policies in the most effective way.

Besides the geographic use of this method, there have also been attempts to apply poverty mapping back and forth in time. Klasen, Grosse, Thiele, Lay, Spatz, and Wiebelt (2007), Grosse, Klasen, and Spatz (2009), and Stifel and Christiaensen (2007) have used LSMS and Demographic and Health Surveys (DHS) for their temporal analysis. The difference between the attempts basically lies the assumption about dynamics when modeling the correlation structure. Whereas Stifel and Christiaensen (2007) assume no dynamics in the regression model, the other studies propose several ways of modeling dynamics.

Explicitly judging or verifying the results whether the estimation is correct (with other data) is hardly possible. In fact, poverty mapping was invented *because* no other data is available, so there is generally no possible way to check the predictive power of these models. One exemption is Mathiassen (2008) who uses several LSMS for Uganda and explicitly tests how well a regression model of one point in time can predict incomes for other points in time. The author has all necessary information at hand to explicitly test the predictive power of poverty maps using static coefficients over time. Demombynes, Elbers, Lanjouw, Lanjouw, Mistiaen, and Özler (2004) also show how well their Census predictions coincide with the LSMS estimates for three countries.

This paper draws heavily from the previous study on poverty mapping in Bolivia by Grosse et al. (2009). Since national household income surveys exist only since 1999 and only urban surveys for earlier years, analysis would leave more than half of the population living in non-urban areas uncovered by data. There are several nationally representative DHS which, however, lack information on incomes. In this paper, we use four nationally representative DHS (1989, 1994, 1998, and 2003),[3] two nationally representative household

[3]As already explained in Grosse et al. (2009), the DHS are nationally representative for households with at least one reproductive woman, i.e., aged between 15 and 49. The effect that this sample selection has on poverty and inequality is also shown in Grosse et al. (2009).

income surveys (1999 and 2002), and two urban household surveys (1989 and 1994). We modify the technique developed for (static) poverty mapping exercises of Hentschel et al. (2000) and Elbers et al. (2003) by combining urban household income surveys with DHS data to (dynamically) extend the time series of household income data back in time until 1989 and 1994 following Grosse et al. (2009). Our technique explicitly estimates the robustness of this backward extension by repeating it for two base periods with two sets of nationally representative data of LSMS and DHS data (1998/9 and 2002/3). We especially take a look at the robustness of the estimations by focussing on two years for which we have complete data for both LSMS and DHS and perform out-of-sample predictions for poverty and inequality assuming that the LSMS of 1999 had also been only urban using 2002 as base year and vice versa. Furthermore, we compare the results of the assumptions of Grosse et al. (2009) with, for example, the ones used by Stifel and Christiaensen (2007). In doing so, we are able to gain insights in the reliability of poverty mapping analysis over time.

2 Approach and Data

The basic idea of this paper applies the approach of Hentschel et al. (2000) and Elbers et al. (2003). The authors use Ecuadorian LSMS data which contains information on income and income determinants and extend their poverty analysis spatially to the whole country using Census data. To be able to do so, they run an expenditure model in the LSMS data using only covariates that are also available in the Census data. Simulating income in the Census data is achieved by simply multiplying the covariates of the Census with the regression coefficients obtained from the LSMS survey (plus adding an error term). With this simulated data, they are able to generate detailed poverty maps of national coverage at the municipality level.[4]

Only very few papers explicitly test how well it works to simulate incomes using the described cross-survey matching. Demombynes et al. (2004) replicate the exercise of Hentschel et al. (2000) and Elbers et al. (2003) for Ecuador and two more countries, Madagascar and South Africa, and compare their results for the strata where both observed expenditures from LSMS data and simulated expenditures from Census are available. For

[4]The second main contribution of Elbers et al. (2003) was to investigate how to correctly estimate standard errors for the estimates by splitting the error term into a spatial and an idiosyncratic component. We cannot do this since the primary sample units (or clusters) of the pre–1999 LSMS are not available for the Bolivian data sets.

most strata, but not for all, they find that the observed poverty levels are similar in LSMS and Census data.[5] Stifel and Christiaensen (2007) use the same methodology, but apply the coefficients obtained from Kenyan LSMS data (of 1997) to several DHS surveys (instead of Census), and they do so back and forth in time (1993 and 2003) instead of across space. In doing so, they assume that the returns to covariates remain constant over time. For the DHS of 1998, which is closest in time, they find a persistent underestimation of poverty when applying the regression coefficients of 1997 to the 1998 data. They correct for it by shifting the poverty line until the simulated headcount in the DHS matches the observed one of the LSMS.

In a study of Mathiassen (2008), the author uses a series of Ugandan LSMS data sets and tests the predictive power of models from one LSMS survey for the other surveys and compares observed poverty with simulated poverty for all years and all surveys. She finds that for 2 of the 7 surveys, the simulations are working very badly. She assumes the reason to be either an unexpected large change of poverty or the time lag between the surveys. In addition, the adequacy of applying the models is much worse for rural areas where nearly half of the simulated poverty levels are statistically different from observed poverty levels.

We draw strongly on previous work by Grosse et al. (2009). It is similar to Stifel and Christiaensen (2007), in that we also use LSMS and DHS data. Differently to Stifel and Christiaensen (2007), we explicitly model dynamics in changes in the regression coefficients instead of assuming that the coefficients stay constant over time. We test the ways of taking dynamics into account (or not) and are able to check which model comes closer to observed values. This is only possible in the case that the two full sets of nationally representative surveys are available. In this respect, this paper is more similar to Mathiassen (2008). With both surveys available for both years, we are able to estimate the stability of poverty analysis by out-of-sample predictions in the dynamic poverty mapping approach. We start presenting the model that follows Grosse et al. (2009).

We construct a 3×3 block diagonal structure of the covariates by interacting them with three regional dummies, and run a weighted standard log-linear OLS regression model where the indices C, T, and R stand for cities, towns, and rural areas, respectively, β are

[5] For Ecuador, 2 out of 8 strata show different results for P0. For Madagascar, all estimates are inside each others confidence intervals, however, due to very high standard errors causing a level difference for point estimates of up to 13 percentage points. The same holds for South Africa with differences up to 6 percentage points.

coefficient vectors, and ε is an independent error term:

$$\begin{pmatrix} \ln y_t^C \\ \ln y_t^T \\ \ln y_t^R \end{pmatrix} = \begin{pmatrix} X_t^C & 0 & 0 \\ 0 & X_t^T & 0 \\ 0 & 0 & X_t^R \end{pmatrix} \begin{pmatrix} \beta_t^C \\ \beta_t^T \\ \beta_t^R \end{pmatrix} + \begin{pmatrix} \varepsilon_t^C \\ \varepsilon_t^T \\ \varepsilon_t^R \end{pmatrix}. \tag{1}$$

Concerning the modeling of dynamics, we test the following assumptions proposed in the literature. The baseline assumption for earlier periods $t-1$, in which the LSMS covers only the cities, is that the absolute differences[6] in the regression coefficients between cities and non-urban areas remain constant between period $t-1$ and t:

$$\beta_{t-1}^T = \beta_{t-1}^C + (\beta_t^T - \beta_t^C) \quad \text{and} \quad \beta_{t-1}^R = \beta_{t-1}^C + (\beta_t^R - \beta_t^C). \tag{2}$$

The second proposed way of capturing "dynamics" in the literature is to assume that there are none, as done in Stifel and Christiaensen (2007)[7] and tested in Mathiassen (2008):[8]

$$\beta_{t-1}^C = \beta_t^C \quad \text{and} \quad \beta_{t-1}^T = \beta_t^T \quad \text{and} \quad \beta_{t-1}^R = \beta_t^R. \tag{3}$$

As the predicted income is obtained from a regression, its variance is too small as compared to observed income. To compensate for this the simulation is run 200 times, where a random variable assumed to be normally distributed with mean zero and with the estimated variance of the error term is added each time to the predicted values. For computing the estimated variance of the error term, we follow Grosse et al. (2009) and assume the constancy of relative changes for the dynamic case

$$\sigma(\varepsilon_{t-1}^T) = \sigma(\varepsilon_{t-1}^C) \cdot \frac{\sigma(\varepsilon_t^T)}{\sigma(\varepsilon_t^C)} \quad \text{and} \quad \sigma(\varepsilon_{t-1}^R) = \sigma(\varepsilon_{t-1}^C) \cdot \frac{\sigma(\varepsilon_t^R)}{\sigma(\varepsilon_t^C)}, \tag{4}$$

and for the no-dynamics case we assume no changes over time:

$$\sigma(\varepsilon_{t-1}^C) = \sigma(\varepsilon_t^C) \quad \text{and} \quad \sigma(\varepsilon_{t-1}^T) = \sigma(\varepsilon_t^T) \quad \text{and} \quad \sigma(\varepsilon_{t-1}^R) = \sigma(\varepsilon_t^R). \tag{5}$$

Different from the studies applying the poverty mapping approach cited above, there is a whole literature addressing the question of changes over time in changing endowments and changing coefficients, as well as changing unobservables. In a multi-country study,

[6]Note that we use the term "absolute" not in the mathematical meaning of $|-1| = 1$, but to contrast it to "relative", i.e., percentage changes.

[7]It must be noted that Stifel and Christiaensen (2007) use in their study assets whose parameters are expected to remain stable over time, and that unlike the case in our paper, Stifel and Christiaensen (2007) use different regressions for the regions they consider. We are forced to use the same regressors for all regions to be able to calculate Equation (2).

[8]Mathiassen (2008) suggests to "update" coefficients in order to take dynamics or "outlier years" into account by averaging coefficients over different years.

edited by Bourguignon, Ferreira, and Lustig (2005), the authors investigate this question of changes and the resulting impacts on inequality (and partly also on poverty and income) for 7 countries.[9] The authors apply generalized Oaxaca-Blinder decomposition methods to investigate how different groups (such as the urban versus rural population) are affected by changes in the distribution of endowments (called population or endowment effect), changes in the returns to these endowments (called price effects), and changes in decision how to use the endowments such as behavior on the labor market (called occupational effects). They point out that these factors are not independent from each other, and that, in addition, they are likely to be affected by external shocks (e.g., international prices) or internal shocks (e.g., government policies). Furthermore, they highlight that both shocks as well as changes are likely to be different for different subgroups (such as the urban-rural divide) of the population.[10]

Only three studies of Bourguignon et al. (2005) address the specific questions on urban versus rural trends. The chapter on Colombia suggests that there are substantial differences between urban and rural areas. For example, the first time period of observation showed increasing inequality in rural areas and stagnating in urban areas. The second period showed exactly the opposite pattern. Especially the effect of increased schooling differed: More and more equally distributed years of schooling in urban areas caused higher inequality, whereas more years of schooling in rural areas cause an inequality decline, due to lower marginal returns in rural areas. Indonesia showed a massive increase in income and a massive reduction in poverty, combined with a increase in inequality. Concerning differences in price effects, returns to education went down in urban areas and caused decreasing inequality, whereas they went up in rural areas and caused increasing inequality. Also other regional differences played a role (such as living on Java or elsewhere). Indonesia experienced in addition a massive urbanization, causing increasing self-employment in urban areas: There was selective migration of the mainly landless poorer wage-workers from rural to urban areas which was in turn profitable for the migrants. For Mexico, the authors find growing negative returns to living in rural areas. Similar to Indonesia, there was a strong rural-to-urban migration as well as from poorer-to-richer-regions migration which lead to increasing inequality. In Mexico, also urban-rural differentials in education

[9] Argentina, Brazil, Colombia, Indonesia, Malaysia, Mexico, Taiwan.

[10] The model which is used similarly to all studies is very different from our model, which is driven by the data they use. The DHS which we use have hardly any employment data except for the few variables listed in Appendix Table **??** so that their approach goes beyond of what we could do with our data.

increased.[11]

In our study, apart from testing the assumptions on different dynamics for urban versus rural areas, we change the original estimation procedure of Grosse et al. (2009) in some ways. As a first test for stability of income, poverty, and inequality estimates, we rerun the Grosse et al. (2009) regression using only covariates that are available in all four DHS and LSMS surveys. To reduce the number of covariates in a more meaningful way, we first statistically test for the equality of means of covariates X of the LSMS of 1999 and the DHS of 1998. This improves the original method since the equality of means is desirable for being able to apply the β coefficients of one survey to the data of the other survey (Stifel and Christiaensen, 2007). Second, we continue to reduce the number of regressors to avoid a large number of insignificant coefficients by redefining covariates, whose representation included disaggregation into many dummy categories, to a simpler categorization,[12] see subsection 3.1 for details.

The data set we use consists of four LSMS: the 2$^{\text{nd}}$ round (1989) and the 7$^{\text{th}}$ round (1994) of the Encuesta Integrada de Hogares (EIH), both only covering urban areas, and the 1$^{\text{st}}$ round (1999) and the 4$^{\text{th}}$ round (2002) of the Encuesta Continua de Hogares (ECH), both being nationally representative. The purpose of the LSMS is to collect individual, household, and community level data to measure the welfare level of the sampled population. In addition to income and/or expenditure data, the LSMS provide information on demographics, asset ownership, education, employment, and health.[13]

Our set of DHS consists of the first four Bolivian rounds which were conducted in 1989, 1994, 1998, and 2003. Two-stage sampling techniques were used to select nationally representative samples of women aged between 15 and 49 who serve as eligible respondents of the DHS, i.e., women in reproductive age. The main objective of the DHS is to collect demographic data on health and fertility trends. Additionally, it includes some questions

[11] The study from Grimm (2004) for Cote d'Ivoire also takes urban-rural differentials into account and separates the model into three different models, for urban men, rural men, and women. Also here, price effects can go into different directions, can be of different sign, and can be of different magnitude for the three models. Grimm (2004) points out the relevance of external and internal factors such as rising world market prices for the main cash crops, freezing of public wages, devaluation of the exchange rate, adjustment policies. Overall, the author finds an urbanization of poverty (combined with decreasing between-region and increasing within-region inequality.).

[12] With the remaining coefficients, we believe to have found a model that is stable. We opted against running stepwise regressions or an alternative data-driven approach as we want to choose the variables to be included based on theory and findings from the literature. It must be noted that some of the variables where equality of means is not given are nevertheless included as dummies that represent one meaningful category that cannot be left out (e.g., some of the departments or education categories).

[13] Note that our monetary variable is mixed: We use income for cities and towns and expenditures for rural areas, see Grosse et al. (2009) for the details.

on the educational attainment and the employment situation of adults and on the asset ownership of the household. They can be grouped into five categories: information on demographics of the household, asset ownership of the household, educational attainment of adult men and women, employment situation of adult men and women, and health situation of children.

3 Results of Out-of-Sample Predictions

3.1 Regression

A first test for the stability of estimation results arises when we change the regression model itself. The idea for the estimation in Grosse et al. (2009) was to use the model with the largest number of possible regressors with the data at hand.[14] The estimation model should transfer the largest possible correlation structure from the LSMS to the DHS data. Since the authors did not want to explain causalities or establish an expenditure or income model (as it is the purpose of standard income regressions), insignificance of coefficients and multicollinearity were left aside. For each year, the largest possible model was used, i.e., with different coefficients for the different years, since the questionnaire designs have changed over time.

We start by comparing the results of the models using all possible variables, i.e., different ones at different points in time, with the results using only the common model. The impact on the regression is presented in Table 1 showing the β coefficients and P-values. As expected, the explanatory power of the common model is lower (shown by a lower R^2). Most coefficients keep the sign, especially the significant ones. Concerning signs and the significance level, the evolution over time is shown in Appendix Table 7. The

[14] Branisa (2006) reproduces the calculations for poverty and inequality for 1989 and 1994 based on Grosse, Klasen, and Spatz (2005) (which is an earlier version of the Grosse et al. (2009) paper on which the pre-test was based), but with 2002 as the base year instead of 1999, and with the largest number of possible regressors. His point estimates for the poverty measures, compared to the results presented in subsection 3.4, are systematically below those computed by Grosse et al. (2005) for both years. Apparently, the change of the base period has in this case an impact on the level of the estimates. Nevertheless, and as was also the case in Grosse et al. (2005), estimations for the year 1989 are higher than the corresponding poverty measures calculated with observed incomes for 1999. Hence, it is confirmed that poverty has decreased between 1989 and 1999. What happened between 1994 and 1999 is unclear, as the point estimates for 1994 by Branisa (2006) are in general closer to the values computed with observed income for 1999. In the case of the point estimates of inequality measures, his estimates are also on every occasion below the ones computed by Grosse et al. (2005) for 1989 and 1994. Once again, the change of the base period appears to have an effect on the level of the estimates. When comparing the inequality measures based on observed incomes for 1999 with those based on simulated incomes for 1994 by Branisa (2006) and by Grosse et al. (2005), it is found that inequality in 1999 is between both estimated values for 1994. Therefore, it is unclear whether inequality has decreased or increased between 1994 and 1999. If one looks at the evolution of inequality between 1989 and 1999, a similar pattern is observed: estimates by Grosse et al. (2005) suggest a decrease in inequality, while estimates by Branisa (2006) suggest hardly any changes.

models of 1989 and 1994 for cities perform way better than for the later years when looking at significance levels. The number of insignificant variables increases for example from 16 in 1989 (12 in 1994) to 30 in 1999 and 28 in 2002 for cities, out of a total of 51 variables. The regressions are weak for rural areas both in 1999 and 2002 and even weaker for towns, which are based on the lowest number of observations. The final regression model is shown in Table 2. Besides reducing the number of regressors to common variables, the main changes are in the variables showing household composition, which are reduced from 6 to 3 variables, for household headship from 5 to 2 variables, for employment from 14 to 8 variables, and for child health from 6 to 3 variables. For department dummies, education, gender, and access to infrastructure, the variables remained the same.[15]

[15] Results and some discussion on other dynamics represented by ϕ in Equation (6) are found in section 4.

