Developing and Validating Regression Models for Predicting Household Consumption to Introduce an Equitable and Sustainable Health Insurance System in Cambodia

Haruyo Nakamura (hnakamura@m.u-tokyo.ac.jp)
Tokyo Daigaku Daigakuin Igakukei Kenkyuka https://orcid.org/0000-0002-1557-0781
Floriano Amimo
University of Tokyo
Siyan Yi
National University Singapore Saw Swee Hock School of Public Health
Sovannary Tuot
KHANA Center for Population Health Research
Tomoya Yoshida
Japan International Cooperation Agency
Makoto Tobe
Japan International Cooperation Agency
Mizanur Rahman
University of Tokyo
Daisuke Yoneoka
University of Tokyo
Aya Ishizuka
University of Tokyo
Shuhei Nomura
University of Tokyo

Research

Keywords: Health financing, Health insurance, Contribution, Household-consumption assessment, Equity, Cambodia

Posted Date: September 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-78787/v1

Version of Record: A version of this preprint was published at The International Journal of Health Planning and Management on July 1st, 2021. See the published version at https://doi.org/10.1002/hpm.3269.

Abstract

Background

Financial protection is a key health system objective and an essential dimension of universal health coverage. However, it is a challenge for low- and middle-income countries, where the general tax revenue is limited, and a majority of the population is engaged in the informal economy. This study developed and validated regression models for Cambodia to predict household consumption, which allows the country to collect insurance contributions according to one's ability to pay. This strategy would maximize the contribution revenue, optimize the government subsidy, and simultaneously ensure equity in healthcare access.

Methods

This study used nationally representative survey data collected annually between 2010 and 2017, involving 38472 households. We developed four alternative prediction models for annual household consumption: ordinary least squares (OLS) method with manually selected predictors, OLS method with stepwise backward variable selection, mixed-effects linear regression, and elastic net regression, which resulted in an adaptive least absolute shrinkage and selection operator (LASSO) regression. Household-level socioeconomic characteristics were also included as the predictors. Subsequently, we performed out-of-sample cross-validation for each model. Finally, we evaluated the prediction performance of the models using mean absolute error, root mean squared error, and mean absolute percentage error (MAPE).

Results

Overall, we found a linearly positive relationship between observed and predicted household consumptions in all four models. While the prediction performance of the four alternative models did not substantially differ, Stepwise Linear Model showed the best performance with the lowest values in all three statistical measurements, including MAPE of 1.376%. The use of regularization and the mixed effects in the regression was not particularly effective in this environment. The household consumption was better predicted for those with lower consumption, and the predictive performance declined as the consumption level increased. Although the richer household consumptions were likely to be overestimated, the trend was less noticeable in Adaptive LASSO Model.

Conclusions

This study suggests the possibility of predicting household consumption at a reasonable level with the existing survey data. Such a prediction would enable the country to raise the secured health insurance revenue equitably. The prediction model should be tested in real settings and continuously improved.

Introduction

Universal health access and protection of the population against catastrophic health expenditures and impoverishment are the key targets in universal health coverage (UHC), a target (3.8) of the Sustainable Development Goal (SDG) 3 (1). Despite extensive efforts of the global community, however, the population incurring catastrophic health spending increased by 3.6% a year between 2000 and 2015 at the 10% threshold and by 5.3% a year at the 25% threshold (2). During the period, the largest concentration of the world population with catastrophic health spending shifted from low-income countries to middle-income countries, while around 70% was persistently concentrated in Asia (2). Evidence has suggested that financial protection can only be universally available if backed by funds from prepaid and pooled sources with subsidies for the indigent. The entitlement to guaranteed services should not be linked to employment status, but instead, it should be universal (3).

Two major prepaid and pooled health financing approaches have been introduced. One is through taxation (the Beveridge model), and the other is contributions collected for social insurance from the insured (known as the Bismarck model) (4). The former poses a challenge for low- and middle-income countries (LMICs), where the general tax revenue is limited. The latter is an alternative method as financial discipline could be maintained by establishing a contribution level to balance revenues and expenditures (5). The revenues contributed by the insured are, in fact, more secure than the general tax that is not guaranteed to be allocated to the health sector. Nevertheless, it is also a challenge for LMICs, where most of the population is engaged in the informal economy, making it difficult to estimate their income levels (6-9). Under such conditions, a flat-rate contribution can be collected. However, it often lacks equity (8) and endangers the financial sustainability of the insurance fund. This is because the contribution rate is usually set at a level that the lowest-income group can afford, and thus limits the total contribution revenue (6). Some countries, however, namely Japan, Korea, and Taiwan, successfully achieved UHC by introducing social insurance for those engaged in the informal economy. Those countries collect insurance contributions based on the household income level (10). Although a clear understanding of household income levels alone would not solve the issue, this seems to be a key and an absolute requirement for a successful introduction of universal social insurance.

