Finding the Poor in Thailand

Andrew J. Healy*           Somchai Jitsuchon
Loyola Marymount University  Thailand Development Research Institute

May 7, 2007†

Abstract

This paper provides the first set of precise micro-level estimates of Thai poverty and inequality. By combining household survey and national census data, we estimate Thai poverty and inequality at the district level. The standard errors on our district-level estimates are smaller than the province-level estimates that come from the household survey by itself. We demonstrate that our results can significantly improve the targeting of policies to poor Thai households. Our estimates can also help researchers to answer a range of questions relating to micro-level poverty and economic development.

JEL codes: O10, O12, O53, C50

*Corresponding author, e-mail: ahealy@lmu.edu; telephone: 310-338-5956; fax: 310-338-1950; address: Loyola Marymount University, Department of Economics, 1 LMU Drive, University Hall 4229, Los Angeles, CA 90045

†We acknowledge financial assistance from the World Bank, and we thank Yos Vajaragupta for dedicated research assistance. Special thanks go to Peter Lanjouw and Khankaew Varakornkarn from the World Bank for their frequent assistance throughout this project. At the Thailand Development Research Institute, K. Nuntaporn and K. Tuang provided invaluable assistance at various points in this research. We acknowledge numerous suggestions from members of the Thai National Economic and Social Development Board and the National Statistical Office, who participated in numerous seminars. Finally, we acknowledge helpful discussions with Nanak Kakwani, Chris Elbers, and Esther Duflö. All errors are our own.
1 Introduction

Until recently, available national data in most developing countries could describe poverty levels only for entire provinces. Household surveys, the best available data source for household income and consumption, are generally representative only at the province level. In Thailand, for example, the changwat (province) is the stratum in the Socio-Economic Survey (SES) and the most specific attainable geographic level for SES-based poverty and inequality measures. There are seventy-six changwat in Thailand, a country of sixty million people. Policymakers interested in doing a better job of targeting policies to poor Thai households have pointed to the need for a more precise spatial description of poverty and inequality. Previously, the federal government directly disbursed nearly all funds, with little local discretion. In the last few years, Thailand has moved toward more local control of policy implementation. As a more even balance develops between local and federal authority, the ability to implement policies on a local level will increasingly be present, making a disaggregated description of poverty and inequality especially valuable. In this paper, we present results that describe Thai poverty and inequality in such a way. This paper also demonstrates that the improved identification of poor households makes possible a substantial improvement in policy.

We obtain these results by combining the 2000 Socio-Economic Survey (SES) with the 2000 Population and Housing Census. With some improvements, we apply the poverty mapping methodology developed in Elbers, Lanjouw, and Lanjouw (2003) to the Thai case.\(^1\) The basic idea contained in that paper is quite simple. The strength of the SES is the comprehensive data it provides for each household in the sample. Specifically, it reports household income and consumption data. It is, however, representative only at the level of the changwat (province). On the other hand, the census covers all households, but it lacks data on household income and consumption.

By combining the two data sources, we can take advantage of the strengths of each. We use the SES to model income and consumption as functions of correlated variables. We choose right-hand side variables that both the SES and census report. The estimated model

\(^1\)For papers that have combined survey and Census data to spatially decompose poverty and inequality estimates, see Hentschel et al. (2000) for Ecuador, Alderman (2002) et al. for South Africa, and Mistiaen et al. (2002) for Madagascar.
for the SES households can then be used to predict income and consumption for the census households. Using careful econometrics to estimate the correct standard errors, we obtain estimates of poverty and inequality at all levels of geographic aggregation, from the changwat all the way down to the village. The standard errors become larger for lower levels of spatial aggregation. We find that we obtain estimates with reasonable standard errors down to the tambon (subdistrict) level. The standard errors for our estimates at the subdistrict level are actually somewhat smaller than the estimates obtained using the SES alone at the province level.

Of course, precise estimates are only valuable to the extent that we can verify that those estimates are unbiased. Fortunately, we can compare our results at the changwat level against the SES’s estimates to test our methods. We show that our estimates of poverty and inequality correspond closely to the changwat level estimates that the SES gives. The estimates found in our poverty map are within the 95% confidence intervals for the changwat poverty and inequality estimates from the SES approximately 95% of the time. When we look at the level of the region, of which there are five in Thailand, the map and the SES show an extremely close correspondence. These results indicate that our estimates are unbiased and can be useful in targeting policies to geographic levels below the changwat. At levels well below 5,000 households, we produce estimates of poverty and inequality that have standard errors of the same size as those obtained at the changwat level using the SES. We now can say a tambon (subdistrict) is poor with the same confidence we could previously apply to an entire changwat.

We present estimates of poverty and inequality at five different geographic levels. The broadest classification is the region. There are five regions in Thailand: North, Northeast, South, Central, and Bangkok. Of these, the Northeast is the poorest and Bangkok the richest. There are 76 changwat in Thailand, where changwat is the Thai word for province. Thai changwat are broken into amphoe, which roughly correspond to American counties. The word amphoe is translated into English as district. Amphoe are further broken down into tambon, which is translated into English as subdistrict. Finally, we estimate poverty and inequality even at the village level, but with less precision.