Table 1: Results of Log-Linear OLS Regression, Full Model versus Common Model, 1999

	City				Town				Rural			
	all		common		all		common		all		common	
	β	P	β	P	β	P	β	P	β	P	β	P
La Paz	0.09	0.39	-0.02	0.88	0.13	0.81	0.11	0.81	0.19	0.04	0.25	0.00
Cochabamba	0.28	0.01	0.27	0.02	0.62	0.22	0.65	0.17	0.28	0.01	0.40	0.00
Oruro	0.04	0.75	-0.06	0.66	-0.26	0.65	-0.27	0.61	0.31	0.03	0.27	0.08
Potosi	0.10	0.45	-0.02	0.89	0.14	0.78	0.16	0.74	0.04	0.65	0.06	0.57
Tarija	0.59	0.00	0.53	0.00	0.37	0.49	0.47	0.33	0.64	0.00	0.63	0.00
Santa Cruz	0.68	0.00	0.70	0.00	0.47	0.35	0.52	0.27	0.74	0.00	0.67	0.00
Beni & Pando	0.70	0.00	0.63	0.00	0.17	0.75	0.18	0.72	0.81	0.00	0.71	0.00
# elderly	-0.08	0.60	0.11	0.41	0.09	0.73	0.09	0.73	-0.08	0.34	-0.11	0.27
# males	-0.07	0.02	-0.07	0.03	0.10	0.22	0.08	0.33	-0.10	0.02	-0.08	0.05
# females	-0.12	0.00	-0.12	0.00	-0.10	0.09	-0.11	0.07	-0.17	0.00	-0.15	0.00
# youngsters	-0.03	0.62	-0.01	0.76	-0.08	0.23	-0.08	0.16	-0.01	0.79	-0.02	0.55
# children	-0.11	0.16	-0.10	0.21	-0.18	0.05	-0.20	0.02	-0.08	0.10	-0.10	0.05
# of working age / hh size	1.02	0.01	1.13	0.00	0.22	0.66	0.14	0.80	0.74	0.01	0.60	0.04
gender hh head	0.03	0.73	0.00	0.97	0.25	0.15	0.27	0.12	-0.02	0.84	0.05	0.57
language of hh head	-0.01	0.86			-0.12	0.30			-0.06	0.32		
hh head age ¡= 24	-0.21	0.31	-0.41	0.05	0.01	0.98	0.04	0.91	0.01	0.98	0.04	0.83
hh head age 25 - 34	-0.25	0.22	-0.27	0.19	0.03	0.94	0.05	0.90	0.05	0.74	0.06	0.73
hh head age 35 - 44	-0.39	0.05	-0.29	0.14	0.01	0.99	0.04	0.92	0.08	0.62	0.12	0.48
hh head age 45 - 54	-0.45	0.03	-0.34	0.09	0.13	0.77	0.14	0.72	-0.04	0.80	-0.06	0.72
hh head age 55 - 65	-0.34	0.09	-0.21	0.31	0.03	0.94	0.04	0.92	0.03	0.84	0.05	0.78
has house	0.07	0.20			-0.07	0.51			0.08	0.25		
floor (cement)	0.17	0.21			0.03	0.86			0.24	0.00		
floor (brick)	0.30	0.05			0.17	0.33			0.00	1.00		
floor (other floor)	0.38	0.01			0.10	0.61			0.24	0.02		
2-3 sleeping rooms	0.21	0.00			-0.18	0.11			0.07	0.24		
¿=4 sleeping rooms	0.22	0.04			0.09	0.73			0.30	0.14		
access to public water	-0.18	0.11	-0.07	0.52	0.06	0.63	0.02	0.91	-0.07	0.22	0.02	0.73
has no toilet	-0.02	0.86	-0.05	0.59	-0.22	0.10	-0.28	0.05	-0.08	0.11	-0.22	0.00
has electricity	-0.32	0.03			-0.19	0.46			0.13	0.05		
cooking material	-0.26	0.02			-0.02	0.91			0.30	0.00		
has phone	0.24	0.00			0.38	0.00			0.30	0.01		
has radio	0.02	0.79			-0.11	0.29			0.10	0.07		
has television	0.18	0.10			0.10	0.54			0.23	0.01		
has fridge	0.23	0.00			0.03	0.77			-0.02	0.85		

continued on next page

Table 1 continued

	City				Town				Rural			
	all		common		all		common		all		common	
	β	P	β	P	β	P	β	P	β	P	β	
no partner in household	0.31	0.15	0.11	0.64	0.52	0.15	0.47	0.17	0.38	0.01	0.41	0.03
com. basic edu. (m.)	-0.12	0.35	-0.16	0.21	-0.01	0.96	0.01	0.96	0.02	0.78	-0.05	0.55
incom. secondary edu. (m.)	0.04	0.70	-0.03	0.78	-0.20	0.25	-0.14	0.44	-0.04	0.56	-0.05	0.47
com. secondary edu. (m.)	-0.04	0.67	-0.06	0.52	0.11	0.48	0.26	0.10	-0.02	0.83	0.03	0.77
tertiary edu. (m.)	0.24	0.03	0.35	0.00	-0.10	0.66	0.01	0.97	0.15	0.49	0.56	0.01
com. basic edu. (w.)	-0.02	0.89	0.10	0.44	0.04	0.81	0.07	0.70	0.20	0.00	0.27	0.00
incom. secondary edu. (w.)	0.05	0.64	0.14	0.17	0.12	0.41	0.16	0.23	0.27	0.00	0.28	0.00
com. secondary edu. (w.)	0.06	0.54	0.26	0.01	0.11	0.50	0.18	0.20	0.18	0.08	0.43	0.00
tertiary edu. (w.)	0.26	0.03	0.53	0.00	0.27	0.19	0.48	0.01	0.28	0.17	0.59	0.02
high skilled white collar (m.)	0.68	0.00	0.74	0.00	1.09	0.01	1.07	0.00	0.60	0.00	0.99	0.00
med. skilled white collar (m.)	0.41	0.03	0.37	0.09	1.02	0.01	0.95	0.01	0.45	0.00	0.64	0.00
skilled manual (m.)	0.44	0.02	0.24	0.25	0.69	0.07	0.55	0.13	0.54	0.00	0.73	0.00
unskilled manual (m.)	0.37	0.08	0.22	0.35	0.45	0.21	0.35	0.34	0.45	0.00	0.54	0.00
agr. employed (m.)	-0.19	0.60	-0.32	0.48	0.47	0.28	0.46	0.25	0.48	0.00	0.57	0.00
agr. self-employed (m.)	0.88	0.01	0.27	0.29	0.07	0.88	-0.09	0.82	0.31	0.01	0.31	0.06
sales and services (m.)	0.51	0.01	0.34	0.12	0.94	0.02	0.86	0.02	0.47	0.00	0.72	0.00
high skilled white collar (w.)	0.35	0.01	0.40	0.01	0.51	0.02	0.65	0.00	0.04	0.90	-0.06	0.83
med. skilled white collar (w.)	0.24	0.01	0.32	0.00	0.77	0.00	0.84	0.00	0.26	0.07	0.17	0.27
skilled manual (w.)	0.03	0.78	-0.07	0.53	0.37	0.02	0.44	0.00	-0.09	0.35	-0.11	0.29
unskilled manual (w.)	0.32	0.00	0.31	0.00	0.61	0.00	0.63	0.00	-0.08	0.51	-0.03	0.85
agr. employed (w.)	1.20	0.02	0.93	0.11	-0.81	0.17	-1.02	0.06	0.07	0.45	-0.04	0.66
agr. self-employed (w.)	0.53	0.00	0.51	0.01	-0.32	0.33	-0.24	0.46	0.03	0.64	-0.05	0.47
sales and services (w.)	0.30	0.00	0.28	0.00	0.67	0.00	0.73	0.00	0.20	0.06	0.31	0.00
has social security	0.09	0.09			0.08	0.48			0.16	0.05		
birth in last 12 month	0.08	0.71	0.18	0.39	-0.32	0.30	-0.25	0.36	-0.05	0.51	-0.08	0.27
attended by doctor	-0.09	0.72	-0.16	0.51	0.63	0.09	0.61	0.07	0.11	0.32	0.21	0.07
delivered in hospital	-0.08	0.64	-0.09	0.57	-0.20	0.37	-0.25	0.25	0.12	0.31	0.10	0.45
child under 4 years	0.02	0.86	0.07	0.48	0.14	0.57	0.12	0.48	0.13	0.29	-0.04	0.65
has first polio vaccination	0.05	0.69			-0.04	0.84			-0.20	0.10		
has triple dpt vaccination	0.06	0.61			-0.02	0.91			0.01	0.85		
has had diarrhea	-0.14	0.14	-0.18	0.07	0.04	0.79	0.00	0.98	0.03	0.60	0.02	0.74
has head cough/fever	0.03	0.67	-0.02	0.85	0.08	0.54	0.06	0.65	0.02	0.71	0.01	0.87
c/t/r dummy/constant	4.57	0.00	4.63	0.00	3.95	0.00	3.82	0.00	3.53	0.00	3.88	0.00
# of observations	1037		1037		332		332		922		922	
R^2	55.74		47.15		58.19		53.43		57.11		48.89	

Notes: For details on the regression and variables, see text and notes of Table 2. β: regression coefficient; P: P-value; all: all possible covariates; common: covariates common over all 4 years.
Source: Own calculations based on ECH and EIH.

Table 2: Results of Log-Linear OLS Regression, Reduced Model, 1989–2002

	City								Town				Rural			
	1989		1994		1999		2002		1999		2002		1999		2002	
	β	P	β	P	β	P	β	P	β	P	β	P	β	P	β	P
La Paz	0.01	0.77	0.15	0.00	-0.04	0.70	-0.03	0.75	0.16	0.75	0.18	0.26	0.25	0.00	0.26	0.00
Cochabamba	0.16	0.00	0.13	0.00	0.24	0.03	0.17	0.06	0.75	0.14	0.17	0.26	0.39	0.00	0.16	0.04
Oruro	-0.17	0.00	-0.20	0.00	-0.06	0.65	-0.23	0.04	-0.18	0.74	-0.11	0.56	0.35	0.01	0.23	0.01
Potosi	-0.26	0.00	-0.21	0.00	-0.07	0.63	-0.08	0.41	0.18	0.72	-0.09	0.62	0.05	0.62	-0.08	0.38
Tarija	-0.03	0.52	0.03	0.53	0.50	0.00	0.16	0.13	0.57	0.27	0.40	0.01	0.71	0.00	0.56	0.00
Santa Cruz	0.43	0.00	0.43	0.00	0.70	0.00	0.44	0.00	0.58	0.25	0.11	0.47	0.71	0.00	0.46	0.00
Beni & Pando	0.44	0.00	0.28	0.00	0.62	0.00	0.31	0.00	0.29	0.57	0.25	0.10	0.74	0.00	0.59	0.00
elderly dependency ratio	-0.23	0.00	-0.28	0.00	-0.20	0.00	-0.32	0.00	-0.18	0.01	-0.24	0.00	-0.03	0.28	-0.08	0.01
child dependency ratio	0.08	0.42	-0.08	0.41	0.17	0.51	0.00	1.00	0.55	0.13	-0.01	0.95	-0.12	0.41	0.05	0.60
hh size	-0.07	0.00	-0.05	0.00	-0.08	0.00	-0.05	0.00	-0.06	0.01	-0.05	0.01	-0.09	0.00	-0.10	0.00
hh head age	0.03	0.00	0.01	0.18	0.00	0.75	0.02	0.14	0.01	0.39	0.01	0.54	0.02	0.18	0.03	0.00
hh head age squared	0.00	0.01	0.00	0.69	0.00	0.91	0.00	0.21	0.00	0.33	0.00	0.66	0.00	0.14	0.00	0.00
gender hh head	-0.12	0.05	-0.12	0.01	-0.05	0.59	0.03	0.73	0.22	0.19	0.06	0.54	-0.01	0.87	-0.12	0.24
access to public water	0.15	0.00	0.03	0.20	-0.08	0.51	-0.05	0.48	0.05	0.70	0.06	0.49	0.00	0.99	0.13	0.00
has no toilet	-0.20	0.00	-0.21	0.00	-0.03	0.77	-0.04	0.58	-0.27	0.06	0.06	0.52	-0.23	0.00	-0.15	0.00
no partner in household	0.35	0.00	0.58	0.00	0.21	0.37	0.34	0.03	0.47	0.15	0.25	0.24	0.39	0.03	0.06	0.77
com. basic edu. (m.)	0.02	0.72	0.03	0.51	-0.16	0.23	-0.04	0.68	-0.02	0.89	0.05	0.65	-0.02	0.79	0.15	0.01
incom. secondary edu. (m.)	0.02	0.65	0.05	0.17	-0.01	0.95	-0.11	0.36	-0.17	0.37	0.07	0.47	-0.01	0.87	0.11	0.04
com. secondary edu. (m.)	0.10	0.03	0.10	0.01	-0.05	0.63	0.06	0.51	0.29	0.07	0.07	0.44	0.05	0.57	0.19	0.01
tertiary edu. (m.)	0.52	0.00	0.40	0.00	0.39	0.00	0.43	0.00	-0.02	0.93	0.29	0.02	0.44	0.06	0.27	0.07
com. basic edu. (w.)	0.00	0.95	0.07	0.07	0.12	0.33	-0.07	0.57	-0.01	0.96	-0.01	0.96	0.26	0.00	0.22	0.00
incom. secondary edu. (w.)	0.14	0.01	0.01	0.74	0.10	0.34	0.01	0.88	0.11	0.41	0.05	0.60	0.27	0.00	0.26	0.00
com. secondary edu. (w.)	0.17	0.00	0.06	0.06	0.24	0.02	0.02	0.77	0.12	0.40	0.31	0.00	0.44	0.00	0.33	0.00
tertiary edu. (w.)	0.39	0.00	0.40	0.00	0.53	0.00	0.44	0.00	0.45	0.01	0.58	0.00	0.75	0.01	0.64	0.00

continued on next page

Table 2 continued

	City								Town				Rural			
	1989		1994		1999		2002		1999		2002		1999		2002	
	β	P	β	P	β	P	β	P	β	P	β	P	β	P	β	P
high & med. skilled white collar (m.)	0.45	0.00	0.79	0.00	0.53	0.01	0.62	0.00	1.03	0.00	0.61	0.00	0.67	0.00	0.21	0.27
skilled & unskilled manual (m.)	0.37	0.00	0.47	0.00	0.23	0.27	0.23	0.05	0.57	0.09	0.41	0.01	0.60	0.00	0.04	0.80
agriculture (m.)	0.42	0.00	0.51	0.00	-0.21	0.61	0.21	0.22	0.19	0.59	0.30	0.14	0.25	0.12	-0.05	0.75
sales and services (m.)	0.42	0.00	0.57	0.00	0.34	0.12	0.53	0.00	0.87	0.01	0.62	0.00	0.67	0.00	0.26	0.15
high & med. skilled white collar (w.)	0.45	0.00	0.46	0.00	0.38	0.00	0.55	0.00	0.73	0.00	0.72	0.00	0.13	0.41	0.43	0.00
skilled & unskilled manual (w.)	0.22	0.00	0.28	0.00	0.14	0.07	0.19	0.01	0.51	0.00	0.17	0.03	-0.12	0.19	0.05	0.54
agriculture (w.)	0.52	0.01	0.10	0.57	0.64	0.04	-0.14	0.61	-0.24	0.49	-0.54	0.00	-0.05	0.42	-0.08	0.19
sales and services (w.)	0.34	0.00	0.30	0.00	0.31	0.00	0.23	0.00	0.71	0.00	0.35	0.00	0.36	0.00	0.25	0.00
birth in last 12 month	-0.13	0.03	-0.13	0.01	0.20	0.23	-0.06	0.64	-0.29	0.36	-0.17	0.26	-0.12	0.09	-0.03	0.71
attended by doctor	0.07	0.45	0.04	0.62	-0.22	0.28	-0.02	0.89	0.68	0.05	0.10	0.55	0.20	0.10	0.02	0.86
delivered in hospital	0.03	0.74	0.00	0.98	-0.10	0.49	-0.26	0.14	-0.38	0.06	0.12	0.37	0.08	0.56	0.15	0.26
c/t/r dummy/constant	4.31	0.00	4.66	0.00	5.05	0.00	4.80	0.00	3.84	0.00	4.41	0.00	4.09	0.00	4.23	0.00

Notes: β: regression coefficient, P: P-value. The variables are defined as follows: Of the nine departments, Beni and Pando are grouped into one single variable (left-out category: Chuquisaca). The elderly (child) dependency ratio is number of elderly (children) divided by number of men and women in working age, and the total number of household members is additionally included (hh size). We include age and age squared of the household head as well as gender of the household head. For infrastructure, due to changes in questionnaire design, we are only able to include access to public water and (not) having a toilet. For education, we include (for men (m.) and women (w.) separately) the categories complete (com.) basic, incomplete (incom.) secondary, complete secondary, and tertiary education (left-out category: no or incomplete basic eduction). For employment, we we include (for men (m.) and women (w.) separately) the categories high and medium skilled white collar, skilled and unskilled manual, employed and self-employed in agriculture (agr.), and sales and services (left-out category: unknown or unemployed). For health, we include only the variables on how the last birth took place (attended by doctor and/or in hospital). Further, the constants for the three regions (city, town, rural) are included (c/t/r).
Source: Own calculations based on ECH and EIH.