A national survey often estimates household income or consumption in LMICs. However, it is usually composed of lengthy questionnaires that are not likely to be utilized regularly by local administrative staff to determine health insurance contributions. Studies have attempted to develop efficient scales to measure households' welfare or poverty status, mainly for social assistance programs (11-18), or a singular value decomposition, such as principle component analysis, for research purposes (19-21). Nonetheless, these tools merely identified poor households or ranked households by their welfare status. A couple of studies have attempted to predict household income or consumption. However, one dichotomized the households at a certain income level as a cut-off point (22). The other predicted national average household income at a country level (23). These attempts implied the possibility of estimating household economic status using a limited number of indices. However, no study has so far focused on predicting household income or consumption on a monetary basis to be applied for health insurance contribution determination.

The present study aims to develop and validate efficient regression models to predict annual household consumption in Cambodia, a lower-middle-income country in Asia (24), using the national survey data. In Cambodia, formal sector workers are enrolled in the National Social Security Fund (NSSF) health insurance, and the poor households are covered by the fully subsidized health protection program, the Health Equity Fund (HEF) (25, 26), but nearly 70% of the population remains uninsured (27). Cambodia's government plans to extend the NSSF health insurance to the currently uninsured population (28), although an additional budget is not guaranteed due to limited fiscal space. This study will help Cambodia implement a financially sustainable health insurance system that allows the insured to pay contributions according to their ability, and the state to redistribute wealth since larger contributions are collected from households with higher ability than those with lower ability. This study findings will also contribute to ensure equity in access to healthcare for the Cambodian population.

Materials And Methods

Data source

This study used the data of the Cambodia Socio-Economic Survey (CSES) conducted between 2010 and 2017 (29-36), publicly accessible upon request. The CSES is a nationally representative cluster sample survey, conducted annually by the National Institute of Statistics (NIS). The CSES uses systematic sampling with probabilities proportional to the size, based on the number of households per village retrieved from the public information source (33). The country's 24 provinces and municipalities, at the time of the surveys, were first divided into 19 separate groups. Each group was further divided into urban and rural strata, and a total of 38 strata were formed. The CSES is designed in the three-stage sampling at primary sampling units (PSUs), enumeration areas (EAs), and households (33). The interview was conducted with the household head, his/her spouse, or any other adult household member if the head and spouse were both absent.

For this study, we used pooled data of 38472 households covered in the CSES 2010–2017: 3592 in 2010 and 2011, 3840 in 2012 and 2013, 12090 in 2014, 3839 in 2015 and 2016, and 3840 in 2017 (29-36). The large data pool increased precision and power. Table 1 shows descriptive statistics and socioeconomic characteristics of the survey respondents.

Table 1 Descriptive statistics and socioeconomic characteristics of the survey respondents

	No of HHs	No of Indi-viduals	HH size[1]	HH head age[1]	F-headed HHs[2]	HH annual consumption[3]	CPI[4]
2010	3592	16510	4.6 (1.9)	46.2 (13.9)	22 (21-24)	678 (666)	100.000
2011	3592	16327	4.5 (1.9)	46.8 (14.0)	23 (21-25)	707 (628)	105.479
2012	3840	17644	4.6 (1.8)	47.3 (13.8)	22 (20-23)	814 (710)	108.572
2013	3840	17225	4.5 (1.8)	47.5 (13.6)	21 (20-23)	895 (695)	111.767
2014	12090	53968	4.5 (1.8)	47.8 (13.8)	22 (21-23)	894 (718)	116.076
2015	3839	17301	4.5 (1.7)	49.2 (13.7)	24 (22-25)	1036 (843)	117.493
2016	3839	16985	4.4 (1.8)	49.3 (13.9)	23 (21-24)	1178 (910)	121.071
2017	3840	16090	4.4 (1.7)	49.2 (13.7)	23 (21-25)	1179 (912)	124.572
Total	38472	172050					

Source: Cambodia Socio-Economic Survey 2010-2017 (29-36)

Notes: No: number, HH: household, F-headed: female-headed, CPI: consumer price index, 1. Mean (standard deviation), 2. Percentage (95% confidence interval), 3. Median (interquartile range) in current US dollars (1 USD = 4065.02 riels as of 23 June, 2020), 4. The 2010 base CPI (37) was used to adjust household consumption data for inflation in the analyses. The annual household consumptions in this table are unadjusted.