The primary usefulness of the estimates we report is as an input to guide policy and
further research. To demonstrate the value of our results, we ask a policy question. If a government's goal is the redistribution of resources to reduce the poverty gap, how much greater of a reduction can be achieved by using our estimates as opposed to the data previously available? Given the precision with which we can estimate poverty at even the subdistrict level, the answer is not surprising. Our results have the ability to lead to a much more effective allocation of resources.

Section 2 describes the Thai data and the methodology that we apply to it. Section 3 compares our results to those produced by the SES for high levels of geographic aggregation. Section 4 reports our estimates of poverty and inequality at the amphoe (district) and tambon (subdistrict) level. Section 5 considers a modified model that can be used in addition to our primary estimates. In Section 6, we demonstrate that these estimates make possible significant improvements in resource allocation. Section 7 concludes.

2 Poverty Mapping Methodology and the Data

2.1 Methodology

Our goal is to predict income and consumption for the census households. To make these predictions, we find variables that are present both in the census and the SES. These variables are then used to estimate models of income and consumption for the survey households.\(^2\) The coefficients (and their standard errors) and the resulting residuals then enable us to estimate income and consumption for the census households.

\(^2\)For India, Deaton and Taorozzi (2000) and Taorozzi (2002) use related methods to estimate consumption for the National Sample Survey (NSS) households. They have a set of variables that they know are well-measured and use them to estimate consumption for the households, leading to estimates of Indian poverty and inequality. Their goal is to resolve problems with the NSS, while ours is to obtain a spatial picture of poverty and inequality.
2.1.1 Finding variables present in both the census and the SES

The first step is to identify the variables that are present and identically defined in both the SES and the census. These variables will form the set from which we choose the variables in our income and consumption models. Before estimating those models, we run tests to confirm that variables are identically defined in the two data sets. We conduct these tests by comparing the changwat means for each variable between the census and the SES. If a variable really is defined in the same way, then we should not reject a test of equality between the census mean and the SES mean.

We consider household asset, demographic, and occupational variables that are plausibly correlated with income or consumption. To ensure that variables are defined the same way in the census and the SES, we test the hypothesis that the variable’s mean in the census and its mean in the SES are equal.\(^3\) We find the variables that pass this test for each changwat, urban and rural. To avoid the possibility that any variable essentially acts as a dummy for a given household or two, we consider only dummy variables that have a mean greater than 0.03 and less than 0.97. For example, consider a changwat that has one hundred households included in the SES. For this changwat, a dummy variable that has mean 0.01 is zero for all households except one. This variable then acts as a dummy for that household in any model. Since the goal is to estimate income and consumption for the census households, this variable is clearly not helpful in constructing an appropriate model. The household represented by the dummy variable cannot be identified in the census. Even if it could, we would not want to use a degree of freedom representing one household when we need to predict income and consumption for all the census households.

\(^3\)To be more precise, we should use more than a variable’s first moment. Most of our variables will be dummies, however, making the mean a sufficient statistic for the distribution. We have performed sensitivity checks for other variables, and find that looking at higher moments adds no change to the set of variables that pass this test.
2.1.2 Selection of household-level correlates

After ascertaining which variables are defined in the same way in the census and SES, we model household income or consumption in the SES as a function of the variables that pass the test described in the previous section. Where \( y_{ch} \) refers to log income of household \( h \) in cluster \( c \) (the cluster is the village in the SES), \( X_{ch} \) denotes the set of available regressors, \( \eta_c \) are village dummies, and \( u_{ch} \) is a household error term, our regression equation is

\[
y_{ch} = E(y_{ch} | X_{ch}^T, \eta_c) + u_{ch} = X_{ch}^T \beta + \eta_c \gamma + u_{ch}.
\]

We improve the previous methodology by including village dummies in the regression. In this way, we pick the household variables that best correlate with the household-level variation in income. Since later steps will account for location-level variation, our aim is to choose the variables that specifically capture household-level variation.

It is important to note that regression equation (1) does not describe a causal relationship. The usual concerns about endogeneity do not apply to our models. We are interested in obtaining consistent estimates of income and consumption for the census households; we are not concerned with any of the individual regression coefficients. We assume that \( u_{ch} \) is distributed with mean zero and variance-covariance matrix \( \Sigma \). A stepwise regression selects all variables that are significant at the 5% level.

There is a question as to whether the regressions should be estimated using the household weights provided in the SES. Note that the usual arguments against weighting do not apply in our context. Our regression is descriptive and thus weights may be needed to provide a consistent estimate of the population regression function (Deaton, 1998). Since a Hausman test (Hausman, 1978) rejects parameter equality at a 5% level for most of the provinces, we use the household weights provided in the SES when estimating our models. Table 1 summarizes the results of the tests that we conducted for consumption in the rural areas.

{Insert Table 1 here}
2.1.3 Selection of location-level correlates

After choosing household-level variables, we then select variables to absorb the location component of the variance. We choose from census means at the village level for rural areas and census means at the tambon level for urban areas.\(^4\) For rural areas, the 1999 Village Survey provides additional candidate variables to model the effect of location on income and consumption. This data set is particularly helpful since poverty in Thailand, as in most developing economies, is concentrated in rural areas. We discuss the Thai Village Survey in greater detail Section 3.