3.2 Method

We consider the following three poverty measures: the headcount ratio (P0), the poverty gap ratio (P1), and the squared poverty gap ratio (P2). These three measures are special cases of the general P(α) family of poverty measures proposed by Foster, Greer, and Thorbecke (1984). The parameter α is an indicator of the degree to which inequality among the poor is considered relevant in assessing poverty. For inequality we consider the Gini coefficient and the Atkinson family of inequality indices (Atkinson, 1970) with a constant inequality aversion parameter γ that allows giving more or less emphasis to redistributions that take place at the lower end of the income distribution. We compute inequality using 0.5 and 2.0 for γ. A higher value of this parameter will give more importance to income transfers that make income differences smaller at the bottom of the distribution relative to those at the top of it. Jenkins (1991) states that the Atkinson measure becomes very bottom-sensitive if γ is larger than 1.0.

As we are mainly concerned with the stability and reliability of the evolution of poverty and inequality in Bolivia in the period 1989–2002, we do not only need point estimates for the relevant figures, but also proper confidence intervals. In Grosse et al. (2009) standard errors were computed for the measures corresponding to predicted income which were based on 200 simulations where an error term was added to the predicted values from the regression. We construct 95 percent confidence intervals as follows.

Concerning poverty and inequality measures using observed income, confidence intervals are constructed based on the asymptotic distribution of the measures.[16] Kakwani (1993) proposes a general method for deriving the distribution of poverty indices that are additively separable and provides formulas for the estimated standard errors of the FGT poverty measures. It is interesting to highlight that Kakwani (1993) finds that the precise estimation of a poverty measure depends on how sensitive the measure is to income transfers among the very poor. For FGT measures, this is reflected in the parameter α. The precision of the poverty measure is a monotonically decreasing function of this parameter. In other words, a higher α means larger standard errors for a given sample size.

The standard errors that Kakwani (1993) proposed are valid under the assumption that the sample was collected under a simple random design. We follow Jolliffe and Semykina (1999) who extend this approach and provide estimated standard errors for the

[16] An alternative would have been to use bootstrap methods for computing the confidence intervals (Biewen, 2002). The accuracy of asymptotic and bootstrap methods for poverty and inequality measures is discussed by Davidson and Flachaire (2007).

FGT poverty measures which are robust to complex survey design, such as stratification and multiple stages.

Inequality measures are usually nonlinear functions of the observations, and for complex surveys their variances are hard to estimate. Approximate variance estimation techniques have been proposed which rely on linearization methods such as a Taylor series approximation. For the Gini coefficient, we use the approximate standard error estimation technique proposed by Kovacevic and Binder (1997) which is based on the theory of estimating equations (Binder and Patak, 1994). They show that for complex survey design the estimated variance of the Gini coefficient can be estimated based on the variance of the estimated population totals.

For the Atkinson measures, we apply the linearization method proposed by Biewen and Jenkins (2006) who draw on an approximation of the variance suggested by Woodruff (1971). Starting from the fact that Atkinson inequality indices can be written in terms of population totals of the variable of interest, they derive an expression for the sampling variance that can handle complex survey design aspects.

With respect to measures based on predicted and simulated income, confidence intervals are also based on the asymptotic distribution of theses measures. We assume that predicted income is similar (in its statistical properties) to observed income and apply the techniques for poverty and inequality confidence intervals as described above. As predicted income is based on information from two surveys, we acknowledge that the confidence intervals are too narrow as they do not explicitly consider sampling error. One main difference between the approach pioneered by Hentschel et al. (2000) and Elbers et al. (2003) and the dynamic extension suggested by Grosse et al. (2009) and Stifel and Christiaensen (2007) is that the former studies combine a Census with a household survey, while the latter combine two surveys with the obvious implication that in the latter case sampling error is an issue.[17]

3.3 Income

Taking a first look at the properties of observed, predicted (i.e., within the LSMS data), and simulated (i.e., over to the DHS data) incomes for the years from 1989 to 2002 reveals that different estimation assumptions, different data sets, and different base years

[17] One could think of using bootstrap methods to account for the sampling error. We have decided not to follow this approach for practical reasons. The computations needed for this paper are already very time consuming taking around half a day to run, and considering doing at least 100 replications seems unfeasible for the time being.

deliver different results for Bolivia. For the four years, we present several different values, depending on the base year and the dynamics of regression coefficients assumed (Table 3). For 1989, the first column shows the observed values as calculated from the LSMS. The next three columns show the within-LSMS predictions. The abbreviation GKS stands for the assumptions of Grosse et al. (2009) following Equations (2) and (4) and the abbreviation SC for Stifel and Christiaensen (2007) following Equations (3) and (5). The number stands for the base year. For example, the column "SC99" refers to the estimation using the Stifel and Christiaensen (2007) method and the base year 1999, i.e., the coefficients of 1999. For the GKS case, the predictions for cities in the LSMS data set is the same independently of the base year because the method always uses the observed coefficients for cities of the respective years instead of the ones from the base year, that is why the column is labeled "GKS9902".

For the years 1989 and 1994, the comparison of observed incomes with predicted incomes in the LSMS is limited to cities.[18] Of course, when comparing the poverty and inequality measures based on observed and simulated incomes for cities, the purpose is not to verify implicitly whether the two assumptions (about the dynamics of regression coefficients and the variances of the error terms in the model) seem to work, as these assumptions are only used with the other two regions (town and rural). The second column "GKS9902" shows how well within-LSMS-sample prediction works using the "true" 1989 coefficients for cities. It slightly underestimates the income mean and overestimates P50. Using the coefficients from 1999 (third column "SC99") more strongly overestimates mean income, and when using the coefficients from 2002 mean income is relatively close to observed values (fourth column "SC02"). Simulating incomes in the DHS surveys overestimates incomes (mean and P50) for both assumptions and for both base years, but worse so for SC case. In nearly all cases, the standard deviation is too low compared to observed values.

[18]The comparison of the summary statistics shown is based on one example, i.e., on one prediction and simulation run. It is not based on the average income of the 200 repetitions. It would be slightly more accurate to present the average of each of the summary statistics, for example the average of all P50 instead of P50 of one income set.

Table 3: Income Properties from LSMS and DHS, 1989–2002

	1989								1994							
	LSMS				DHS				LSMS				DHS			
	obs.	predicted			simulated				obs.	predicted			simulated			
		GKS9902	SC99	SC02	GKS99	GKS02	SC99	SC02		GKS9902	SC99	SC02	GKS99	GKS02	SC99	SC02
Total Bolivia																
Mean y					215	215	239	232					231	241	266	255
Min y					3	5	5	4					4	11	6	6
Max y					4953	4953	4813	5358					4229	4229	4619	6273
P5 y					27	30	30	30					24	38	32	32
P50 y					126	128	135	130					142	160	154	143
P95 y					714	701	814	761					738	714	857	819
SD y					290	285	326	346					281	270	349	369
City																
Mean y	310	296	323	305	311	311	350	337	326	316	339	341	337	337	382	370
Min y	3	6	9	4	10	10	11	5	5	13	13	10	11	11	10	9
Max y	7988	4274	5549	5530	4953	4953	4813	5358	7752	5691	7216	10570	4229	4229	4619	6273
P5 y	48	45	50	40	47	47	51	42	61	56	59	41	57	57	58	43
P50 y	190	196	214	186	200	200	217	198	208	214	227	199	236	236	250	228
P95 y	922	891	977	990	943	943	1066	1061	963	910	977	1071	922	922	1124	1153
SD y	444	318	355	371	366	366	414	454	408	335	382	472	338	338	430	477
Town																
Mean y					184	206	208	218					219	222	263	238
Min y					5	17	9	18					4	18	8	15
Max y					2163	2166	2137	3270					1996	1534	3243	2384
P5 y					21	41	26	41					26	53	31	42
P50 y					108	144	128	150					138	169	158	162
P95 y					696	576	688	626					709	568	871	713
SD y					235	192	246	224					245	186	326	257
Rural Areas																
Mean y					112	106	119	115					107	132	129	123
Min y					3	5	5	4					4	12	6	6
Max y					1932	1247	1726	1575					1194	1113	1770	1317
P5 y					23	23	25	26					17	31	26	27
P50 y					79	75	85	84					71	99	90	88
P95 y					301	293	326	304					308	332	368	349
SD y					113	104	121	108					112	110	127	117

continued on next page

Table 3 continued

	1999							2002						
	LSMS				DHS			LSMS				DHS		
	obs.	predicted			simulated			obs.	predicted			simulated		
		GKSSC99	GKS02	SC02	GKSSC99	GKS02	SC02		GKSSC02	GKS99	SC99	GKSSC02	GKS99	SC99
	Total Bolivia													
Mean y	266	278	297	267	303	320	294	256	263	258	278	304	301	314
Min y	1	7	12	8	3	5	3	0	7	2	6	5	2	2
Max y	7347	6950	6950	7185	5720	5720	6479	6858	9406	9406	7301	12655	12655	10853
P5 y	31	32	44	29	36	43	36	31	30	20	31	36	24	35
P50 y	159	157	184	146	180	204	165	142	140	128	150	168	152	176
P95 y	901	923	921	902	948	957	971	784	866	898	878	975	1033	1013
SD y	355	410	405	418	385	377	412	421	416	428	401	500	514	459
	City													
Mean y	378	404	404	393	412	412	402	361	377	377	393	441	441	454
Min y	1	12	12	8	11	11	6	0	12	12	16	6	6	15
Max y	7347	6950	6950	7185	5720	5720	6479	6858	9406	9406	7301	12655	12655	10853
P5 y	66	57	57	45	63	63	50	51	41	41	52	50	50	62
P50 y	247	247	247	236	270	270	239	207	213	213	233	255	255	285
P95 y	1100	1188	1188	1204	1208	1208	1248	1267	1298	1298	1233	1379	1379	1362
SD y	442	525	525	548	450	450	504	546	539	539	500	675	675	600
	Town													
Mean y	258	248	272	229	289	314	274	244	248	261	285	281	295	297
Min y	2	7	13	14	10	10	7	3	10	3	6	6	2	2
Max y	2703	2504	2598	2386	3507	3067	3455	4131	3245	4236	2926	4354	3801	3393
P5 y	39	46	56	44	30	54	44	42	48	22	38	44	21	32
P50 y	167	164	209	160	182	219	180	169	164	143	174	183	159	181
P95 y	724	716	730	626	863	929	832	713	764	812	872	791	1035	977
SD y	267	255	241	210	357	306	310	268	261	352	336	347	408	344
	Rural Areas													
Mean y	112	114	158	108	129	169	124	115	109	91	115	141	124	142
Min y	8	12	15	10	3	5	3	6	7	2	9	5	3	5
Max y	1391	1178	1078	895	1885	1910	2018	2496	2697	2091	2005	3651	3769	4700
P5 y	24	24	35	23	27	31	27	24	24	13	23	28	17	27
P50 y	81	78	119	79	89	120	90	84	84	59	83	98	77	97
P95 y	297	301	412	272	353	485	331	301	284	271	310	394	388	418
SD y	108	112	134	99	129	168	120	126	104	105	110	144	154	154

Notes: See subsection 3.3 for explanation.
Source: Own calculations based on ECH, EIH, and DHS.

Nearly the same holds for 1994, where the within-LSMS-survey predictions using the GKS assumption comes relatively close to observed values (tenth column) whereas the SC assumption of no dynamics overestimates mean income (eleventh and twelfth column). P50 is overestimated for the 1999 coefficients and underestimated for the 2002 coefficients. The trend from 1989 to 1994 is, in general, nearly always the same: an increase in mean income in all regions for all specifications except in one case. The income level for each of the assumptions, however, is different. For example, the observed income in cities increases from 310 to 326, whereas for the two predictions using the SC assumption, the increase would be from 323 to 339 ("SC99") or 305 to 341 ("SC02").

For the later years 1999 and 2002 (second part of Table 3), more comparisons are possible. First, there are observed incomes in all three regions. In cities and towns, income goes down whereas it goes slightly up in rural areas between 1999 and 2002. Overall, this leads to a decrease of income at the national level. The column "GKSSC99" shows the within-survey prediction for the year 1999, i.e., applying the coefficients to the same data and predicting incomes. Stronger differences arise if 2002 is taken as a base year, and different assumptions are used to estimate the 1999 value. "GKS02" presents the results of out-of-sample predictions in pretending that the LSMS survey was only urban, and applying Equation (2) (to both LSMS and DHS) whereas "SC02" applies Equation (3), i.e., the coefficients from 2002 (also to both LSMS and DHS). Even within the LSMS data set, there is a tendency for overestimation of incomes, especially for cities. In addition, using the GKS assumption and 2002 as a base year, the overestimation becomes stronger in rural areas. Going to the DHS the results are even less encouraging. Both assumptions and both base years overestimate incomes in all regions. Yet, the trends in the DHS data from 1994 to 1999 remain similar: again we find increasing mean incomes, but to a different level. For 2002, the income results resemble the ones of 1999. However, within-LSMS results are slightly better, and also the ones using the GKS assumption and 1999 as the base year (comparing "obs." with "GKS99" in the LSMS columns). Here, the assumption of no dynamics of SC delivers the worst results for within-LSMS-predictions. But again, going to the DHS data also renders an overall overestimation of incomes, independently of the assumption and base year. Worst results for simulations over-to-DHS are also achieved using "SC99".

Besides the question of how close estimates come to observed values, the within-country differences become clear. Mean income in rural areas is only about one-third of the value of cities. Towns show also lower values than cities and are most of the time close to the national average. These relations remain over time, leading to the alarming question on how to avoid that rural areas become more and more detached from overall growth.

3.4 Evolution of Poverty and Inequality

We discuss here the main results on poverty and inequality using a graphical presentation: Figures 1 to 8 show moderate poverty for P0, P1, and P2, and inequality for Gini and two Atkinson indices (with the inequality-aversion parameter γ of 0.5 and 2.0). The results for extreme poverty are shown in Appendix Figures 9 to 12. For the test of the different assumptions on dynamics, we focus on the three regions and on the two years for which we have complete data (1999 or 2002) so that we can compare the measures based on simulated income, for both GKS and SC, with the measures based on observed income. When we refer to simulated values in a given year, we mean that the other year has been used as the base year for the model in both GKS and SC, with numbers after the abbreviation showing the base year. For the base year, and by construction, GKS and SC yield the same results.

Results for total Bolivia show that results differ depending on the method and base year chosen. Moderate poverty (Figure 1) declines from 1989 to 1999, independently of the method and base year used. The level and dynamics, however, differ substantially depending on both base year and method. For example, using 1999 as base year, the SC assumption gives significantly lower values as the GKS assumption. The trend of poverty in the crisis-driven years between 1999 and 2002 is not clear. P0 clearly increases when looking at observed values, whereas P1 and P2 stagnate (P1 with a slightly increasing trend, P2 with a slightly decreasing trend). The SC method delivers a steady downward trend for all measures, whereas GKS is able to reproduce the increase in P0 slightly better. Worth noting furthermore is the level difference in the base year between LSMS and DHS data. The P0 estimation is 5 percentage points below the observed values for 1999 and even 7 for 2002,[19] without overlapping confidence intervals.[20] Inequality results (Figure 2) also strongly depend on base year and method. For 1989, results are nearly the same independently of year and method. For 1994, the same holds for base year 1999. In this case, inequality seems to have remained constant in the 1990s and only increased

[19]Remember that for the base year, and by construction, GKS and SC yield the same results.

[20]The picture for extreme poverty (Appendix Figure 9) is very similar, however with a less strong increase in P0 and a clearer downward trend of P1 and P2.

from 1999 onwards. But for 2002 as base year, differences become stronger, since the GKS assumption shows a decrease until 1994 followed by an increase in the next observation years. Overall, GKS better reflects observed trends.