Analyses

Equitable health insurance contribution should be determined based on one's ability to pay, which is not simply defined as a current income function. It should, however, be more precisely defined as a non-subsistence effective income (38). Effective income is further defined as the income that households would behave as if they have when making consumption decisions (38). Households tend to smooth consumption over time by saving and borrowing (39), taking into account expected variations in income over the year, their assets and future earning potentials (38). Additionally, a policy paper suggested that consumption-based measure is more relevant in a lower-income setting where many households are borrowers, rather than savers (40). Therefore, we used annual household consumption as the basis to

estimate household's ability to pay. The household consumption in each year was transformed into the value of 2010 based on the consumer price index (37) to adjust for inflation in the eight-year study period.

Table 2 shows the household consumption aggregates, including food, non-food, and housing consumption items. The CSES household questionnaire is designed to collect consumption data on purchase in cash, consumption of own production, and consumption of items received in kind. We aggregated the data following the World Bank's guideline (41, 42), the most widely referenced guideline of household consumption aggregates, albeit excluding consumer durables due to insufficient information.

Table 2 Composition of household consumption aggregates

1. Food consumption (20 items)
Rice/cereals, meats, dairy products, vegetables, fruits, seasonings, non-alcoholic beverages, food taken away from home, purchased meals, etc.
2. Non-food consumption (18 items)
Clothing and footwear, personal care, communication, transportation, household equipment, recreation, education, domestic salaries, etc.
3. Housing consumption (12 items)
Utility, house rent, maintenance of dwelling, etc.

Source: Cambodia Socio-Economic Survey 2014 (33), Guidance for Constructing Consumption Aggregates for Welfare Analysis (41) and User's Manual for Handling Resampled Micro Data of CSES 2009 (42)

Based on the previous discussions in similar studies (11-23), 369 predictor variables were created with the CSES data. Table 3 shows a summary of the predictor variables.

Table 3 Summary of predictor variables

Residential area
Province; urban/rural settings
Household members' characteristics
Sex, age, ethnicity and educational level of household members; household size; dependent rate; total working hours
Real estate property
Number, area and use of own land; number, area, use and price value of own buildings; investment on buildings
Housing/living conditions
Size and construction materials of the dwelling; source of lightening; source of drinking water; type of toilet; utility charges; consumption of luxury food
Land use
Number, area and use of land parcels operated
Farming activities
Harvested land area; production; type of livestock, fishery and forestry activities
Durable goods
Possession, number and newness of durable goods in both urban and rural settings
Work
Type of employer; employment status; occupation; type of industry
Income and liabilities
Type of income; number and amount of loans
Survey year

The data were divided into a training set and a test set using the 8:2 ratio randomly. Subsequently, the analyses were conducted in two steps.

In the first step, using the training set, we constructed linear regression models that related a set of predictor variables (X) to observed household consumption (y), the value reported in the CSES, as follows:

$$y_t = \sum_k \beta_k X_{k,t} + e_t$$

where βk is a coefficient parameter to be estimated using ordinary least squares (OLS) method and is the error term, which is assumed to follow the normal distribution. With the new information on the predictor

variables in time t + 1, the corresponding household consumption $(\hat{y}_{t+1})$ can be predicted by plugging the estimated parameters into the above equation.

For the linear and mixed-effects models, we screened all the variables using a partial correlation coefficient with significance at the 90% or higher level as the cut-off point (23). We manually selected predictor variables using a backward-elimination technique to construct Model A (Manually-selected Linear Model). We also used the backward-selection technique within a stepwise regression analytical framework with a 0.1 level of significance as the cut-off point for removing variables (23) to construct Model B (Stepwise Linear Model). Subsequently, we constructed Model C (Mixed-effects Linear Model) with the remainder of the stepwise selection, considering a random effect across the same province. To avoid overfitting, we constructed Model D with elastic net regression, which was finally functioned with L1 penalty term of the regression coefficients, which was known as least absolute shrinkage and selection operator (LASSO) regression. In addition, we made it adaptive LASSO by adding data-dependent weights to obtain more unbiased estimates. Ten-fold cross-validation was used to select the regularization parameter in the LASSO model (43, 44). We used all the available predictor variables for Model D since adaptive LASSO can automatically perform the variable selection to improve the prediction performance and interpretability of the statistical model while ensuring the model parsimony.