We cannot use location dummies to predict income in the census because we can only estimate parameters for the SES villages. We can certainly include village dummies in the model that we use in the SES. However, doing so would not be useful since we need to model income and consumption for all the census households and only a very small number of the census villages are contained within the SES sample.

Consequently, we approximate the location effect with location-level regressors that are available in the census. To do so, we estimate a village effect from regression (1) and regress it on the set of location variables. Our regression equation, where \(\hat{\eta}_c\) is the estimated village effect from (1), is

\[
\hat{\eta}_c = Z^T_i \alpha + v_{ch},
\]

where \(Z^T_i\) are the location variables, \(i = t\) (tambon) for urban areas, and \(i = c\) (village) for rural areas. Here we employ a forward stepwise regression procedure that ensures all the previously selected household variables remain and then adds any location variables that are significant at the 5\% level.

\(^4\)In urban areas, we cannot use village-level means due to a limitation in our ability to match village identifiers between the SES and the census.
2.1.4 Modeling the residual distribution

The stepwise regressions (1) and (2) select the set of household and location variables that we use to model income and consumption. To simulate income and consumption for the census households, we also need to consider the distribution of the residuals from the regression of log income and consumption on the selected household and location variables.

Where the superscript $S$ indicates the set of selected variables, $X_{ch}^{T,S}$ denotes the set of selected household variables, and $Z_{i}^{T,S}$ denotes the set of selected location variables, we have

$$y_{ch} = X_{ch}^{T,S} \beta + Z_{i}^{T,S} \alpha + u_{ch} \quad (3)$$

Both due to the presence of a significant cluster component and the fact that the variance of $e_{ch}$ depends on observables, we can easily reject homoskedasticity of $u_{ch}$. As a result, we model the residual as having a cluster component and a household component. We take the error term from the regression of $y_{ch}$ on the selected household and location variables, $u_{ch}$, and break it up into a cluster component and a household component, $e_{ch}$.

$$u_{ch} = \eta_{c} + e_{ch}$$

Following Elbers, Lanjouw, and Lanjouw (2003), we then use a logistic model to find the conditional variance of the household component. Where $H_{ch}^{T}$ denotes cluster and household-level variables, the regression we consider is

$$\ln \left[ \frac{e_{ch}^{2}}{A - e_{ch}^{2}} + \phi \right] = H_{ch}^{T} \theta + r_{ch} \quad (4)$$

The constant $A$ is $1.05 \times \max\{e_{ch}^{2}\}$, and bounds the predicted variance between 0 and $1.05A$. We apply $\phi = 0.001$. For Thailand, we found that failing to use 0.001 or some other small constant causes too much weight to be put on a couple of households in some cases. Even with our safeguards to ensure that the model includes no variable that essentially acts as a dummy for a given household, it is still our experience that sometimes a household will have a very small residual from (3). With this adjustment, we find that neither households
with large residuals nor those with small residuals interfere with the choice of appropriate variables for regression (4). The delta method and a first-order Taylor approximation gives the following estimate for the conditional variance of $e_{ch}$, where $B = \exp(H_{ch}^T \hat{\theta})$ and $\hat{\sigma}_r^2$ is the estimated variance of $r_{ch}$:

$$\hat{\sigma}^2_{e,ch} \approx \left[ \frac{AB - \phi A}{1 - \phi + B} \right] + \frac{1}{2} \hat{\sigma}_r^2 \left[ \frac{AB(1 - \phi - B)}{(1 - \phi + B)^3} \right]$$

(5)

We use the above variance estimator to generate a set of standardized household residuals. We then draw the location component and the household component either directly from the standardized residuals or from the most appropriate parametric distribution.\(^6\) We consider normal and $t$ distributions for the household error. Whether one of these parametric distributions is employed or we draw from the empirical distribution has little effect on the resulting poverty and inequality estimates.

We obtain GLS estimates of the parameters and then conduct simulations to assign error terms to households. This procedure enables us to generate estimates and confidence intervals for the Foster-Greer-Thorbecke (FGT) indices (Foster, Greer, and Thorbecke, 1984), the Atkinson inequality measures (Atkinson, 1970), and the Gini coefficient at any level of geographic aggregation. We focus on the headcount index and the Gini in this paper.

2.2 The data

The income data we use comes from the 2000 Socio-Economic Survey. The SES is a stratified random sample of 24,747 Thai households. There are 76 strata, one for each changwat. Each of these strata is divided into three categories: urban, sanitary district, and rural. The census has only two categories, rural and urban, and we employ the census definition when classifying households as either urban or rural.

\(^5\)We thank Chris Elbers for helping with this derivation.