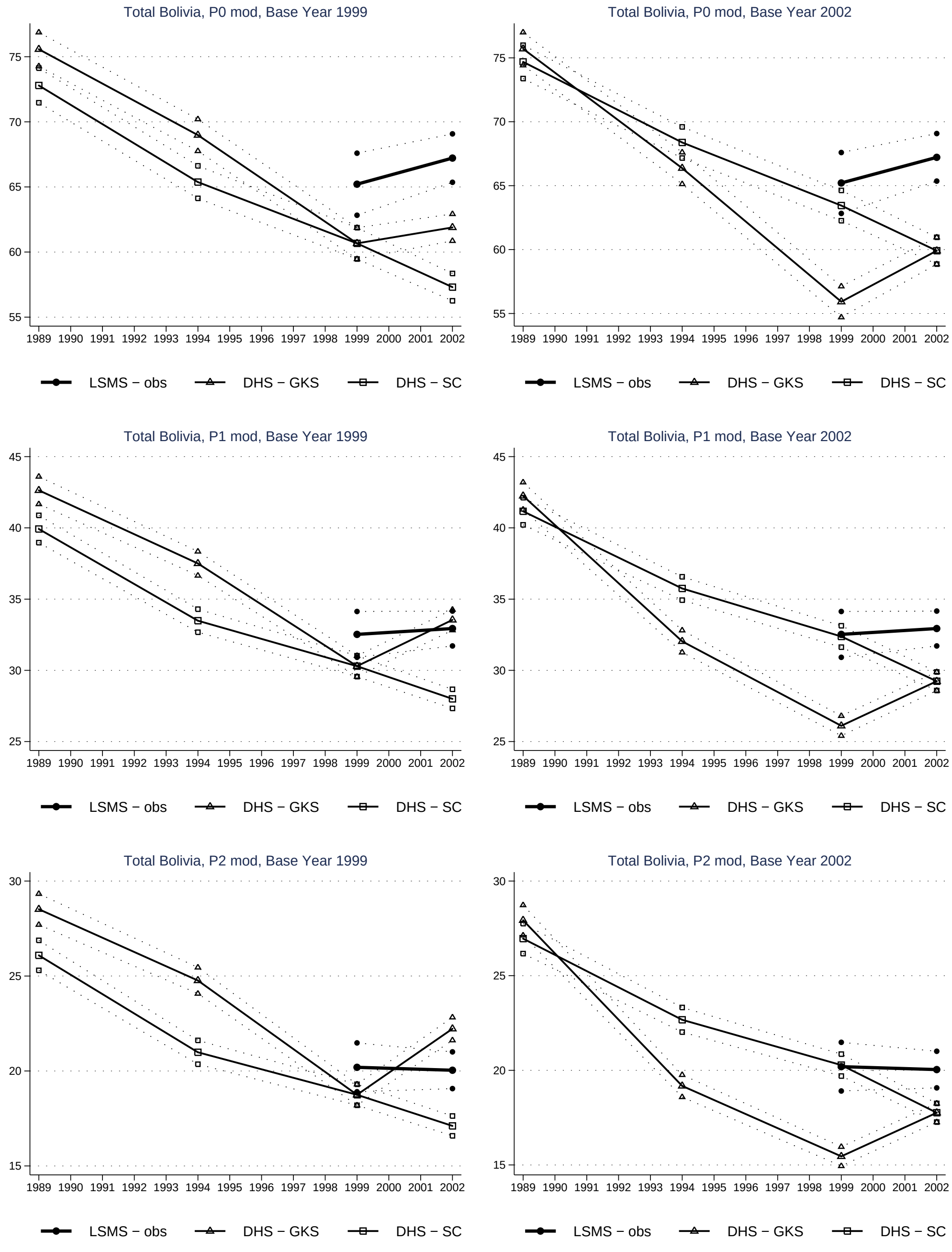
Looking at cities reveals, first, how well the cross-survey-matching method works. Here, the GKS method uses the actual regression coefficients in the LSMS data to simulate incomes in the DHS data. The SC method uses either 1999 (left part of Figure 3) or 2002 (right part). For P0, the GKS assumption delivers lower point estimates, sometimes not with overlapping confidence intervals. Results for SC are worse, with significantly lower levels for 1999 and mixed results for 2002. For P1 and P2, results of GKS are very close to observed values, whereas SC significantly underestimates P1 and P2 with 1999 as the base year and significantly overestimates P1 and P2 with 2002 as the base year. This result advocates the doubts about the accuracy of using regression coefficients of one year for estimations of other years back and forth in time, as SC does, without taking dynamics into account.

Taking a closer look at the end of the observation period reveals that the three poverty measures based on observed income seem to have increased between 1999 and 2002. With 1999 as the base year, this trend is not well replicated using SC, which suggests a decline in poverty. GKS suggests relative little changes between both years. With 2002 as the base year, results are similar. If we focus on levels, it is clear that the estimates based on GKS are closer to the results based on observed values than the estimates based on SC. This is not surprising, as SC implies that the results from the regression of the base year are used for the other year, while GKS uses the results from the regression of the corresponding year to simulated values for cities. It is nevertheless surprising that for P0 (moderate poverty) the point estimate of GKS for 2002 is below the lower bound of the 95 percent confidence interval of poverty computed with observed income, and that both confidence intervals, the one for observed income and the one for simulated income, do not overlap. The situation is better for P1 and P2, as at least the confidence intervals overlap. The simulated result come closer to observed values when using GKS for the year 1999.[21]

Turning to inequality, Figure (4) reveals that the corresponding coefficients are able to reproduce the inequality trend (GKS), first decreasing than increasing (even if the levels

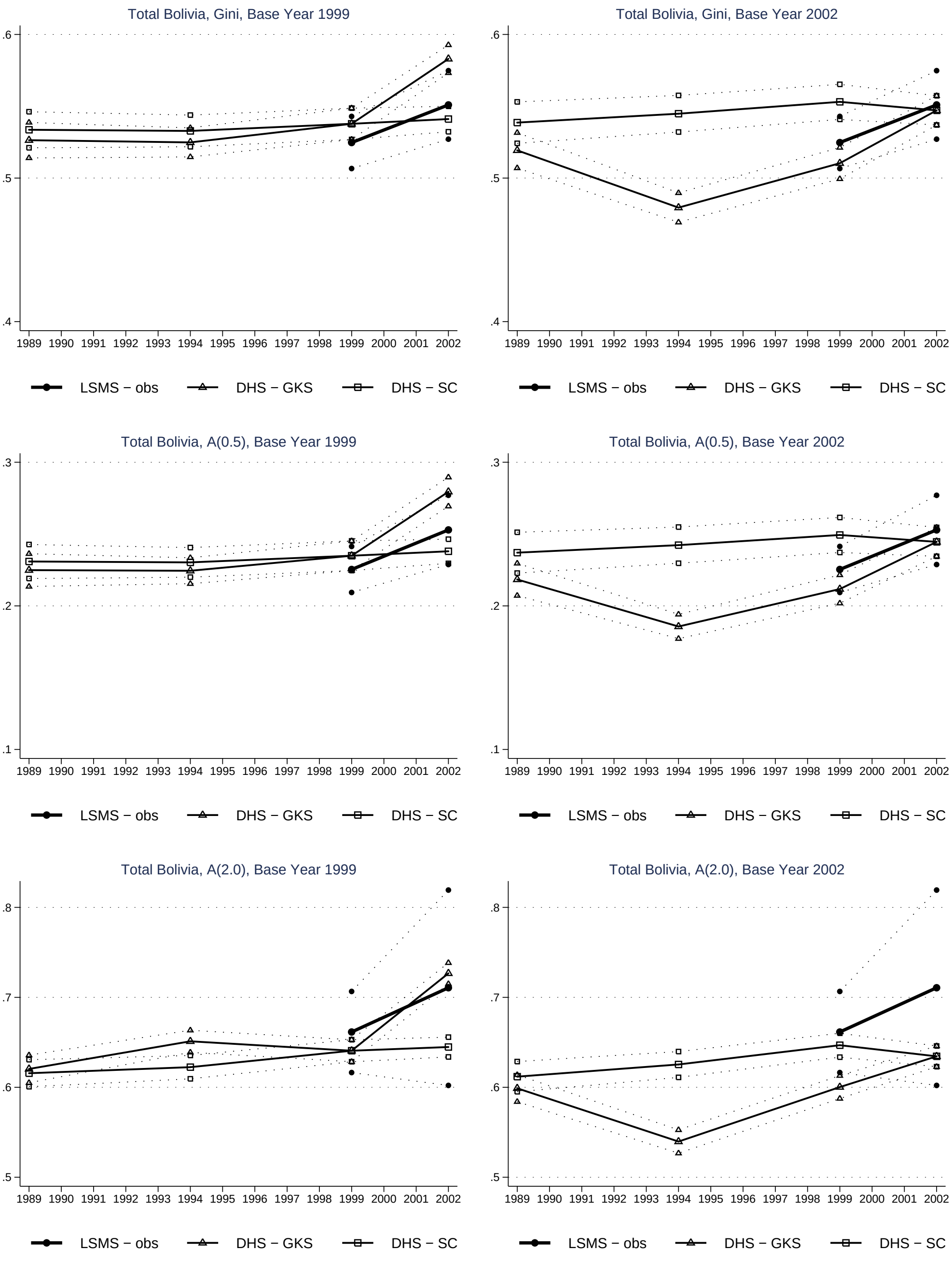
[21] For extreme poverty (Appendix Figure 10), the differences using 1999 as base year are smaller using GKS for all poverty measures, but SC fails to reproduce the increase in poverty from 1999 to 2002. The results for base year 2002 using GKS deliver better simulations within the same year (observed and simulated confidence intervals overlap) compared to moderate poverty, and again SC overestimates poverty significantly.

Figure 1: Moderate Poverty, Total Bolivia, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 2: Inequality, Total Bolivia, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

tend to be smaller than the observed ones), whereas the constant coefficients (SC) deliver a constant picture on inequality with hardly any change. This again calls for some no-constancy modeling to take dynamics into account. As for cities data for all 4 years are available, observed income suggests that inequality as measured by Gini and A(0.5) decreased between 1989 and 1994, remain stable until 1999 and then had an important increase until 2002. This is consistent with the fact that Bolivia experienced a macroeconomic crisis in 1999 which was related to a pronounced deterioration of terms of trade and with the Brazilian devaluation. For A(2.0) inequality increases already from 1994 onwards.

Concerning moderate poverty in towns (Figure 5), observed income points to a small decrease in poverty between 1999 and 2002. With 1999 as the base year, GKS shows a slight increase for P0, and a more important increase for P1 and P2. SC follows better the trend of poverty computed with observed values. Similar results are found when using 2002 as the base year. If one looks at levels, the point estimates for P0 in 2002 based on simulated income using GKS is closer to the result based on observed income than SC, but the situation is reversed when looking at P1 and P2. With 2002 as the base year, and focusing on 1999, SC is closer to the figure based on observed income than GKS for P0, P1, and P2.[22]

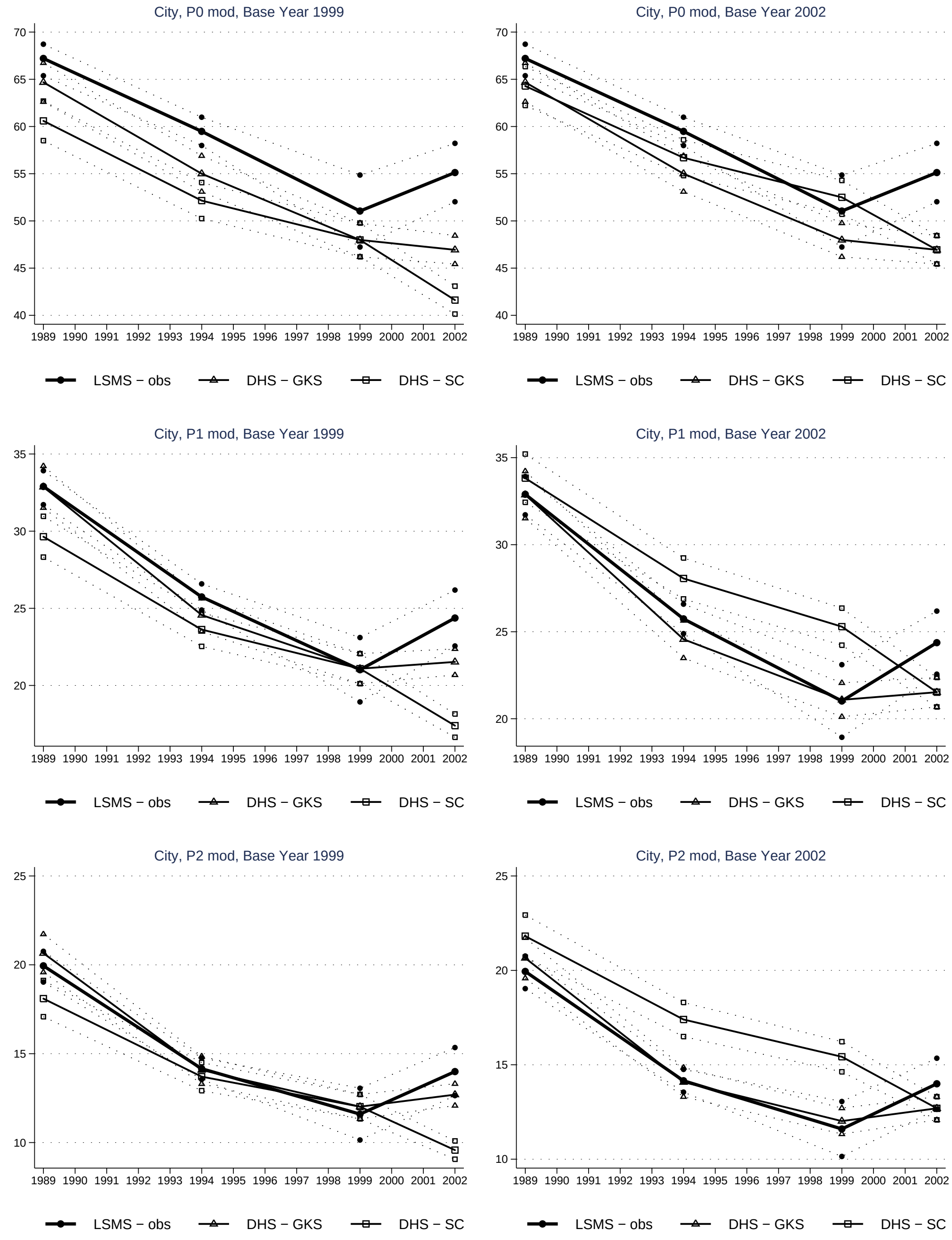
For inequality in towns (Figure 6), we focus again on the years 1999 and 2002. Gini and A(0.5) based on observed income show very little change between these two years, while A(2.0) suggests a decrease. It must be noted that the confidence intervals are especially large for A(2.0).[23] As for the case of cities, SC shows estimates that are almost unchanged between 1999 and 2002, irrespective of the base year used. GKS, on the other hand, shows a sharp increase in inequality for all measures if 1999 is used as the base year. With 2002 as the base year, GKS suggests a small increase in inequality. For towns, the overall level difference between the two base years is again most pronounced.

Moderate poverty in rural areas (Figure 7) shows an interesting picture. The levels

[22] Extreme poverty (Appendix Figure 11) shows one interesting difference. The headcount increases which is only replicated by the GKS method. Striking are also the overall level differences for the earlier years of 5 to 10 percentage points lower when using 2002 as base year which is even more relevant for extreme poverty.

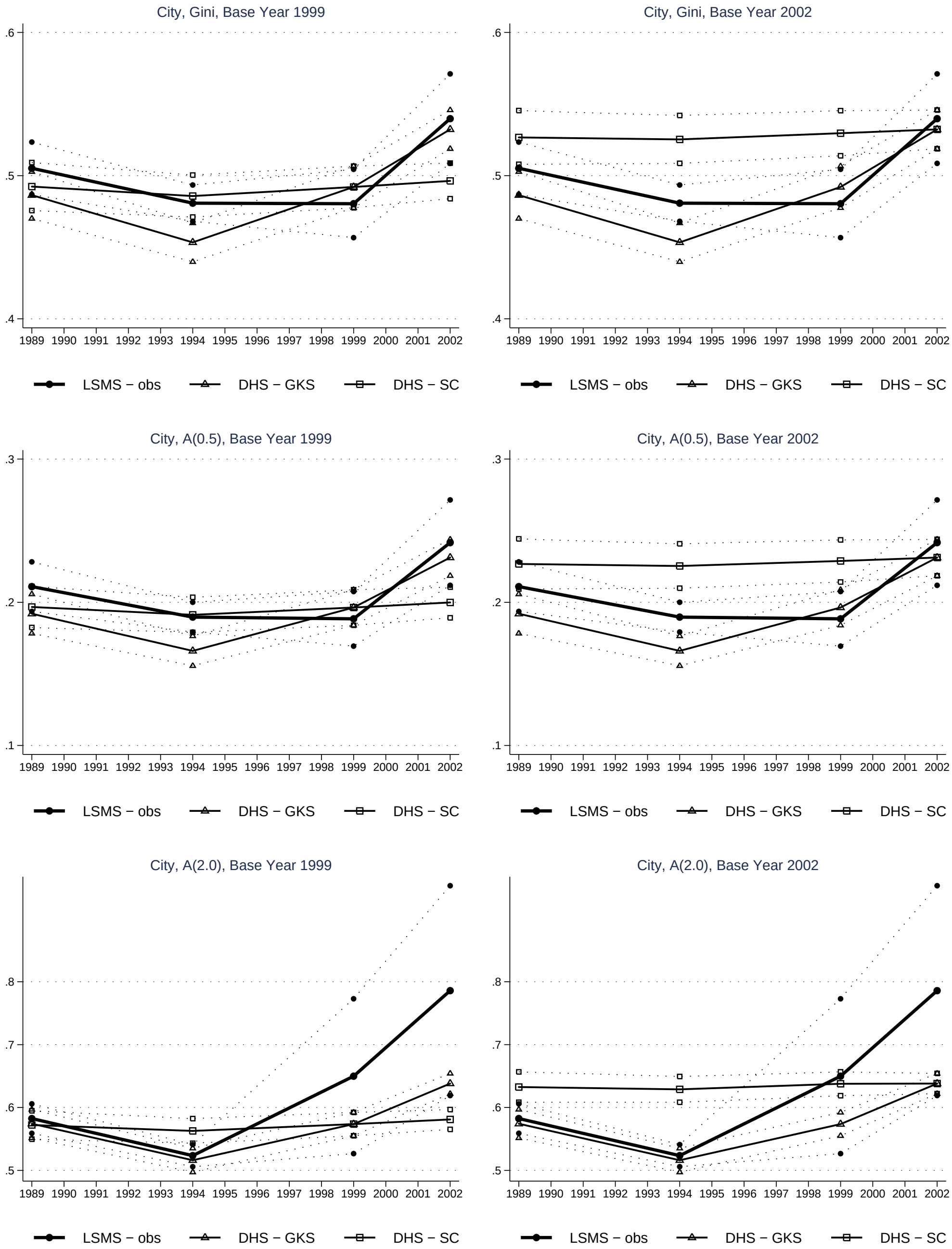
[23] Beyond the relatively small sample for towns, this could have something to do with non-linearities of the measure. As mentioned before, Atkinson (1970) inequality measures explicitly consider a constant inequality aversion parameter, which allows giving more or less emphasis to redistributions that take place at the lower end of the income distribution. A parameter value such as 2.0 gives much more importance to income transfers that make income differences smaller at the bottom of the distribution relative to those at the top of it (Jenkins, 1991).