In the second step, the trained models were applied to the test data. With this subset, we predicted the household consumption values, 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 three measures, namely mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). We used MAE because it evaluates prediction performance of the model most simply by taking the absolute difference between the actual and predicted values and finds the average as follows (45):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

RMSE squares the difference, finds the average of all the squares, and then finds the square root, as shown below. RMSE was additionally used because it is more sensitive to larger errors as it creates an exponential change in the base number by squaring the difference (45).

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}t)^2}$$

While MAE and RMSE are useful methods to compare the prediction performance of different models for the same dataset, they do not tell the relative performance of the prediction model itself. MAPE is the percentage of the error compared to the actual value according to the following equation (46), which provides more context to explain the model's average performance.

$$MAPE = \left\{\frac{1}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{|y_t|}\right\} \times 100$$

All analyses were conducted in Stata 16.0. A *P*-value <0.05 was considered statistically significant. The protocol of this study has been published elsewhere (47).

Results

Figure 1 shows the conceptual framework of predictor variable selection. Out of 369 predictor variables, 98 remained after removing variables with 0.1 or greater partial correlation coefficients. Subsequently, 51 predictor variables were selected for Model A, 86 for Model B and C, and 162 predictor variables remained for Model D. Supplementary Table 1 shows details of the remaining predictor variables and the coefficients in each model.

(Please insert Figure 1 here)

Figure 1. Conceptual framework of predictor variable selection

Figure 2 shows scatter plots of observed versus predicted household annual consumption values with the four alternative prediction models. Logarithmic transformation was performed for the outcome values. Overall, a positive linear relationship between observed and predicted household annual consumption was found in all four models, with the data points concentrated along the regression fit lines. The relationship was stronger for households with lower consumption, and it declined as the level of household consumption increased. There was a subtle trend that the middle-class households' consumption was likely to be underestimated. In contrast, that of the high-class households' consumption was over-estimated in all four models, while the trend was less noticeable in Model D.

(Please insert Figure 2 here)

Figure 2. Observed vs. predicted household annual consumption in Cambodia in 2010-2017

Table 4 shows the MAE, RMSE, and MAPE values of the four alternative prediction models. It should be noted that MAE and RMSE are expressed in logarithmically converted Cambodian riel. All these statistical measurements with smaller values are preferred. The smallest mean absolute error, MAE of 0.227 was calculated for Model B, followed by Model C with 0.228, Model D with 0.230, and Model A with 0.242. The trend was not different for RMSE, which should react more pronouncedly to larger errors, with the values of 0.301 for Model B, 0.302 for Model C, 0.305 for Model D, and 0.320 for Model A. The percentage of the predictive error compared to the observed value, MAPE was 1.376% for Model B, 1.380% for Model C, 1.394% for Model C, and 1.469% for Model A. The rank was consistent with all three statistical measurements.

Table 4 Prediction performance of alternative predictive models (95% confidence intervals)

	Model A	Model B	Model C	Model D
	Manually-selected Linear Regression	Stepwise Linear Regression	Mixed-effects Linear Regression	Adaptive LASSO Regression
MAE (95% CI)	0.242 (0.238 0.247)	0.227 (0.223 0.231)	0.228 (0.223 0.232)	0.230 (0.225 0.234)
RMSE (95% CI)	0.320 (0.312 0.327)	0.301 (0.293 0.309)	0.302 (0.294 0.309)	0.305 (0.297 0.313)
MAPE (95% CI)	1.469 (1.441 1.497)	1.376 (1.349 1.402)	1.380 (1.354 1.406)	1.394 (1.367 1.421)

Notes: LASSO: least absolute shrinkage and selection operator, MAE: mean absolute error, RMSE: root mean squared error, MAPE: mean absolute percentage error, CI: confidence interval

Discussion

This study found that it is possible to predict household consumption at a reasonable level, with a pool of highly predictive indices. The final product of the study will be an automated tool with selected predictor variables and respective regression coefficients, which will further determine the optimal amount of health insurance contribution for each household. Moreover, our approach would suggest a possibility of an equitable contribution collection from all socioeconomic groups of the society, while ensuring the feasibility of the insurance fund by allowing informed planning through an accurate estimation of the revenue pool. Incorporating our predictive model into the existing social insurance system in Cambodia will enhance the country's current efforts to prevent catastrophic health expenditure and achieve UHC targets.