\(^6\)We model a location component of the residual wherever possible. If the location variables capture the location effect particularly well, the fact that the estimate of the household variance is not the true value means that we can get a negative variance estimate for the location part of the residual. For these cases, we estimate the model without a location component to the residual.
Providing the greater coverage we need to obtain spatially disaggregated estimates of poverty and inequality is the 2000 Population and Housing census. The short form of the census completely covers the population. It has data on household composition, education, and occupations, but not assets. The 20% subsample also has information on assets. We have produced Thai maps using both the 100% sample and the 20% sample. We found that the greater precision that comes from the richer set of variables in the 20% sample dominates the loss of coverage. As a result, we use the 20% subsample to construct our estimates.\footnote{While the 20% sample gives the most precise results down to the tambon level, this should add an extra bit of caution in considering results at the village level. Using the 20% sample introduces error in the Census means, since they are now only an estimate of the population mean. We estimate that this only becomes an issue for population units of 25 households or less. Details of these calculations are available upon request.}

We break Bangkok up into four parts due to its size. We report results that aggregate the map estimates for the four parts.

For rural areas, we also include variables from the 1999 Village Survey. The data comes from a questionnaire sent out to the village headperson in each Thai village. Included is data on irrigation, the total number of households in a village, and farm assets. In our results, we find a somewhat better fit for rural areas compared to urban areas, and the extra information provided by the Village Survey is the primary reason.

\subsection*{2.3 First-stage results}

Compared to poverty maps that have been done in other countries, we find that the set of chosen regressors can account for a larger part of the variance in income and consumption. This finding speaks to the high quality of the Thai data. The SES is known for its reliability among the household surveys available outside the industrialized countries (Townsend and Gine, 2003).

Table 2 shows the median number of regressors and the median number of households per stratum, for each of the regions. It also describes the quality of fit of the regression model in the last two columns, which describe the median and average $R^2$ for the models, again by region. Notice that a relatively small number of regressors are capable of accounting for a large share of the variation in rural consumption. In Thailand as a whole, the median...
number of included regressors is 11 with a median of 147 households in the models. On average, 70.9% of the variation in consumption is accounted for by the regressors that are included in the models. This means that most of the variation in our simulations will come from variation in the chosen regressors and those coefficients, as opposed to that variation introduced by the random error term draws.

{Insert Table 2 here}

The qualitative nature of the results seen in Table 2 for rural consumption holds true for urban areas and for income. The set of regressors, on average, accounts for over 70% of the variance in consumption. This ability to precisely model income and consumption leads to the small standard errors for tambon-level poverty and inequality estimates that we report in Section 4.

3 Region and Changwat Results: Comparisons with the SES

Our contribution is to provide estimates of poverty and inequality at geographical levels like the amphoe and the tambon. Since our work has provided the first of these estimates for Thailand,\(^8\) we cannot check our estimates at these levels. Here, though, we show our estimates at the region and changwat level and compare to the estimates given by the SES. The close tracking confirms that our estimates perform well.\(^9\)

Of course, these results are not are primary goal; our primary contribution comes from the amphoe and tambon estimates that we report in Section 4. The close correspondence between our results and the SES estimates at the changwat level, though, gives us confidence

---

\(^8\)More precisely, our work provides the first of these estimates based on the directly collected, national data sets collected by the Thai government. Another data set, based on letters sent to village leaders, gives a partial list of poor villages in Thailand. The limitations of, and concerns about, this list helped to motivate Thai officials into seeking the results reported here.

\(^9\)We focus on the headcount and the Gini in the discussion. The correspondence between the map and the SES carries over to other measures, such as the Atkinson inequality index and the poverty gap. Details are available upon request.
about our results at lower geographic levels. In other words, the fact that our estimates
match the SES at the region and changwat levels means that we can trust our estimates
of poverty and inequality at the amphoe and tambon levels. The reported results refer to
normally distributed error terms, but the qualitative nature of the results is the same with
other error distributions.

3.1 Poverty

Consider first the comparison at the region level. Table 3 shows the comparisons for the
headcount index for rural areas. In all cases, the map estimate of poverty is well within
the 95% confidence interval generated by combining the point estimates and standard errors
for both the map and SES estimates. What is particularly encouraging is that the tracking
seen in Table 3 comes not from large standard errors but from a close tracking in the point
estimates.

{Insert Tables 3 and 4 here}

Nationally, the estimates of poverty differ by less than a percentage point. The tracking
is also very close for the North, the South, and for income in the North. The tracking is
somewhat less precise for the Central region and for consumption in the Northeast.

We also find a close correspondence between our results and the SES estimates for the
headcount in urban areas. Table 4 shows this comparison. Again, the national estimates
differ by less than a percentage point. In addition, the estimate for the South, the Northeast,
and Bangkok are all quite close. The Central estimates are somewhat lower in the SES, but
the test for equality is still not close to rejecting at a 5% level. We do reject the test for
equality for consumption in the North, with an asymptotic z-statistic of 1.98. In other words,
the test for equality cannot be rejected for 21 of the 22 comparisons presented in Tables 3
and 4.
At the changwat level, the correspondence between the map and the SES is also encouraging.\textsuperscript{10} At very low poverty levels, comparisons become difficult. For example, rural Samutprakan has zero poverty in the SES. This estimate has a standard error of zero, and thus we will reject equality for any non-zero map estimate of poverty.