Figure 3: Moderate Poverty, Cities, 1989–2002



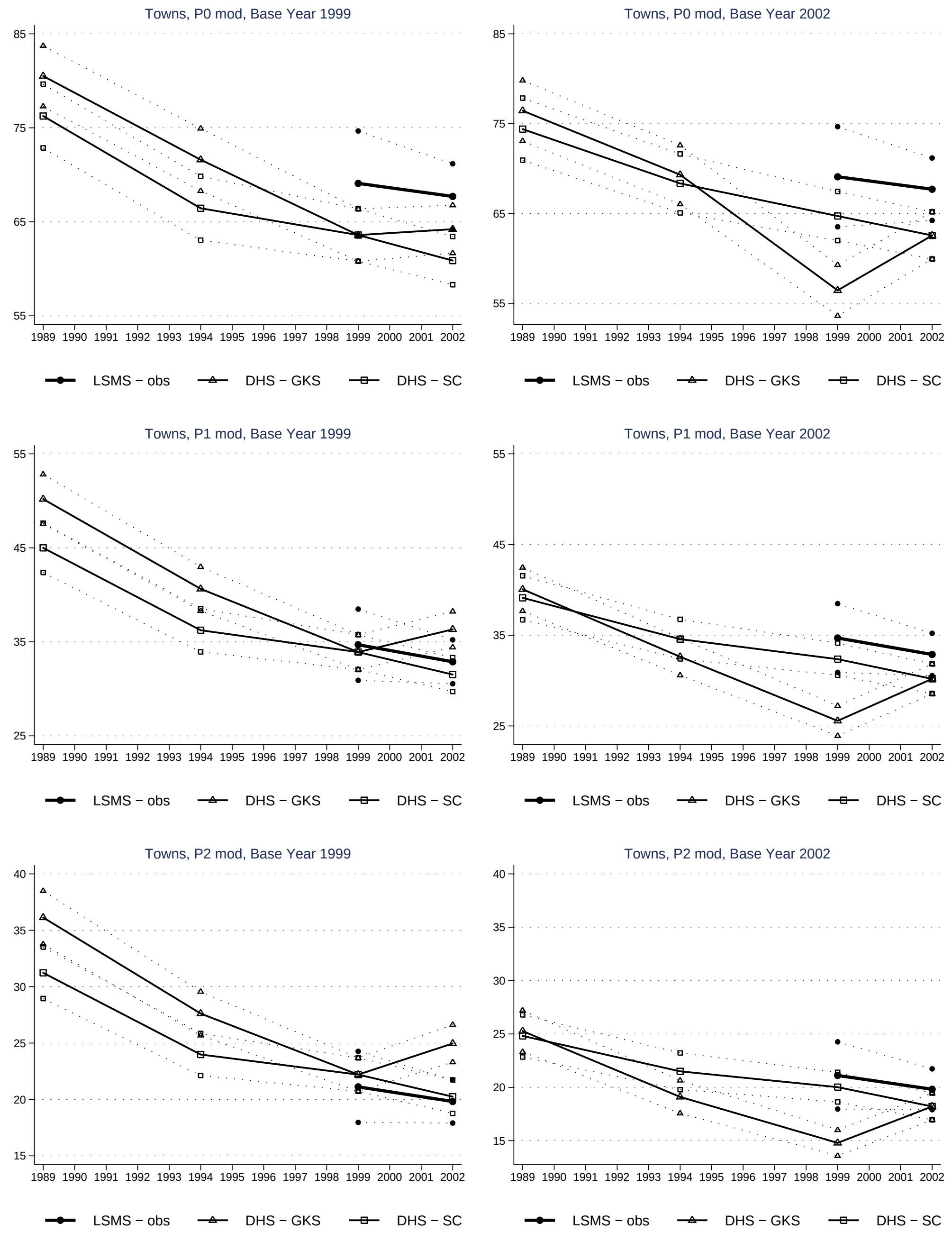
Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 4: Inequality, Cities, 1989–2002



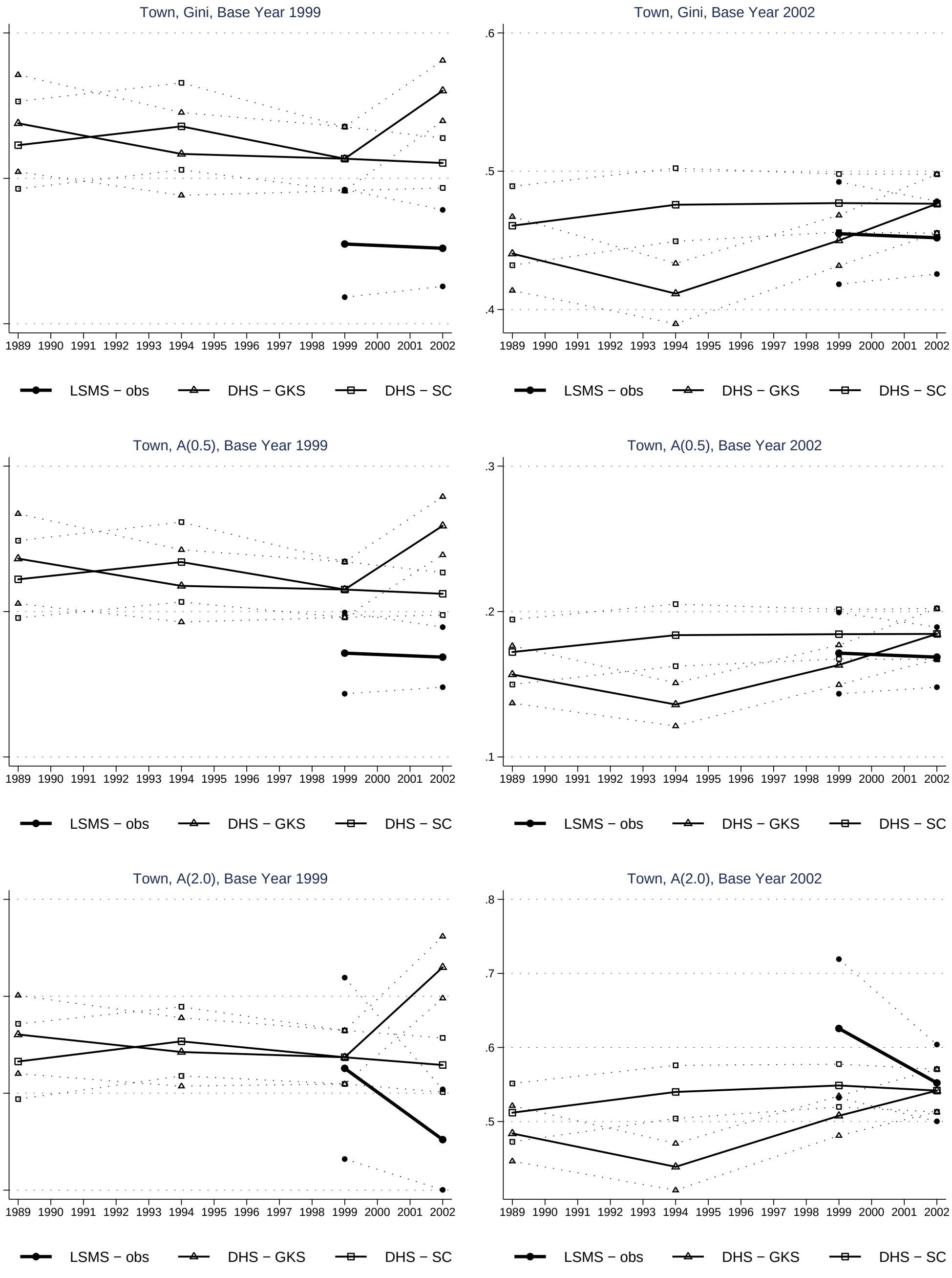
Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 5: Moderate Poverty, Towns, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 6: Inequality, Towns, 1989–2002



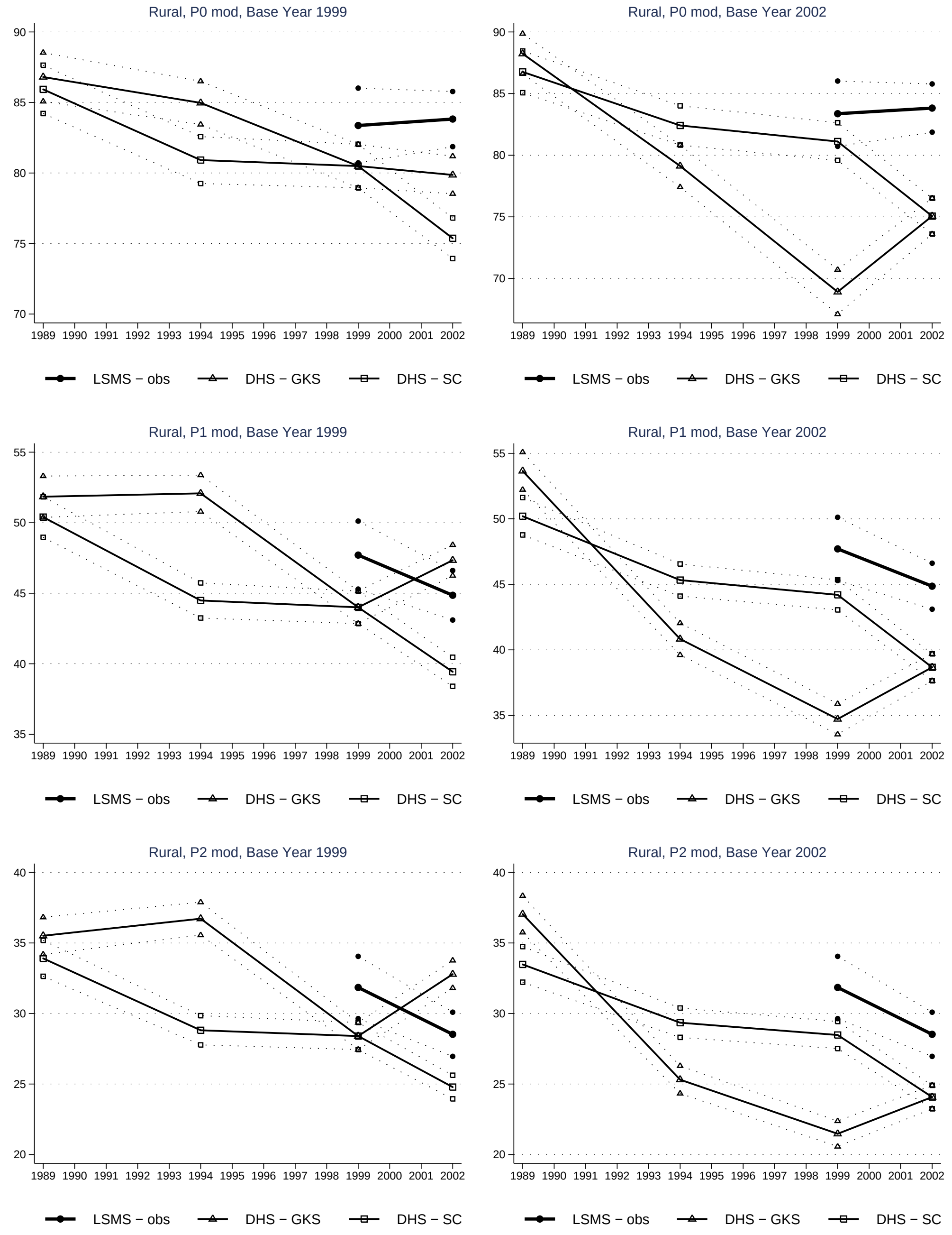
Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

of poverty are quite different between figures based on observed income and simulated income. Observed income suggests that P0 in 2002 has remained very close to its value in 1999. With 1999 as the base year, GKS also shows little changes, but SC suggests an important decrease in P0. For P1, while observed income points to a decrease, GKS suggests an important increase, while SC shows a trend more in line with observed income. Results with 2002 as the base year are different. For P0, GKS now suggests a sharp increase in poverty, while SC shows an important decrease. For P1 and P2, the downwards trend shown by observed income is replicated by SC, but not by GKS which rather suggests an important increase in poverty.[24]

Inequality measures in rural areas (Figure 8) present a similar picture as towns. Observed income shows almost no changes in inequality between 1999 and 2002, which highlights the fact that rural areas were less affected by the 1999 crisis. As before, SC suggests very stable figures for both years and they are relatively close to the measures based on observed income. GKS with base year 1999 shows a very different picture in 2002 as it points to an important increase in inequality, with levels well above the confidence interval for observed income. With 2002 as the base year results are closer to those with observed income, even if the levels remain higher than the observed ones.

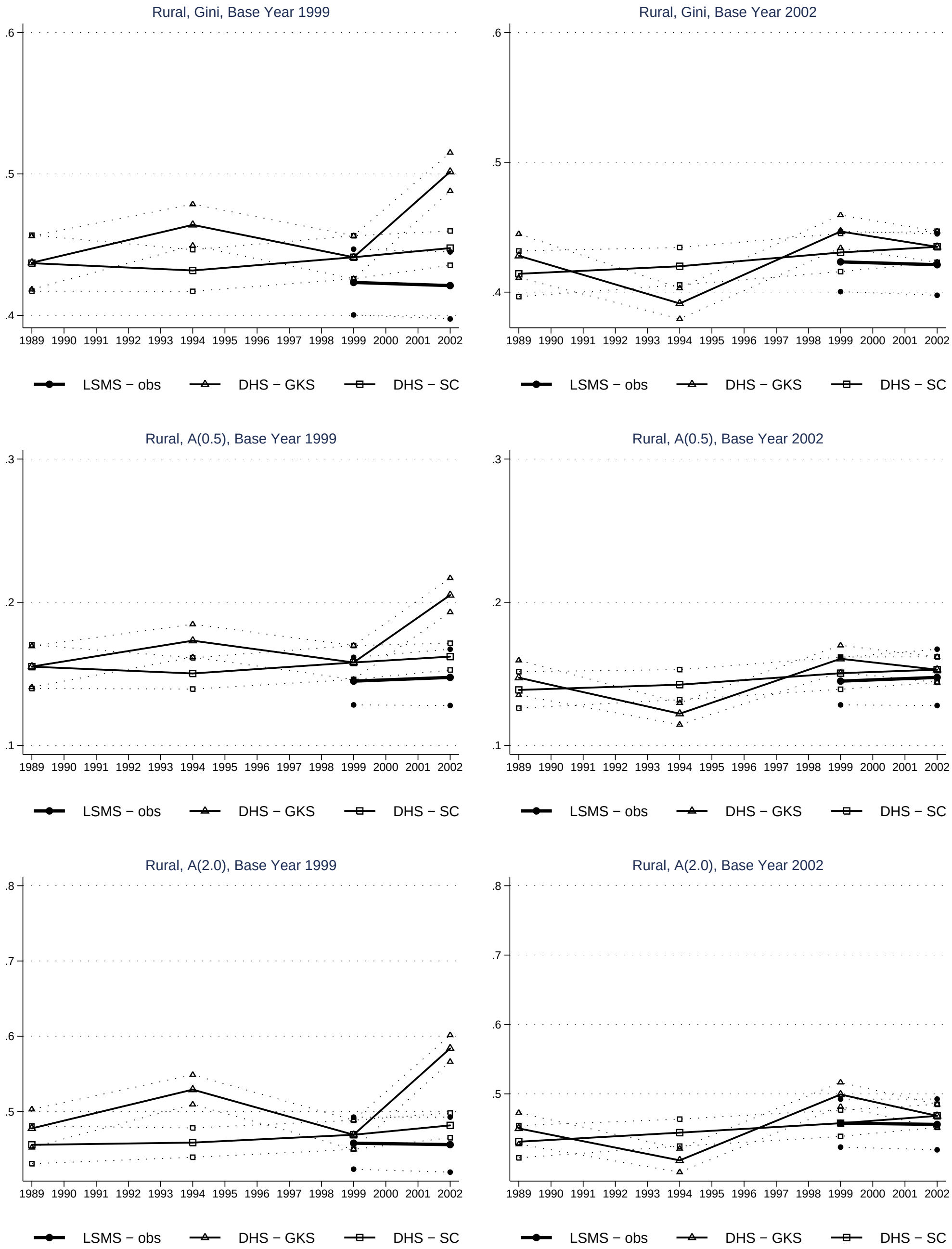
[24] The results for extreme poverty show the same picture (Appendix Figure 12).

