

## 論文の内容の要旨

論文題目 Developing and validating regression models for predicting household consumption  
for the health insurance in Cambodia  
(カンボジアの医療保険のための家計消費推計モデルの構築と検証)

氏名 中村 治代

**Background:** Health financial protection is a challenge for low- and middle-income countries (LMICs), where the general tax revenue is limited. Although social insurance allows LMICs to effectively mobilize domestic financial resources, it requires compulsory participation realized through capacity-based contribution, that is another challenge for LMICs where most of the population is engaged in the informal economy. This study developed and validated household consumption predictive models with the existing survey data for Cambodia, where 70% of the population is uninsured and nearly 3% of households spend over 10% of their total consumption for healthcare, to allow the country to estimate the insurance contributions according to household ability to pay. Furthermore, this study estimated insurance contribution options, and evaluated equity and progressivity of each option.

**Methods:** This study used data of the Cambodia Socio-Economic Survey (CSES), conducted annually between 2010 and 2017, involving a total of 38472 households. Four alternative linear models were developed with the pooled data which increased precision and power of the models. The outcome variable of the models was annual household consumption that were aggregated by food, non-food, and housing consumption items in the CSES data. Based on the previous discussions in similar studies, 369 predictor variables were also created with the CSES data to explore every possibility to find the best predictors of the outcome. To avoid recall bias and underreporting, the models were mostly comprised of proxy indicators, such as information on occupation and educational attainment, instead of direct measurement of consumption, such as food expenditure.

The data were divided into a training set and a test set using the 8:2 ratio randomly. Using the training set, four regression models were constructed. For the linear and mixed-effects models, the predictor variables were initially screened using a partial correlation coefficient. The predictor variables were manually selected to construct Model A, while backward-selection technique within a stepwise regression analytical framework was employed to construct Model B. For Model C, a mixed-effects liner regression was

employed with the remainder of the stepwise selection, considering a random effect across the same province. To avoid overfitting and obtain unbiased estimates, Model D was constructed with adaptive least absolute shrinkage and selection operator (LASSO) regression. All the available predictor variables were used for Model D since adaptive LASSO can automatically select variables to optimize the prediction performance and interpretability of the statistical model, while ensuring the model parsimony. Subsequently, the trained models were applied to the test data. With this subset, the household consumption values were predicted, and the results were compared with the values reported by the CSES, which used the full-length questionnaires. Finally, the prediction performance was evaluated with mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Smaller values of the statistical measurements were preferred.

In the second step, ten options of hypothetical health insurance contributions were estimated for the currently uninsured population in Cambodia within the existing health insurance framework, by calculating the insurance costs and household capacity-to-pay (CTP), on the premise of compulsory participation. The ten options of household insurance contributions were estimated as ① 80% of household-based flat-rate contribution (HFC) and 20% of individual-based flat-rate contribution (IFC), ② 50% of HFC and 50% of IFC, ③ 20% of HFC and 80% of IFC, ④ 34% of contribution based on household capacity to pay (CTP), 33% of HFC and 33% of IFC, ⑤ 50% of CTP, 25% of HFC and 25% of IFC, ⑥ 50% of CTP and 50% of IFC, ⑦ 80% of CTP, 10% of HFC and 10% of IFC, ⑧ 100% of proportionate CTP, ⑨ 100% of progressive CTP at 1.0% for Q2, 1.2% for Q3, 1.4% for Q4 and 1.9% for Q5, and ⑩ 100% of progressive CTP at 0.3% for Q2, 1.0% for Q3, 1.2% for Q4 and 2.0% for Q5. The insurance contributions were estimated on the bases of both observed and predicted household consumptions, using the household consumption predictive model. The appropriateness of the insurance contributions was further assessed, regarding the amount and share in household CTP, and presented by economic quintiles (Q1 to Q5: Q1 is the poorest and Q5 is the richest). The amounts of household health insurance contribution options on a prediction basis were compared with those on an observation basis. The amounts and CTP shares of the insurance contribution with government subsidy were also compared with those without subsidies.