Consider those provinces with at least 2% poverty estimates in the SES. In rural areas, 97% of the comparisons out of 63 possible accept the hypothesis of equality between the map and the SES for income. The hypothesis cannot be rejected for 90% of the comparisons out of 59 for consumption. In urban areas, 89% of the comparisons out of 61 cannot reject for income and 95% out of 56 cannot reject for consumption.

In sum, we find a very close tracking between the map and the SES estimates of the headcount index for income and consumption in both rural and urban areas. This finding holds true at the national, regional, and provincial levels. The robustness of this correspondence should inspire confidence in our estimates for the headcount at the amphoe and tambon levels.

### 3.2 Mean income

For mean income, the poverty map matches the SES with less frequency. Consider Tables 5 and 6, which contain the comparisons for rural and urban areas. In both urban and rural areas, we cannot reject equality for any of the regions, but we easily reject equality for Thailand as a whole for both income and consumption. Moreover, some of the failure to reject is driven by large standard errors rather than close tracking between the map and the SES.\textsuperscript{11} The correspondence is closest for consumption in the rural and urban North, income in the rural South, and for income and consumption in Bangkok. The map and the SES are particularly disparate in their estimates of mean income in the urban Central and North.

\{Insert Tables 5 and 6 here\}

\textsuperscript{10}Details of all changwat-level comparisons between our estimates and the SES are available from the authors upon request.

\textsuperscript{11}These standard errors are driven in large part by the substantial heterogeneity observed among the changwat in any region.
Most strikingly, our estimate of mean income is higher in all cases, revealing some sort of bias in either the SES or in our estimate. While it is not possible to make any certain conclusions, this finding is consistent with a frequent criticism of household surveys: non-response among wealthy households that is not reflected in the household weights. For example, Mistiaen and Ravallion (2003) report that non-response is higher among wealthy households in the US Current Population Survey and thus that the survey seriously underestimates American inequality. Deaton has documented this concern for a variety of countries, most notably India.\textsuperscript{12} This sort of non-random non-response is consistent with the finding that our estimates for mean income are higher than those produced by the SES alone.

To see this, notice that our estimates would not be biased by the presence of non-random non-response in the SES. The model of income or consumption that we estimate for the changwat is valid for the households that are included in the sample. The coefficients and residuals that we use to model income or consumption for the census households continue to be valid. Since the census has universal coverage, our estimates are not biased by non-response in the SES.

The finding that our estimates of mean income are higher than the SES estimates carries over to the changwat results. For rural areas, the test for equality between the map and the SES passes 89\% of the time for income and 84\% of the time for consumption. For urban areas, the equality test passes only 77\% of the time for income and 87\% of the time for consumption. In 46 of the 47 times in which the equality test fails to pass, the map estimate of income or consumption is higher than the SES estimate.

Supporting the idea that wealthy non-response is driving the discrepancies seen in Tables 5 and 6 is the fact that the poverty map matches the SES quite closely for measures where this source of bias is not relevant (the headcount), somewhat closely when it is of lesser importance (the Gini), and much less closely for the measure where it is of greatest importance (mean income). All results are consistent with the hypothesis that our estimates are unbiased and that there is non-response by the wealthy in the SES.

\textsuperscript{12} It is believed in Thailand’s National Statistical Office that the SES may suffer from this problem. The NSO is presently considering a research effort into the importance of wealthy non-response in the SES.
3.3 Inequality

The Gini coefficient is broadly the same in the comparisons between the map and the SES. For all regions, we cannot reject the hypothesis that the Gini is the same in the SES and the map. Tables 7 and 8 describe all the comparisons for the Gini. The tracking is very close in most cases. The only note of caution to add to this result is the fact that 15 of the 18 regional comparisons give higher Ginis in our estimates than in the SES. Again, the results are consistent with some underrepresentation of the wealthy in the SES. Such non-response by the wealthy would explain the somewhat higher inequality estimates higher in our estimates compared with the SES.

In the aggregate, income equality is similar in urban and rural areas. Estimated from the SES, the Gini coefficient in urban areas is 0.453 and the Gini in rural areas is 0.429. Our corresponding estimates are 0.463 for urban areas and 0.437 for rural areas. For rural inequality, we estimate that inequality is highest in the North (0.458) and lowest in the South (0.422). For urban inequality, the differences are somewhat larger. The Gini is highest in the Northeast (0.513) and lowest in Bangkok (0.400).

As Tables 5 and 6 showed, mean income is considerably higher in urban areas than in rural areas. As a result, inequality for Thailand as a whole is significantly larger than inequality for either rural areas or urban areas by themselves. The Gini coefficient for Thailand as a whole in 2000 was 0.53 (Jitsuchon and Plangprahan, 2001; Sunantha, 2002). The differences between the regions would also be somewhat larger if urban and rural were considered together. We estimate that the difference in mean income between rural and urban areas is smallest in the Central region, a region in which inequality is relatively low and where urban mean income is about 1.5 times larger than rural mean income, as shown in Tables 5 and 6. At the same time, the difference in mean income between rural and urban areas is largest in the Northeast, a region in which inequality is relatively high and where urban mean income is about 2.5 times larger than rural mean income. Our estimates thus show that the Northeast has the highest inequality within its urban areas and the Northeast also has the highest urban-rural disparity in mean income.