Figure 7: Moderate Poverty, Rural Areas, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 8: Inequality, Rural Areas, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own alculations based on ECH, EIH, and DHS.

4 Discussion and Outlook

It is certainly difficult to interpret the differences of using both approaches, i.e., SC and GKS, as even using measures based on observed income does not always show a clear picture of the evolution of poverty and inequality in towns and rural areas in Bolivia between 1999 and 2002. The confidence intervals for the measures are relatively large, e.g., P0 (moderate poverty) based on observed income in rural areas is estimated to be between 81 percent and 86 percent in 1999.[25]

The relevant regions for out-of-sample predictions are towns and rural areas for which one has to compare the measures based on simulated income with those based on observed income. We have systematically compared the performance of GKS and SC in Table 4 for poverty and inequality. The idea is to check whether potential problems arise comparing the "true value", i.e., the one computed using observed income, with the values based on simulated income following either GKS or SC. The table shows (i) a simple judgment on the over-/underestimation of the true values, i.e., to see if the estimates are systematically or randomly above or below the observed values; (ii) whether the estimated numbers lie outside the 95 percent confidence interval of the true value; (iii) whether confidence intervals fail to overlap; and (iv) whether the simulated trend (between 1999 and 2002) is different than the one computed with true values. In general, the latter two problems are the most relevant ones.

[25] For the data used here, it is important to highlight that for both base years used (1999 and 2002), there is a one year lag between the LSMS, which are used for estimating the models, and the DHS, which are used to simulate income for towns and rural areas. This means that the previous consideration about the stability of the results from a model does not only apply for two years, but for four years (1998, 1999, 2002, and 2003). Furthermore, the period itself of the late 1990s and early 2000s is characterized by significant ups and in the economic performance.

Table 4: Observed and Simulated Poverty and Inequality Levels and Trends, 1999–2002

	Total				City				Town				Rural			
	1999		2002		1999		2002		1999		2002		1999		2002	
	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC
Moderate poverty																
PO																
Overestimate (level)						x										
Underestimate (level)	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x
Estimates are NOT in 95% CI	x		x	x			x	x	x			x	x		x	x
CI do NOT overlap	x		x	x			x	x	x			x	x		x	x
Different trend from observed		x		x	x	x	x	x	x		x			x	x	x
P1																
Overestimate (level)			x			x					x				x	
Underestimate (level)	x	x		x			x	x	x	x		x	x	x		x
Estimates are NOT in 95% CI	x			x		x	x	x	x		x		x	x	x	x
CI do NOT overlap	x			x		x	x	x	x				x		x	x
Different trend from observed		x		x		x		x	x		x		x		x	
P2																
Overestimate (level)		x	x		x	x					x	x			x	
Underestimate (level)	x			x			x	x	x	x			x	x		x
Estimates are NOT in 95% CI	x		x	x		x		x	x		x		x	x	x	x
CI do NOT overlap	x		x	x		x		x	x		x		x	x	x	x
Different trend from observed	x		x			x		x	x		x		x		x	
Extreme poverty																
PO																
Overestimate (level)			x			x				x	x				x	
Underestimate (level)	x			x	x		x	x	x			x	x	x		x
Estimates are NOT in 95% CI	x			x		x	x	x	x		x		x	x		x
CI do NOT overlap	x			x				x					x			x
Different trend from observed		x		x		x		x		x		x	x		x	
P1																
Overestimate (level)			x			x				x	x	x			x	
Underestimate (level)	x			x			x	x	x				x	x		x
Estimates are NOT in 95% CI	x		x	x		x		x	x		x	x	x	x	x	x
CI do NOT overlap	x		x	x		x		x	x		x		x	x	x	x
Different trend from observed	x		x			x		x	x		x		x		x	

continued on next page

Table 4 continued

	Total				City				Town				Rural			
	1999		2002		1999		2002		1999		2002		1999		2002	
	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC	GKS	SC
P2																
Overestimate (level)			x			x					x	x			x	
Underestimate (level)	x	x		x			x	x	x	x			x	x		x
Estimates are NOT in 95% CI	x		x	x		x		x	x		x	x	x	x	x	x
CI do NOT overlap	x		x	x		x		x			x		x	x	x	x
Different trend from observed	x		x			x		x	x		x		x		x	
Inequality																
Gini																
Overestimate (level)		x	x		x	x				x	x	x	x	x	x	x
Underestimate (level)	x			x			x	x	x							
Estimates are NOT in 95% CI		x	x			x		x			x	x			x	x
CI do NOT overlap						x					x	x			x	
Different trend from observed		x		x		x		x	x		x			x	x	x
A(0.5)																
Overestimate (level)		x	x		x	x				x	x	x	x	x	x	x
Underestimate (level)	x			x			x	x	x							
Estimates are NOT in 95% CI		x	x			x		x			x	x			x	
CI do NOT overlap						x					x	x			x	
Different trend from observed		x		x		x		x	x		x		x			
A(2.0)																
Overestimate (level)			x								x	x	x		x	x
Underestimate (level)	x	x		x	x	x	x	x	x	x						
Estimates are NOT in 95% CI	x							x	x		x	x	x		x	
CI do NOT overlap	x							x			x	x			x	
Different trend from observed		x		x		x		x	x		x			x	x	x

Notes: For explanation, see subsection 3.4.
Source: Own calculations based on ECH and DHS.

For our data, it is difficult to come to an overall judgement on wether SC performs better than GKS because the results differ for poverty and inequality measures as well as for income. In general, both methods do not yield very good results. For example, P0 is nearly always underestimated. For towns and rural areas, SC gives slightly better results. However, this does not hold for cities, where clearly GKS outperforms SC, and total Bolivia, where results are mixed. Similar results were obtained when looking at income in subsection 3.3.

It can be argued that depending on the changes in regression coefficients, error terms, and endowments, one or the other method performs better. First, if the regression model is very stable from period to period, the assumption of SC of using the coefficients and error terms obtained from one period and apply them for another one should not face many problems. However, if this is not the case, it is possible that a more flexible approach, such as assuming any kind of dynamics as GKS may yield better results, assuming that the changes between periods are adequately modeled. If this is not the case, then GKS does not yield good results for towns and rural areas. Table 5 systematically compares which assumption for 1999 comes closer in terms of distance to the observed coefficients. The column "closer" indicates which estimated coefficient, β_{GKS} or β_{SC} following the Equations 2 and 3, respectively, comes closer to the "true" β_{99}. Of the total of 72 coefficients for towns and rural areas, the GKS coefficients are closer in 39 cases and the SC coefficients in 33 cases.

Table 5: Regression Coefficients Using two Different Assumptions

	City		Town				Rural			
	β_{99}	β_{02}	β_{99}	β_{GKS}	β_{SC}	closer	β_{99}	β_{GKS}	β_{SC}	closer
La Paz	-0.04	-0.03	0.16	0.18	0.18	GKS	0.25	0.27	0.26	SC
Cochabamba	0.24	0.17	0.75	0.68	0.17	GKS	0.39	0.33	0.16	GKS
Oruro	-0.06	-0.23	-0.18	-0.35	-0.11	SC	0.35	0.18	0.23	SC
Potosi	-0.07	-0.08	0.18	0.17	-0.09	GKS	0.05	0.04	-0.08	GKS
Tarija	0.50	0.16	0.57	0.23	0.40	SC	0.71	0.37	0.56	SC
Santa Cruz	0.70	0.44	0.58	0.32	0.11	GKS	0.71	0.45	0.46	SC
Beni & Pando	0.62	0.31	0.29	-0.02	0.25	SC	0.74	0.44	0.59	SC
elderly dependency ratio	-0.20	-0.32	-0.18	-0.30	-0.24	SC	-0.03	-0.15	-0.08	SC
child dependency ratio	0.17	0.00	0.55	0.38	-0.01	GKS	-0.12	-0.29	0.05	GKS
hh size	-0.08	-0.05	-0.06	-0.03	-0.05	SC	-0.09	-0.05	-0.10	SC
hh head age	0.00	0.02	0.01	0.03	0.01	SC	0.02	0.03	0.03	GKS
hh head age squared	0.00	0.00	0.00	0.00	0.00	SC	0.00	0.00	0.00	SC
gender hh head	-0.05	0.03	0.22	0.30	0.06	GKS	-0.01	0.07	-0.12	GKS
access to public water	-0.08	-0.05	0.05	0.07	0.06	SC	0.00	0.03	0.13	GKS
has no toilet	-0.03	-0.04	-0.27	-0.28	0.06	GKS	-0.23	-0.25	-0.15	GKS
no partner in household	0.21	0.34	0.47	0.60	0.25	GKS	0.39	0.53	0.06	GKS
com. basic edu. (m.)	-0.16	-0.04	-0.02	0.09	0.05	SC	-0.02	0.09	0.15	GKS
incom. secondary edu. (m.)	-0.01	-0.11	-0.17	-0.27	0.07	GKS	-0.01	-0.11	0.11	GKS
com. secondary edu. (m.)	-0.05	0.06	0.29	0.39	0.07	GKS	0.05	0.16	0.19	GKS
tertiary edu. (m.)	0.39	0.43	-0.02	0.03	0.29	GKS	0.44	0.49	0.27	GKS
com. basic edu. (w.)	0.12	-0.07	-0.01	-0.20	-0.01	SC	0.26	0.06	0.22	SC
incom. secondary edu. (w.)	0.10	0.01	0.11	0.03	0.05	SC	0.27	0.18	0.26	SC
com. secondary edu. (w.)	0.24	0.02	0.12	-0.10	0.31	SC	0.44	0.22	0.33	SC
tertiary edu. (w.)	0.53	0.44	0.45	0.36	0.58	GKS	0.75	0.66	0.64	GKS

continued on next page

Table 5 continued

	City		Town				Rural			
	β_{99}	β_{02}	β_{99}	β_{GKS}	β_{SC}	closer	β_{99}	β_{GKS}	β_{SC}	closer
high & med. skilled white collar (m.)	0.53	0.62	1.03	1.12	0.61	GKS	0.67	0.76	0.21	GKS
skilled & unskilled manual (m.)	0.23	0.23	0.57	0.57	0.41	GKS	0.60	0.60	0.04	GKS
agriculture (m.)	-0.21	0.21	0.19	0.61	0.30	SC	0.25	0.67	-0.05	SC
sales and services (m.)	0.34	0.53	0.87	1.06	0.62	GKS	0.67	0.86	0.26	GKS
high & med. skilled white collar (w.)	0.38	0.55	0.73	0.90	0.72	SC	0.13	0.30	0.43	GKS
skilled & unskilled manual (w.)	0.14	0.19	0.51	0.57	0.17	GKS	-0.12	-0.07	0.05	GKS
agriculture (w.)	0.64	-0.14	-0.24	-1.02	-0.54	SC	-0.05	-0.83	-0.08	SC
sales and services (w.)	0.31	0.23	0.71	0.63	0.35	GKS	0.36	0.28	0.25	GKS
birth in last 12 month	0.20	-0.06	-0.29	-0.55	-0.17	SC	-0.12	-0.38	-0.03	SC
attended by doctor	-0.22	-0.02	0.68	0.88	0.10	GKS	0.20	0.40	0.02	SC
delivered in hospital	-0.10	-0.26	-0.38	-0.54	0.12	GKS	0.08	-0.08	0.15	SC
c/t/r dummy/constant	5.05	4.80	3.84	3.59	4.41	GKS	4.09	3.84	4.23	SC

Notes: See text for details.
Source: Own calculations based on ECH and EIH.

Second, as SC by definition always use the same coefficients and error terms, all changes in poverty and inequality measures between years are explained by changes in endowments, which yields as a general result that if endowments do not change much SC will provide measures that are rather stable over time. If true values (of income, poverty, and inequality) are not changing too much over time, SC will perform well. This is why nearly all poverty graphs show a monotonic (downward) trend and the inequality graphs nearly no trend for the SC case. As was mentioned before, Bolivia experienced an important crisis in 1999, where per capita GDP decreased almost by 2 percent after several years of positive growth. After the crisis, the economy recovered slowly, and it was only in 2004 that growth of per capita GDP was again larger than 1 percent. Therefore, one could expect that the four surveys considered in our study (1998, 1999, 2002, and 2003) depict rather different economic situations.

Coming back to the question of how to model dynamics, the constancy-of-differences assumption in Equation (2) can alternatively be relaxed following Grosse et al. (2009) and rearranged to :

$$(\beta_{t-1}^{T} - \beta_{t-1}^{C}) = \phi(\beta_{t}^{T} - \beta_{t}^{C}) \quad \text{and} \quad (\beta_{t-1}^{R} - \beta_{t-1}^{C}) = \phi(\beta_{t}^{R} - \beta_{t}^{C}) \tag{6}$$

where ϕ can be understood as a kind of "mobility parameter" that measures if the coefficients estimated separately for urban and non-urban areas become more similar towards each other or not. We present the evolution of ϕ to gain insights of the mobility of parameters over time in subsection 3.1.[26]

Table 6 gives a first insight on this question. For example, it calculates ϕ of Equation (6) for towns and rural areas which reveals that it is neither constant nor of the same sign or magnitude for each coefficient. In addition, it reveals that coefficients can be of different magnitude and even sign (exemplarily shown for the coefficient for cities in the last column). Of the 36 coefficients, one-third is not even stable in sign for cities. Of the remaining, many change considerably in terms of magnitude. What also becomes evident is, however, that a constant ϕ can also not be confirmed.

Additionally worth noting is that results would have changed if we had followed the way Stifel and Christiaensen (2007) deal with the issue of underestimating poverty (by shifting the poverty line until observed and simulated poverty levels coincide). In this case the picture would look different, as the level of simulated real income would change

[26]However, we do not show the results of the whole estimation procedure on poverty and inequality using this assumption.

in order to match observed levels. Such a modification would have changed the results for, for example, moderate poverty (P0) for total Bolivia (Figure 1) in that the GKS assumption would have nearly exactly coincided in level and trend for both base years whereas the level for 1999 (using 2002 as base year) for SC, that without shifting comes close to observed values, would overestimate P0 clearly. We suspect that the level results are driven by the regional dummy (i.e., the regional constant). Stifel and Christiaensen (2007) for example have a very large constant that is taken back and forth in time. The share of the remaining few coefficients (e.g., only 3 for the Nairobi sample) is rather negligible, besides being selected to ensure stability in themselves. In selecting variables, the authors explicitly use the ones that are expected to remain stable over time and not respond to economic conditions or policy changes.[27] One way of dealing with such problems is suggested by Grosse et al. (2009) and consists in shifting real per capita mean income (both observed and simulated) to levels observed by national accounts.[28] This data is available for all countries (sometimes even for regional disaggregation) and can serve as a kind of neutral anchor for the level.

[27]The suggestion of Mathiassen (2008) to update coefficients seems worth trying in this regard.