While the four alternative prediction models had different functions, there was no significant difference in the results, particularly among Model B, C and D. The regularization technique with consideration of data-dependent weights in Model D, and the inclusion of random effects in Model C were not particularly effective in this environment. Among the three models with better predictability, Model B and C were more parsimonious with 86 predictors, as compared to Model D with 162 predictors. Parsimoniousness of the model is an important criterion in the model selection because the number of covariates yields the size of the questionnaire. These findings suggest that Model B would best suit the situation in Cambodia. Although Model B was not the most parsimonious, the number of questions could be curtailed as multiple variables are attributed to one information source. For example, the question about the floor material of one's dwelling was used to create five predictor variables. It is expected that the number of questions required in Model B could be boiled down to as few as 56.

In this study, the household consumption was expressed in the logarithmic scale. When the logarithmic transformation is returned in practice, the MAE in Model B is interpreted to be 4151 thousand Cambodian riels, which is equivalent to USD 1021.15, meaning that there is a mean error of USD 1021.14 in the average household annual consumption of USD 4231.22. This result of the error is further interpreted in the context of insurance contribution. Suppose the contribution rate is 3% of the non-subsistence household consumption, the MAE on the annual household contribution would be USD 30.63, which is USD 0.57 per person per month. Also, the model predictability is generally better for poorer households who could be threatened by overestimating their ability to pay. Therefore, the negative impacts of using this tool on the insurance contribution determination are not expected to be large.

The proxy means test has been practiced in Cambodia to identify poor households as beneficiaries of the social assistance programs, including the above-mentioned Health Equity Fund (17). The proxy means test is carried out based on the questionnaires that consist of scoring and non-scoring proxy indicators that differentiate poor households from non-poor households (17). The household consumption prediction model developed in this study is essentially different from the proxy means test. While the former predicts household consumption in monetary form, the latter merely assesses the level of poverty by scoring households. In addition, the results of the proxy means test are verified through the discussions in the community (26), but prediction performance of the tool has not been regularly evaluated. Therefore, the proxy means test cannot be used for the insurance contribution determination because it does not provide the reliable information on how much a household earns or spends, which is necessary when the insurance contribution is equitably collected. On the other hand, the reverse might be possible. That is, the household consumption prediction model could be used for both the poor household identification and the insurance contribution determination. If the feasibility of this model is proved, it is worth trying to use the model for the dual purposes to make the Cambodian social security system more efficient.

Despite our innovative methodology to estimate household consumption on a monetary basis, there are some practical limitations. First, this study compared the predicted household consumption with the observed values. However, the observed values were not real household consumption, but estimated

values based on the survey data. Therefore, the out-of-sample validation was performed based on the assumption that the survey data were reliable. Thus, the validity of the model is dependent on prediction performance of the survey data. Second, this study has developed and validated predictive models. However, the models have not been applied in the real world. Therefore, the applicability of the models, whether the predicted values are accepted by people in the community or key stakeholders, should be further assessed in practical settings. Finally, recall and social desirability biases may be an issue when the information was collected for household consumption prediction. Future studies should estimate the health insurance contribution of each household using this household consumption prediction model and compare the expected health expenses with the current practice of out-of-pocket payment.

Conclusion

Estimating the general population's household consumption is an important step towards achieving UHC, particularly for LMICs that wish to adapt contributory social health insurance universally. This study suggests the possibility of developing a regression model with the existing survey data in Cambodia to predict the household consumption at a reasonable level. The strategy would enable the country to universally introduce contributory social health insurance with the equitably collected revenue that has less political influence than the general tax revenue. The prediction model should be tested in real settings, periodically re-evaluated, and continuously improved. It is expected that the experience in Cambodia will help other LMICs improve their financial protection policies.

Declarations

Ethics approval and consent to participate

Not applicable. Ethical approval is not required because the study will use data, documents, and publicly accessible records, and all individual data are non-identifiable.

Consent to publish

Not applicable.