{Insert Tables 7 and 8 here}
Of primary interest in Tables 7 and 8 illustrate is the close correspondence between our estimates of region-level inequality with those estimates produced using the SES alone. Confirming the close correspondence seen in Tables 7 and 8 are the changwat comparisons. In rural areas, the equality test cannot be rejected for 95% of the provinces for both income and consumption. The equality test passes in urban areas 95% of the time for income and 97% of the time for consumption. As was the case for the headcount, the map produces estimates of inequality that correspond very closely to the SES estimates.

These results and their robustness provide the foundation for the amphoe and tambon estimates of poverty and inequality that we report in the next section. This section has used the SES estimates of poverty and inequality at the region and changwat levels to check that the map performs well. Given that validation, the next section demonstrates how our results add to the state of knowledge about Thai poverty and inequality by giving precise estimates at geographic levels below the changwat. Even though a changwat is on average sixty times larger than an tambon, we estimate poverty and inequality more precisely at the tambon level than the SES can estimate poverty and inequality at the changwat level.

4 Thai Poverty and Inequality at the Tambon Level

There are 7411 tambon in Thailand, 80.2% of which are either entirely urban or entirely rural. On average, a tambon is approximately 1/60 the size of a changwat.\(^{13}\) Despite the much lower level of geographic aggregation, the standard errors at the tambon level in the poverty map are actually lower than at the changwat level in the SES. This holds true for the headcount, mean income, and the Gini, both for income and consumption. Table 9 shows the comparisons for the mean of the standard errors over the changwat.

\(^{13}\)Specifically, there are 5498 tambon that are only rural, 447 that are only urban, and 1466 tambon that have both urban and rural parts, for a total of 8877 tambon map estimates. This compares to 151 map estimates at the changwat level, of which 75 are rural and 76 are urban.
For all variables, the map provides more precise estimates at the tambon level than the SES does at the changwat level. The tambon level is precisely the administrative level that is presently assuming greater authority in Thailand. Given the performance of the poverty and inequality measures described in the previous section, Thai officials can use this new information to target policies at the tambon level with similar confidence to the changwat-level targeting that was previously possible.

In addition, we use Geographic Information Systems (GIS) maps to report our estimates of poverty and inequality. Figure 1 shows the headcount for rural consumption at the tambon level. Figure 2 displays the headcount for rural consumption at the changwat level. Due to the lesser detail this map contains, it is easier to process visually. Similar maps describe, for example, the headcount for urban consumption and the Gini coefficient for rural income. These maps and the estimates behind them provide a more detailed picture of poverty and inequality in Thailand than was previously available. As shown in Table 9, the complete coverage of the census enables us to achieve disaggregated estimates of poverty and inequality with no loss in statistical precision.

5 Checking Our Estimates with Field Research

To supplement the validation exercise described in Section 3, we attempted to check our results with field research. To do this, we conducted field visits to villages in Nakhon Sithammarat, a province in southern Thailand. Throughout this paper, we have emphasized our results at the amphoe and tambon levels. At the level of the village, not surprisingly, our estimates are considerably less precise. We feel, as a result, that our estimates are primarily useful at the amphoe and tambon levels.

Still, the observations of the field team largely accorded with the predictions of our estimates, even at the village level. The instances in which our estimates appeared not to match the actual picture also provided an opportunity for us to investigate an alternative

---

14 All other maps are available from the authors, upon request.
model for more precise prediction. Specifically, we found one village where an apparently important variable did not make it into our models. We discovered that this variable was omitted because our regression models applied to the changwat as a whole. In some cases, it may be possible to run our models for an amphoe instead. For some large amphoe, the SES has enough households to allow estimation of income and consumption models. In the few cases where our map fails to match the SES, we find that estimating a model at the amphoe level makes it possible to obtain better estimates of poverty and inequality.

5.1 Validation at the village level

Before the poverty map, the Thai government compiled a partial list of poor villages using a data set called the NRD2C that came from responses to surveys sent out to village leaders. Several problems hamper this data. For example, the list classifies some villages as poor in part due to their lack of irrigated farmland, even though most villagers had ceased to be farmers. We also found that the list suffers from some inaccuracies in the reporting that seem to be intentional, possibly to manipulate the allocation of funds.

By comparison, the poverty map has several advantages. First, it accounts for all localities that might be absent from the NRD2C list because of the different or outdated categorization of rural villages and urban communities or unreturned questionnaires from village leaders. Second, it takes advantage of data that the government directly and comprehensively collects (the SES and the census). Perhaps due to these strengths, the poverty map appears to perform at least as well as the NRD2C at measuring poverty in the validation villages, despite a lack of any direct income or consumption data at the village level.

The poverty map does not, however, give the correct picture in all the villages. Since the map standard errors for mean income/consumption, the headcount, and the Gini all become large for the village-level estimates, this finding is not surprising. Still, we attempted to ascertain why the model gave a poor prediction in some of the villages. Our investigation revealed that the problem may have occurred because the income/consumption models do not use some village characteristics that were relevant in predicting the general income of the villages. Most notably the models do not treat adequately well the fact that many
villages in the Southern region grow para-rubber trees, and prices of raw rubber are always very important in determining the average income of the villagers. This problem could be overcome by using secondary data.