[28]Note that, in general, shifting per capita mean income, shifting the poverty line, or changing the intercept of the regression are all equivalent transformations. The only difference could be availability (for example, having two or more different poverty lines) or disaggregation (for example, having national account data at the departmental level)

Table 6: Stability of Regression Coefficients over Time

	City				Town		Rural		ϕ		$\lvert\beta\rvert$ City
	β_{89}	β_{94}	β_{99}	β_{02}	β_{99}	β_{02}	β_{99}	β_{02}	Town	Rural	Constant?
La Paz	0.01	0.15	-0.04	-0.03	0.16	0.18	0.25	0.26	0.97	1.04	0
Cochabamba	0.16	0.13	0.24	0.17	0.75	0.17	0.39	0.16	-222.34	-13.85	1
Oruro	-0.17	-0.20	-0.06	-0.23	-0.18	-0.11	0.35	0.23	-0.94	0.89	1
Potosi	-0.26	-0.21	-0.07	-0.08	0.18	-0.09	0.05	-0.08	-22.22	-113.91	1
Tarija	-0.03	0.03	0.50	0.16	0.57	0.40	0.71	0.56	0.29	0.53	0
Santa Cruz	0.43	0.43	0.70	0.44	0.58	0.11	0.71	0.46	0.35	0.65	1
Beni & Pando	0.44	0.28	0.62	0.31	0.29	0.25	0.74	0.59	5.56	0.46	1
elderly dependency ratio	-0.23	-0.28	-0.20	-0.32	-0.18	-0.24	-0.03	-0.08	0.25	0.68	1
child dependency ratio	0.08	-0.08	0.17	0.00	0.55	-0.01	-0.12	0.05	-31.02	-5.39	0
hh size	-0.07	-0.05	-0.08	-0.05	-0.06	-0.05	-0.09	-0.10	12.26	0.07	1
hh head age	0.03	0.01	0.00	0.02	0.01	0.01	0.02	0.03	-1.30	0.84	1
hh head age squared	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-2.30	1.32	0
gender hh head	-0.12	-0.12	-0.05	0.03	0.22	0.06	-0.01	-0.12	8.60	-0.25	0
access to public water	0.15	0.03	-0.08	-0.05	0.05	0.06	0.00	0.13	1.15	0.42	0
has no toilet	-0.20	-0.21	-0.03	-0.04	-0.27	0.06	-0.23	-0.15	-2.41	1.94	1
no partner in household	0.35	0.58	0.21	0.34	0.47	0.25	0.39	0.06	-2.77	-0.64	1
com. basic edu. (m.)	0.02	0.03	-0.16	-0.04	-0.02	0.05	-0.02	0.15	1.45	0.70	0
incom. secondary edu. (m.)	0.02	0.05	-0.01	-0.11	-0.17	0.07	-0.01	0.11	-0.94	-0.02	0
com. secondary edu. (m.)	0.10	0.10	-0.05	0.06	0.29	0.07	0.05	0.19	39.21	0.77	0
tertiary edu. (m.)	0.52	0.40	0.39	0.43	-0.02	0.29	0.44	0.27	2.83	-0.35	1
com. basic edu. (w.)	0.00	0.07	0.12	-0.07	-0.01	-0.01	0.26	0.22	-1.94	0.47	0
incom. secondary edu. (w.)	0.14	0.01	0.10	0.01	0.11	0.05	0.27	0.26	0.40	0.70	1
com. secondary edu. (w.)	0.17	0.06	0.24	0.02	0.12	0.31	0.44	0.33	-0.43	0.63	1
tertiary edu. (w.)	0.39	0.40	0.53	0.44	0.45	0.58	0.75	0.64	-0.55	1.09	1

continued on next page

Table 6 continued

	City				Town		Rural		ϕ		$\lvert\beta\rvert$ City
	β_{89}	β_{94}	β_{99}	β_{02}	β_{99}	β_{02}	β_{99}	β_{02}	Town	Rural	Constant?
high & med. skilled white collar (m.)	0.45	0.79	0.53	0.62	1.03	0.61	0.67	0.21	-77.30	-0.33	1
skilled & unskilled manual (m.)	0.37	0.47	0.23	0.23	0.57	0.41	0.60	0.04	1.89	-1.98	1
agriculture (m.)	0.42	0.51	-0.21	0.21	0.19	0.30	0.25	-0.05	4.63	-1.74	0
sales and services (m.)	0.42	0.57	0.34	0.53	0.87	0.62	0.67	0.26	5.69	-1.20	1
high & med. skilled white collar (w.)	0.45	0.46	0.38	0.55	0.73	0.72	0.13	0.43	2.00	2.13	1
skilled & unskilled manual (w.)	0.22	0.28	0.14	0.19	0.51	0.17	-0.12	0.05	-15.70	1.81	1
agriculture (w.)	0.52	0.10	0.64	-0.14	-0.24	-0.54	-0.05	-0.08	2.18	-10.88	0
sales and services (w.)	0.34	0.30	0.31	0.23	0.71	0.35	0.36	0.25	3.37	4.07	1
birth in last 12 month	-0.13	-0.13	0.20	-0.06	-0.29	-0.17	-0.12	-0.03	4.53	-10.13	0
attended by doctor	0.07	0.04	-0.22	-0.02	0.68	0.10	0.20	0.02	7.21	9.51	0
delivered in hospital	0.03	0.00	-0.10	-0.26	-0.38	0.12	0.08	0.15	-0.72	0.44	0
c/t/r dummy/constant	4.31	4.66	5.05	4.80	3.84	4.41	4.09	4.23	3.14	1.68	1

Notes: See text for details.
Source: Own calculations based on ECH and EIH.

5 Conclusion

This paper aims at estimating the stability of dynamic poverty mapping approaches. Since the poverty mapping approach was established to generate data by regression-based cross-survey mapping where otherwise no other data would have been available, the results can generally not be compared to true data. With the data for Bolivia, we were able to undertake out of sample predictions and compare simulated data with true data. This becomes extremely relevant when using the method not only in space but also over time. Our method finds that results crucially depend on the assumptions in the regressions underlying the poverty mapping. Keeping coefficients constant over time is not the advised option. How to correctly model the coefficients in a dynamic way, however, needs to be investigated in much more detail.

References

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory 2*(3), 244–263.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics 108*, 317–342.

Biewen, M. and S. P. Jenkins (2006). Variance estimation for generalized entropy and Atkinson inequality indices: The complex survey data case. *Oxford Bulletin of Economics and Statistics 68*(3), 371–383.

Binder, D. and Z. Patak (1994). Use of estimating functions for interval estimation from complex surveys. *Journal of the American Statistical Association 89*, 1035–1043.

Bourguignon, F., F. H. G. Ferreira, and N. Lustig (2005). *The microeconomics of income distribution dynamics in East Asia and Latin America.* World Bank and Oxford University Press: New York.

Branisa, B. (2006). Wealth, poverty, and inequality in Bolivia using different methods and data sets. Master's thesis, University of Göttingen, Göttingen, Germany.

Davidson, R. and E. Flachaire (2007). Asymptotic and bootstrap inference for inequality and poverty measurement. *Journal of Econometrics 141*, 141–166.

Demombynes, G., C. Elbers, J. O. Lanjouw, P. Lanjouw, J. Mistiaen, and B. Özler (2004). Producing an improved geographic profile of poverty. Methodology and evidence from three developing countries. In A. Shorrocks and R. van der Hoeven (Eds.), *Growth, Inequality and Poverty: Prospects for Pro-Poor Economic Development.* New York: Oxford University Press.

Elbers, C., J. O. Lanjouw, and P. Lanjouw (2003). Micro-level estimation of poverty and inequality. *Econometrica 71*(1), 355–364.

Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica 52*(3), 761–766.

Grimm, M. (2004). A decomposition of inequality and poverty changes in the context of macroeconomic adjustment: A microsumulation study for Côte d'Ivoire. In A. Shorrocks and R. van der Hoeven (Eds.), *Growth, Inequality, and Poverty. Prospects for Pro-Poor Economic Development.* Oxford University Press: New York.

Grosse, M., S. Klasen, and J. Spatz (2005). Creating national poverty profiles and growth incidence curves with incomplete income or consumption expenditure data: An application to Bolivia. Ibero-America Institute for Economic Research Discussion Paper 129, University of Göttingen.

Grosse, M., S. Klasen, and J. Spatz (2009). Matching household surveys with DHS data to create nationally representative time series of poverty: An application to Bolivia. Courant Research Centre 'Poverty, Equity and Growth' Discussion Paper 21, University of Göttingen.

Hentschel, J., J. O. Lanjouw, P. Lanjouw, and J. Poggi (2000). Combining census and survey data to trace the spatial dimensions of poverty: A case study of Ecuador. *World Bank Economic Review 14*(1), 147–165.

Jenkins, S. P. (1991). The measurement of income inequality. In L. Osberg (Ed.), *The Measurement of Income Inequality*, pp. 3–38. Armonk: M. E. Sharpe.

Jolliffe, D. and A. Semykina (1999). Robust standard errors for the Foster-Greer-Thorbecke class of poverty indices: SEPOV. *Stata Technical Bulletin 51*, 34–36.

Kakwani, N. (1993). Statistical inference in the measurement of poverty. *Review of Economics and Econometrics 75*(4), 632–639.

Klasen, S., M. Grosse, R. Thiele, J. Lay, J. Spatz, and M. Wiebelt (2007). Analyzing pro-poor growth in Bolivia: Addressing data gaps and modeling policy choices. In S. Klasen, M. Grimm, and A. McKay (Eds.), *Determinants of Pro-Poor Growth: Analytical Issues and Findings from Country Cases.* New York: Palgrave Macmillan.

Kovacevic, M. S. and D. A. Binder (1997). Variance estimation for measures of income inequality and polarization – The estimating equations approach. *Journal of Official Statistics 13*(1), 41–58.

Mathiassen, A. (2008). The predictive ability of poverty models. Empirical evidence from Uganda. Discussion Paper 560, Statistics Norway, Division for Development Cooperation.

Morales, J. A. (1994). Hiperinflación y políticas de estabilización. Serie Docente 15, CIEPLAN.

Morales, J. A. (2001). El modelo de desarrollo económico visto de nuevo. In O. Zegada (Ed.), *Bolivia: Modelo de Desarrollo Y Crisis*. La Paz: Offset Boliviana.

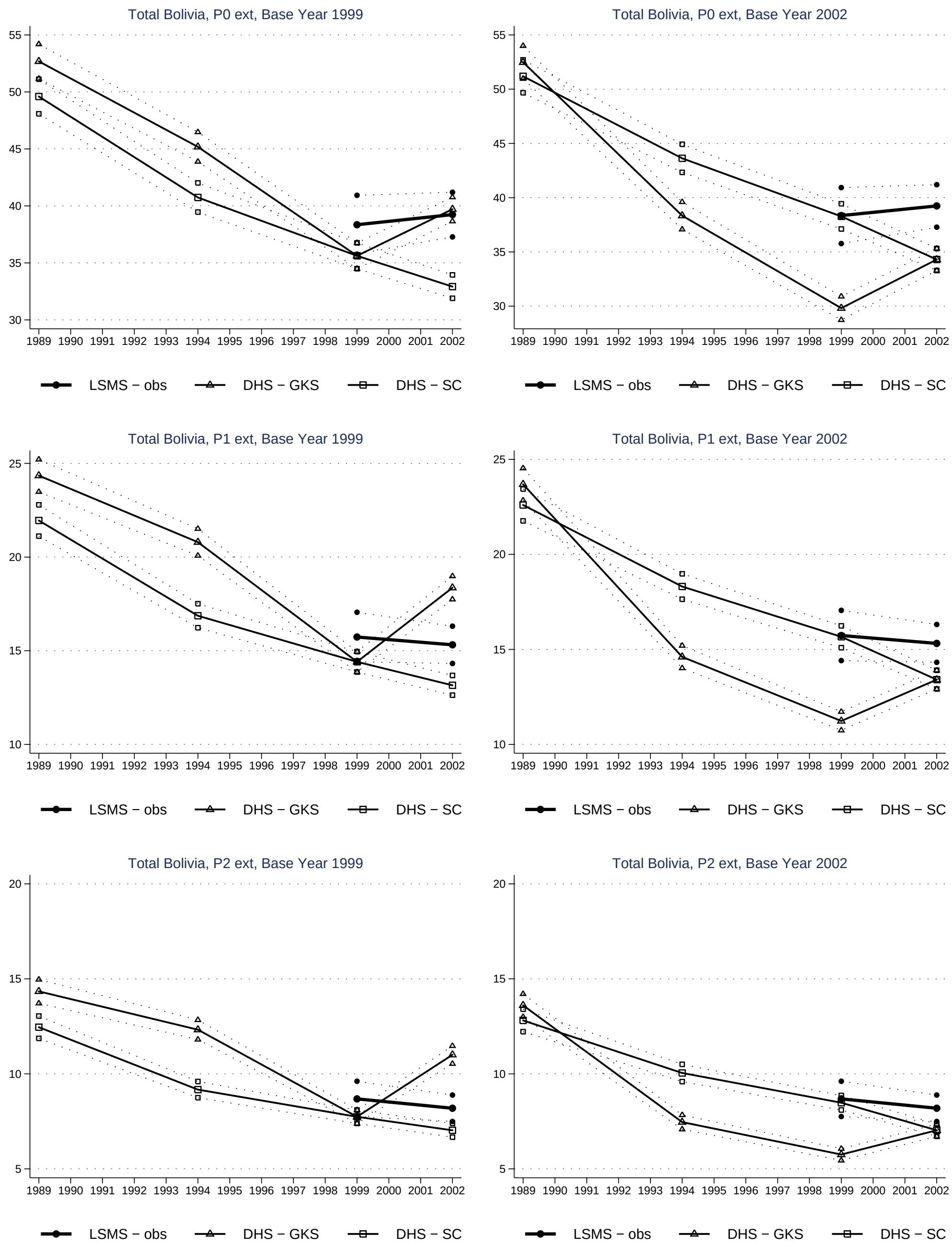
Sachs, J. and J. A. Morales (1988). *Bolivia: 1952-1986*. San Francisco.: ICS Press.

Stifel, D. and L. Christiaensen (2007). Tracking poverty over time in the absence of comparable consumption data. *World Bank Economic Review 2*(21), 317–341.

Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association 66*(334), 411–414.