Finally, the equity and progressivity of the hypothetical health insurance contribution were assessed by examining how the cumulative proportion of the insurance contribution payment changed with the cumulative proportion of the households ranked by per capita household CTP, using the Gini coefficient, the concentration index, and the Kakwani index. The equity and progressivity were assessed for the insurance contribution options on both observation and prediction bases, and with and without

government subsidy. Additionally, dominance tests were conducted to determine whether the insurance contribution would help reduce economic inequality in Cambodia.

**Results:** 51 predictor variables were selected for Model A, 86 for Models B and C, and 162 remained for Model D. A positive linear relationship between observed and predicted household annual consumption was found in all four models. There was a subtle trend that the middle-class household consumption was likely to be underestimated. In contrast, the high-class household consumption was over-estimated in all four models, while the trend was less noticeable in Model D. The smallest values of MAE, RMSE and MAPE were calculated for Model B, followed by Model C, Model D and Model A. The rank was consistent for all three statistical measurements.

The total health insurance cost per capita per year was estimated to be 14.57 USD. Overall, the median household contributions were slightly lower on a prediction basis than those on an observation basis, while the IQRs were slightly higher on a prediction basis than that on an observation basis. When the insurance contribution was collected at a flat rate on a household or individual basis, the amount of household contribution increased as a fraction of the household-based contribution increased. For the insurance options with both flat-rate and capacity-based contribution fractions, the amount of government subsidy declined as the capacity-based revenue fraction increased. When contribution was entirely collected on a capacity basis, no government subsidy was required. It was inferred that prediction errors could erroneously impose contribution on some of the poor households, while erroneously exempt some of the rich households from contribution payment.

Median shares of all the ten hypothetical health insurance contribution options in household CTP were under three percent with government subsidy. For the flat-rate options, the CTP share of insurance contribution decreased as the economic status progressed. For options that included both flat-rate and capacity-based contributions, burden on the poor declined as the capacity-based revenue fraction increased. When contribution was entirely collected on a capacity basis, burden on the poor was further reduced. When contribution is progressively collected, burden on the poor became minimum, while excessive amount of contribution was calculated for the rich in return.

It was found that the richer households were currently contributing a larger proportion of healthcare

payments than the poorer households in Cambodia. Meanwhile, it was also found that the richer households were contributing a smaller proportion of healthcare payment in comparison to their CTP, and they would continue to do so unless progressive contribution is collected in the country. It was suggested that health insurance with progressive contribution collection would help redistribute income and reduce inequality in Cambodia.

**Conclusion and recommendations:** This study suggested that it is possible to develop a regression model using population survey data to make a reasonable prediction of household consumption in Cambodia. The predicted household consumption could further estimate capacity-based health insurance contribution for all households, including the informal sector. Moreover, if the health insurance becomes compulsory, the capacity-based contribution could increase the insurance revenue, reduce burden on the poor, and eventually reduce regressivity of the healthcare financing.

The final product of the household consumption predictive model will be an automated tool with selected predictor variables and respective regression coefficients. Although various issues remain for UHC achievement, incorporating the developed tool into the existing health financing system in Cambodia may enhance its current efforts to reduce catastrophic health expenditure and make progress towards UHC achievement.

The household consumption predictive model should be gradually tested in real settings. Subsequently, a well-designed feasibility study should be conducted to evaluate and improve the model. The feasibility study could also provide empirical data of how much administration cost would be required in what settings. Simultaneously, rigorous discussions should be held among the stakeholders to decide what social values should be considered to make the best health financing policy in the country, and how the personal data should be protected.

Future study could explore developing non-linear and region-specific models in Cambodia, particularly when more data become available. Similar studies could be conducted in other LMICs to apply the attempts of this study with some modification as the model development and application in health financing are highly context-specific. Although this study has merely provided a tool to promote a possible health financing measure, it is expected that these efforts would help Cambodia and other LMICs to strengthen their healthcare financing, and ultimately move towards UHC goals.