In the following section, we explore the former of these possibilities. The comparisons reported in this paper show that the poverty map generally performs well throughout Thailand. Finding a more refined model of income or expenditure should be capable of solving any problems that remain.

5.2 Using a model at a lower geographic level

Until now, we have discussed estimates of poverty and inequality that have come from a model of consumption that applied to an entire changwat. The regression coefficients that generated each household’s estimated income and expenditure were the same throughout the changwat. But, if one district in a changwat produces rubber and another produces textiles, the variables that correlate with consumption in each are likely to be different in each district. So we end up choosing between sample size and model precision in our regression models. We have already attempted to deal with this trade-off by allowing for conditional heteroskedasticity. Here, we also consider using a regression model based at the amphoe level rather than the changwat level.

The close correspondence between the poverty map and the SES at the changwat level makes it feasible for us to look at income and expenditure models for some large amphoe, to supplement the results from the changwat models that we have already discussed. Within some provinces, enough of the households are located within one amphoe to enable us to run a regression at that level. We cannot generate changwat-level estimates to compare with the SES. Earlier, though, we generated amphoe-level estimates using the changwat-level regressions. We can compare those estimates with what we obtain by running amphoe-level regressions.

We consider an amphoe model for one amphoe in each of four rural changwat and four urban changwat.\textsuperscript{15} Table 10 shows the comparisons for the headcount between our original

\textsuperscript{15} The rural changwat are Amnatcharoen, Kamphaengphet, Phuket, and Samutprakan. The urban changwat
results from the changwat model and the SES, at the level of the changwat. It also shows the comparisons for the headcount between our original results and the results obtained by using an amphoe model, at the level of the amphoe.

{Insert Table 10 here}

Consider Chachoengsao, a province where the poverty map and the SES disagree quite sharply at the changwat level. For the amphoe that we look at in Chachoengsao, the amphoe model gives a result that would lead to closer correspondence between the map and the SES. Our original estimate of the headcount was higher than the SES, and the amphoe model leads to a lower estimate of the headcount for the amphoe than does the changwat model. On average, the smaller sample size for the amphoe model and its greater precision cancel, and we get similar standard errors whether we use the changwat or the amphoe model. In the few places where the map struggles, consideration of more precise models offers hope that the poverty map methods can still yield accurate estimates of poverty and inequality at the amphoe and tambon levels. If researchers consider suspect some survey data from one part of a changwat, they can still generate poverty map estimates for other amphoe by looking at more amphoe-level income and consumption models.

6 Policy Targeting: A Demonstration

In this section, we consider an example of how policymakers could use the estimates reported in this paper. Specifically, we ask: If the goal is reduction of the poverty gap, how much of an improvement in resource allocation does our estimates make possible? We find that our estimates make possible a much more efficient allocation of resources and that they can guide even simple policies in subtle ways.

The poverty gap is the total distance between household income (or consumption) and the poverty line, summed over all households. Where $H$ is the number of households, $i$ are Chachoengsao, Lampang, Loei, and Satun. Here we just consider consumption poverty. Additional results for mean income and the Gini coefficient are available upon request.
indexes the households, $x_i$ is household income (or consumption), and $z$ is the poverty line, the poverty gap in changwat $j$ is

\[ PG_j = \frac{1}{H} \sum_{i=1}^{H} (1 - \frac{x_i}{z}) \mid (x_i > z) . \]

Consider a policymaker who wants to reduce the poverty gap as much as possible with $1$ of resources. We consider the marginal dollar to simplify the policymaker’s problem. Assume utility to be the expected reduction in the poverty gap. The policymaker achieves a reduction of $1$ in the poverty gap if she can target the dollar to any poor household, where we ignore the possibility of households being within $1$ of the poverty line.

Here we consider three kinds of possible targeting: changwat-level, amphoe-level, and tambon-level. Changwat-level targeting involves the $1$ being spread evenly to all the households in the changwat, with amphoe-level and tambon-level targeting having the analogous definition.

Using the changwat-level targeting possible with the SES alone, a policymaker would target the changwat with the highest headcount. Since the poorest tambon must be at least as poor as the changwat as a whole, targeting the poorest tambon will lead to at least as large an expected reduction in the poverty gap as targeting the changwat as a whole. In reality, there are likely to be costs associated with attempting to implement policy at a lower geographic level. Corruption, most notably, may become more severe. These costs may outweigh the benefits of more precise targeting.

Even if corruption costs are the same across changwat, different policies will be optimal in different places. Our estimates can help policymakers decide whether it is worth it to attempt targeting at lower geographic levels. Some changwat have relatively uniform poverty rates. For these changwat, the benefits of targeting at the tambon level are relatively small. Other changwat have large heterogeneity in poverty rates, and thus large benefits can be obtained by focusing resources on the poorer areas. To find out how our estimates can improve policies aimed at reducing the poverty gap, we compare our estimates for the headcount ratio at the changwat, amphoe, and tambon levels. The difference between the changwat-level headcount estimate and the amphoe-level headcount for the poorest amphoe gives the
Likewise, the difference between the amphoe-level headcount estimate for the poorest amphoe and the tambon-level headcount for the poorest tambon gives the benefit of targeting at the amphoe level compared to the tambon level. If the benefits from more precisely targeting outweigh the corruption costs, the policymaker should target to the lower level. Here we focus on the results for the Southern region of Thailand, which consists of 17 changwat. We run 200 simulations to estimate the distribution of the improvements in effectiveness achieved by targeting at the amphoe level compared to the changwat level, and at the tambon level compared to the amphoe level. Table 11 shows the potential improvements from targeting policies to the amphoe and tambon levels.