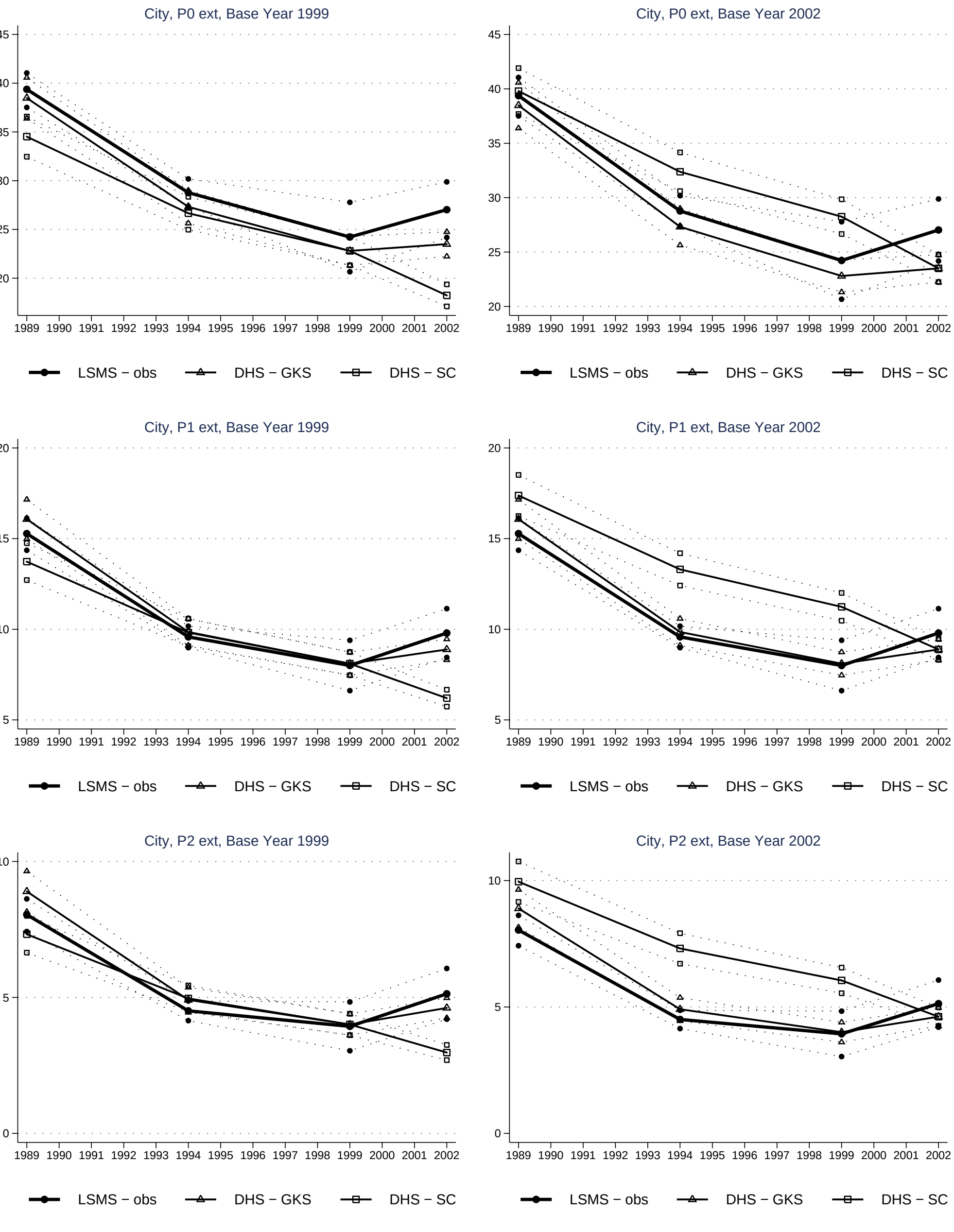
6 Appendix

Figure 9: Extreme Poverty, Total Bolivia, 1989–2002



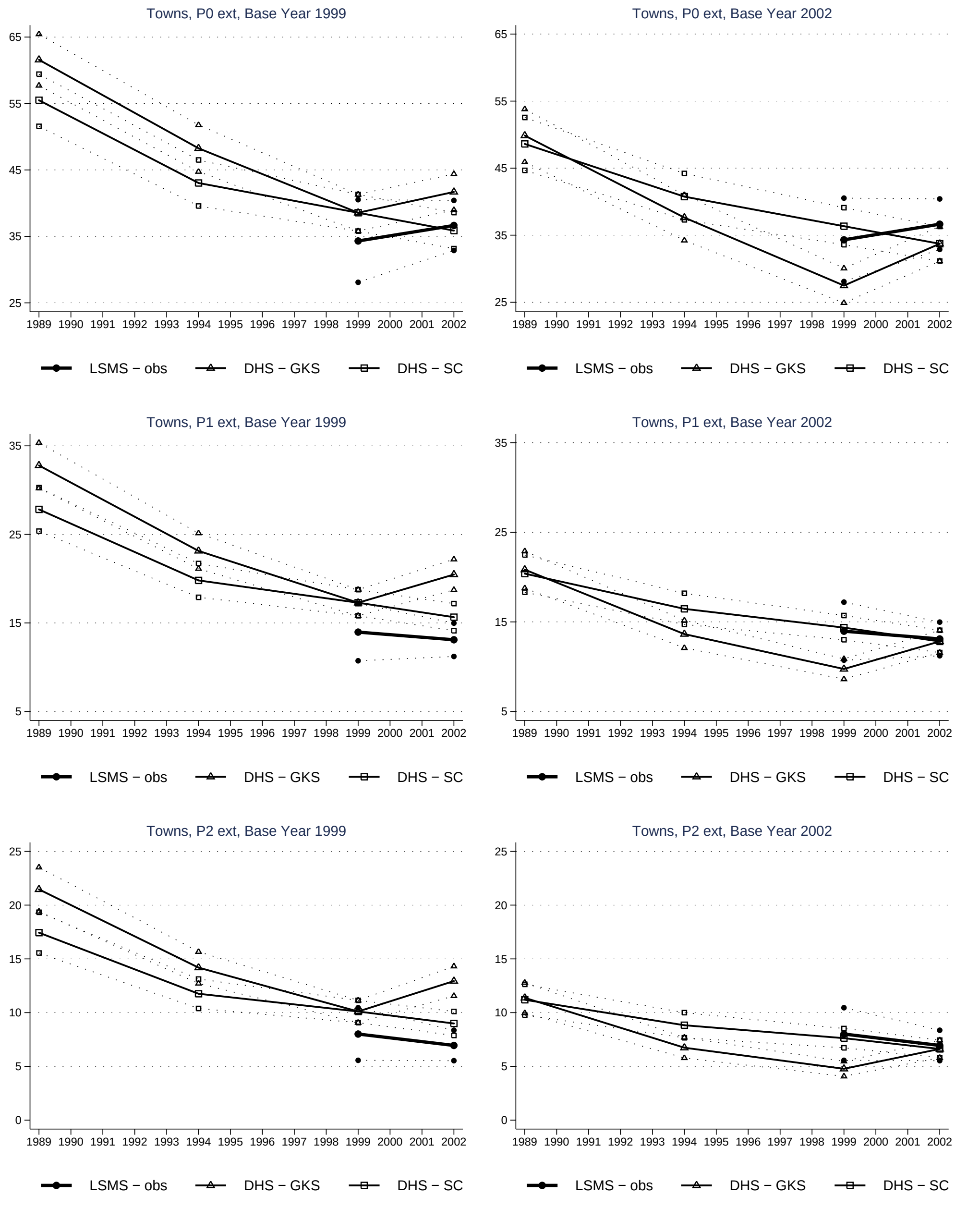
Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 10: Extreme Poverty, Cities, 1989–2002



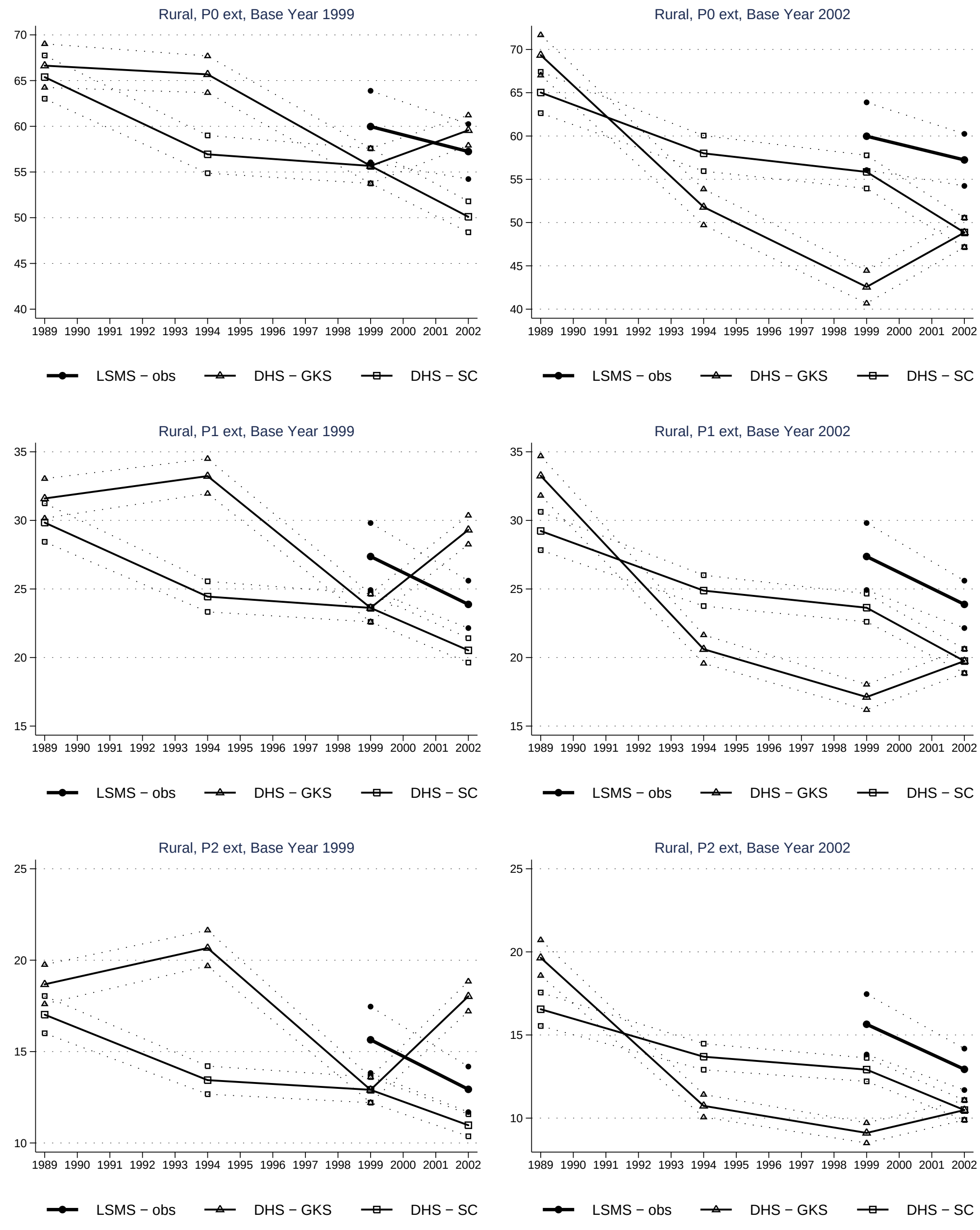
Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 11: Extreme Poverty, Towns, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.
Source: Own calculations based on ECH, EIH, and DHS.

Figure 12: Extreme Poverty, Rural Areas, 1989–2002



Notes: LSMS–obs: Data from LSMS using observed income; DHS–GKS: Data from DHS using GKS assumptions on dynamics; DHS–SC: Data from DHS using SC assumptions on dynamics. See text for details.

Source: Own calculations based on ECH, EIH, and DHS.

Table 7: Results of log-linear OLS Regression, Common Model, 1989–2002

	City								Town				Rural			
	1989		1994		1999		2002		1999		2002		1999		2002	
	β	P	β	P	β	P	β	P	β	P	β	P	β	P	β	P
La Paz	0.01	0.89	0.14	0.00	-0.02	0.88	0.00	0.98	0.11	0.81	0.23	0.17	0.25	0.00	0.23	0.00
Cochabamba	0.16	0.00	0.13	0.00	0.27	0.02	0.18	0.06	0.65	0.17	0.19	0.22	0.40	0.00	0.13	0.07
Oruro	-0.17	0.00	-0.18	0.00	-0.06	0.66	-0.22	0.05	-0.27	0.61	-0.06	0.75	0.27	0.08	0.18	0.04
Potosi	-0.25	0.00	-0.20	0.00	-0.02	0.89	-0.04	0.70	0.16	0.74	-0.02	0.92	0.06	0.57	-0.09	0.31
Tarija	-0.02	0.61	0.03	0.52	0.53	0.00	0.18	0.08	0.47	0.33	0.48	0.00	0.63	0.00	0.53	0.00
Santa Cruz	0.42	0.00	0.43	0.00	0.70	0.00	0.48	0.00	0.52	0.27	0.17	0.28	0.67	0.00	0.44	0.00
Beni & Pando	0.43	0.00	0.26	0.00	0.63	0.00	0.29	0.01	0.18	0.72	0.32	0.05	0.71	0.00	0.52	0.00
# elderly	0.01	0.84	0.04	0.29	0.11	0.41	0.07	0.58	0.09	0.73	0.09	0.39	-0.11	0.27	-0.01	0.89
# males	-0.06	0.00	-0.03	0.05	-0.07	0.03	-0.09	0.01	0.08	0.33	0.00	0.91	-0.08	0.05	-0.11	0.00
# females	-0.02	0.21	-0.05	0.00	-0.12	0.00	-0.03	0.31	-0.11	0.07	-0.01	0.80	-0.15	0.00	-0.21	0.00
# youngsters	-0.09	0.00	-0.04	0.00	-0.01	0.76	0.00	0.96	-0.08	0.16	-0.10	0.02	-0.02	0.55	-0.01	0.79
# children	-0.08	0.00	-0.05	0.00	-0.10	0.21	-0.12	0.08	-0.20	0.02	-0.08	0.08	-0.10	0.05	-0.09	0.01
# of working age / # all	0.70	0.00	1.06	0.00	1.13	0.00	1.35	0.00	0.14	0.80	0.65	0.04	0.60	0.04	1.05	0.00
gender hh head	-0.12	0.05	-0.08	0.05	0.00	0.97	0.01	0.94	0.27	0.12	0.06	0.50	0.05	0.57	-0.06	0.51
hh head age ¡= 24	-0.33	0.01	-0.18	0.03	-0.41	0.05	-0.09	0.65	0.04	0.91	0.00	0.99	0.04	0.83	-0.15	0.30
hh head age 25 - 34	-0.18	0.11	-0.14	0.06	-0.27	0.19	-0.05	0.81	0.05	0.90	0.21	0.22	0.06	0.73	0.00	1.00
hh head age 35 - 44	-0.10	0.34	-0.14	0.06	-0.29	0.14	-0.03	0.87	0.04	0.92	0.16	0.36	0.12	0.48	0.01	0.95
hh head age 45 - 54	-0.10	0.37	-0.12	0.10	-0.34	0.09	-0.01	0.98	0.14	0.72	0.16	0.35	-0.06	0.72	0.07	0.59
hh head age 55 - 65	-0.06	0.62	-0.07	0.40	-0.21	0.31	-0.04	0.82	0.04	0.92	0.34	0.04	0.05	0.78	-0.08	0.56
access to public water	0.14	0.00	0.03	0.23	-0.07	0.52	-0.04	0.53	0.02	0.91	0.06	0.49	0.02	0.73	0.13	0.00
has no toilet	-0.20	0.00	-0.19	0.00	-0.05	0.59	-0.04	0.58	-0.28	0.05	0.03	0.71	-0.22	0.00	-0.15	0.00

continued on next page

Table 7 continued

	City								Town				Rural			
	1989		1994		1999		2002		1999		2002		1999		2002	
	β	P	β	P	β	P	β	P	β	P	β	P	β	P	β	P
no partner in household	0.32	0.00	0.52	0.00	0.11	0.64	0.28	0.08	0.47	0.17	0.27	0.20	0.41	0.03	0.07	0.72
com. basic edu. (m.)	0.00	0.98	0.02	0.66	-0.16	0.21	-0.03	0.74	0.01	0.96	0.07	0.49	-0.05	0.55	0.13	0.02
incom. secondary edu. (m.)	0.02	0.76	0.04	0.24	-0.03	0.78	-0.08	0.47	-0.14	0.44	0.08	0.37	-0.05	0.47	0.09	0.07
com. secondary edu. (m.)	0.10	0.03	0.10	0.01	-0.06	0.52	0.09	0.31	0.26	0.10	0.07	0.42	0.03	0.77	0.17	0.02
tertiary edu. (m.)	0.51	0.00	0.31	0.00	0.35	0.00	0.38	0.00	0.01	0.97	0.34	0.01	0.56	0.01	0.33	0.05
com. basic edu. (w.)	-0.01	0.87	0.07	0.08	0.10	0.44	-0.08	0.52	0.07	0.70	-0.03	0.74	0.27	0.00	0.24	0.00
incom. secondary edu. (w.)	0.14	0.01	0.02	0.61	0.14	0.17	0.04	0.66	0.16	0.23	0.07	0.50	0.28	0.00	0.29	0.00
com. secondary edu. (w.)	0.17	0.00	0.10	0.00	0.26	0.01	0.04	0.57	0.18	0.20	0.29	0.00	0.43	0.00	0.36	0.00
tertiary edu. (w.)	0.34	0.00	0.35	0.00	0.53	0.00	0.42	0.00	0.48	0.01	0.53	0.00	0.59	0.02	0.59	0.01
high skilled white collar (m.)	0.59	0.00	1.10	0.00	0.74	0.00	0.87	0.00	1.07	0.00	0.78	0.00	0.99	0.00	0.30	0.15
med. skilled white collar (m.)	0.32	0.00	0.55	0.00	0.37	0.09	0.38	0.01	0.95	0.01	0.53	0.00	0.64	0.00	0.18	0.43
skilled manual (m.)	0.37	0.00	0.51	0.00	0.24	0.25	0.20	0.12	0.55	0.13	0.45	0.01	0.73	0.00	0.10	0.57
unskilled manual (m.)	0.43	0.00	0.33	0.00	0.22	0.35	0.30	0.05	0.35	0.34	0.41	0.02	0.54	0.00	0.04	0.81
agr. employed (m.)	0.51	0.00	0.59	0.00	-0.32	0.48	0.38	0.07	0.46	0.25	0.75	0.00	0.57	0.00	0.53	0.02
agr. self-employed (m.)	0.38	0.00	0.39	0.02	0.27	0.29	0.09	0.68	-0.09	0.82	0.22	0.27	0.31	0.06	-0.05	0.78
sales and services (m.)	0.44	0.00	0.58	0.00	0.34	0.12	0.50	0.00	0.86	0.02	0.63	0.00	0.72	0.00	0.34	0.06
high skilled white collar (w.)	0.90	0.00	0.91	0.00	0.40	0.01	0.72	0.00	0.65	0.00	0.92	0.00	-0.06	0.83	0.68	0.00
med. skilled white collar (w.)	0.37	0.00	0.39	0.00	0.32	0.00	0.38	0.00	0.84	0.00	0.52	0.00	0.17	0.27	0.23	0.24
skilled manual (w.)	0.21	0.00	0.20	0.00	-0.07	0.53	0.09	0.33	0.44	0.00	0.18	0.06	-0.11	0.29	0.03	0.72
unskilled manual (w.)	0.43	0.01	0.34	0.00	0.31	0.00	0.33	0.00	0.63	0.00	0.15	0.12	-0.03	0.85	0.05	0.60
agr. employed (w.)	0.64	0.00	0.33	0.03	0.93	0.11	0.12	0.80	-1.02	0.06	0.00	0.00	-0.04	0.66	-0.21	0.46
agr. self-employed (w.)	0.46	0.11	-0.40	0.16	0.51	0.01	-0.08	0.79	-0.24	0.46	-0.48	0.00	-0.05	0.47	-0.05	0.37
sales and services (w.)	0.35	0.00	0.30	0.00	0.28	0.00	0.26	0.00	0.73	0.00	0.34	0.00	0.31	0.00	0.24	0.00
birth in last 12 month	-0.14	0.02	-0.14	0.01	0.18	0.39	-0.13	0.35	-0.25	0.36	-0.13	0.36	-0.08	0.27	0.02	0.81
attended by doctor	0.10	0.27	0.13	0.10	-0.16	0.51	0.06	0.72	0.61	0.07	0.04	0.82	0.21	0.07	0.03	0.83
delivered in hospital	0.01	0.94	-0.02	0.73	-0.09	0.57	-0.30	0.09	-0.25	0.25	0.15	0.34	0.10	0.45	0.13	0.32
child under 4 years	-0.02	0.62	0.00	0.88	0.07	0.48	0.18	0.08	0.12	0.48	-0.01	0.92	-0.04	0.65	0.06	0.35
has had diarrhea	-0.09	0.08	-0.05	0.42	-0.18	0.07	0.04	0.66	0.00	0.98	0.01	0.91	0.02	0.74	-0.16	0.02
has head cough/fever	-0.04	0.46	-0.12	0.00	-0.02	0.85	0.01	0.91	0.06	0.65	-0.02	0.82	0.01	0.87	0.09	0.12
c/t/r dummy/constant	4.48	0.00	4.10	0.00	4.63	0.00	4.04	0.00	3.82	0.00	3.76	0.00	3.88	0.00	4.13	0.00
# of observations	4607		5131		1037		1506		332		1120		922		1709	
R^2	42.10		50.33		47.15		43.94		53.43		44.64		48.89		42.46	

Notes: For details on the regression and variables, see text and notes of Table 2. β: regression coefficient; P: P-value; all: all possible covariates; common: covariates common over all 4 years.
Source: Own calculations based on ECH and EIH.