{Insert Table 11 here}

In general, we find that the potential improvements are larger for poorer places, which is not surprising. A changwat with near-zero poverty will have near zero poverty in its amphoe and tambon, and so there is less to be gained by targeting at the amphoe or tambon levels. For income in Narathiwat, we would expect to achieve a further reduction in the poverty gap of 16.7% by targeting at the amphoe level rather than the changwat level. We could achieve an additional improvement of 13.6% by targeting tambon rather than amphoe.

Other results show that sometimes the primary improvement is obtained by targeting at the amphoe level and other times the primary improvement is obtained by targeting the tambon. In Yala, we would achieve three times the improvement by going from changwat to amphoe (17.0%) as we would achieve by going from amphoe to tambon (5.5%). On the other hand, in Pattani, we would achieve a much larger improvement by going all the way to the tambon level (a 17.7% improvement compared to a 7.4% improvement).

Moreover, while the greatest potential for the map to improve resource allocation lies in the poorest areas, there are other changwat where the map can lead to large improvements in policy. Take Ranong, for example. Even though the headcount estimate for the changwat is only 9.6%, we can achieve a total improvement of 27.9% by targeting policy to the tambon instead of the changwat (with 22.8% of the improvement coming from moving from the
amphoe to the tambon level). In other words, $1 directed to Ranong at the changwat level would reduce the poverty gap by only 9.6 cents. Using our estimates to direct $1 to Ranong’s poorest tambon would reduce the poverty gap by 37.5 cents. Our estimates improve resource allocation in this case by a factor of four.

7 Conclusion

This paper provides the first comprehensive estimates of Thai poverty and inequality at a geographic level below the changwat (province). We obtain these estimates by applying the methods developed in Elbers, Lanjouw, and Lanjouw (2003), with some improvements that have proven helpful in the Thai case. We combine household survey data with national census data to utilize the comprehensive income data in the former and the complete coverage of the latter. To confirm that our methods work appropriately, we compare our estimates of poverty and inequality at the changwat level to those provided by the household survey alone. These tests confirm that the models used to predict income and consumption for the census households are appropriate ones.

This validation provides the necessary foundation for our most important results: our estimates of poverty and inequality and the amphoe and tambon levels. We provide estimates of poverty and inequality at the level of the tambon that actually have slightly smaller standard errors than the previously available estimates at the level of the changwat. In other words, even though a changwat is on average 60 times larger than a tambon, using the SES and the Census together makes it possible for us to estimate poverty just as precisely at the much more specific geographic level.

Our more precise spatial analysis of Thai poverty and inequality is of obvious interest to policymakers. This project arose through an interest on the part of Thai officials on the National Social and Economic Development Board to better target policies to their poor citizens. In addition, our estimates can be used as an input into other research. As one example, a current project (Sims, 2007) uses our estimates of poverty and inequality to investigate whether designating areas as protected forest helps promote the economic development of those areas. Our results could be useful to any economist interested in the
causes and effects of micro-level poverty, income, or inequality in Thailand.

Moreover, our estimates describe Thai poverty and inequality at a time of considerable interest in Thailand’s economic development. In 2000, Thailand was emerging from the recession that followed the 1997 financial crisis. The crisis itself followed a decade (1986-96) in which the Thai economy averaged annual growth of 9.1% (Jitsuchon, 2001). During that time, aggregate poverty in Thailand fell precipitously from a headcount rate of 29.5% in 1986 to a rate of 11.4% in 1996 (Thailand National Economic and Social Development Board, 2001). With the onset of the crisis in 1997, poverty had increased to 15.9% by 1999 before falling thereafter.

Most of the research into the crisis in Thailand has focused on the macroeconomic policies that precipitated the crisis and the macroeconomic responses to the crisis, with much less attention paid to the microeconomic effects of the crisis. Our estimates make it possible to estimate micro-level Thai poverty and inequality at the time when Thailand was just beginning to rebound from the crisis. By applying similar methods to previous and future years, it will be possible to see how poverty and inequality changed not only at the aggregate level, but also at the micro level as a result of the crisis and the growth that preceded it. As a result, obtaining micro-level estimates of poverty and inequality for a range of years should be an important topic for future research. The micro-level estimates of Thai poverty and inequality in 2000 that are reported in this paper improve the ability of policymakers to target poor households and make it possible for researchers to address a range of questions. Extending our analysis to other years will improve our understanding of the reasons that poverty and inequality change over time.

References


\(^{16}\)For a comprehensive review of the East Asian growth experience and crisis, see Rao (2001).