Availability of data and materials

The datasets that support this study's findings are available from the National Institute of Statistics (NIS), Ministry of Planning of Cambodia. Data are available upon reasonable request and with permission of the NIS.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

HN conceived of and designed the study, drafted the protocol, and acquired the data. HN analysed the data, interpreted and discussed the results, and drafted the final manuscript. FA and DY designed the statistical framework and supported data analyses. SY and ST confirmed the quality of the data and study design considering local contexts and previous studies in the field. TY and MT ensured that any part of the Cambodian social protection policy framework is accurately and appropriately described. MR and AI confirmed the study contexts in universal health coverage in LMICs. SN was responsible for the integrity of the data and prediction performance of the data analyses, and oversaw the study. All the authors made critical revisions to the manuscript for important intellectual content, approved the final version of the manuscript, and supported the interpretation and discussion of findings.

References

1. The United Nations General Assembly. Resolution adopted by the General Assembly on 25 September 2015 - Transforming our world: the 2030 Agenda for Sustainable Development. New York. 2015
2. World Health Organization, The World Bank. Global Monitoring Report on Financial Protection in Health 2019 Advance Copy. 2019.
3. International Bank for Reconstruction and Development, The World Bank. High-Performance Health Financing Universal Health Coverage: Driving Sustainable, Inclusive Growth in the 21st Century. Washington DC; 2019.
4. Nitayarumphong S. Universal Coverage of Health Care : Challenges for th Developing Countries
5. Shimazaki K. The Path to Universal Health Coverage - Experiences and Lessons from Japan for Policy Actions -. National Graduate Institute for Policy Studies; 2013.
6. Maeda A, Araujo E, Cashin C, Harris J, Ikegami N, Reich MR. Universal Health Coverage for Inclusive and Sustainable Development A Synthesis of 11 Country Case Studies. Washington D.C.: The World Bank; 2014.
7. Gumber A. Health insurance for the informal sector: Problems and prospects. New Delhi: Indian Council for Research on International Economic Relations; 2002. Contract No.: Working Paper No. 90.
8. Chen M, Palmer AJ, Si L. Improving equity in health care financing in China during the progression towards Universal Health Coverage. BMC Health Serv Res. 2017;17(852).
9. Mukangendo M, Nzayirambaho M, Hitimana R, Yamuragiye A. Factors Contributing to Low Adherence to Community-Based Health Insurance in Rural Nyanza District, Southern Rwanda. J Environ Public Health. 2018;2018.
10. Shimazaki K. Health Care in Japan: Institutions and Policies [revised edition]. Tokyo: University of Tokyo Press; 2020.

11. El-Gilany A-H, El-Wehady A, El-Wasify MA. Updating and Validation of the Socio-economic Status Scale for Health Research in Egypt. Eastern Mediterranean Health Journal. 2012;18(9):962-8.
12. Gaur KL. Socio-Economic Status Measurement Scale: Thirst Area With Changing Concept For Socio-Economic Status. International Journal of Innovative Research and Development. 2013;2(9):139-45.
13. Singh T, Sharma S, Nagesh S. Socio-economic status scales updated for 2017. International Journal of Research in Medical Sciences. 2017;5(7):3264-7.
14. Naga RA, Burgess R. Prediction and Determination of Household Permanent Income. 2001.
15. Vollmer F, Alkire S. Towards a Global Assets Indicator: Re-assessing the Assets Indicator in the Global Multidimentional Poverty Index. OPHI Research in Progress 53a, Oxford Poverty and Human Development Initiative2018.
16. Ngo DKL. A theory-based living standards index for measuring poverty in developing countries. Journal of Development Economics. 2018;130:190-202.
17. Identification of Poor Households Programme, Ministry of Planning. Cambodia national poverty identification system 2019 [Available from: https://www.idpoor.gov.kh/reporting/builder.
18. The Social Protection & Labor Team, The World Bank Group. Measuring income and poverty using Proxy Means Test [August 4, 2020]. Available from: https://olc.worldbank.org/sites/default/files/1.pdf.
19. Tajik P, Majdzadeh R. Constructing Pragmatic Socioeconomic Status Assessment Tools to Address Health Equality Challenges. Int J Prev Med. 2014;5(1):46-51.
20. Chasekwa B, Maluccio JA, Ntozini R, Moulton LH, Wu F, Smith LE, et al. Measuring wealth in rural communities: Lessons from the Sanitation, Hygene, Infant Nutrition Efficiency (SHINE) trial. PLoS One. 2018;13(6).
21. Ferguson BD, Tandon A, Gakidou E, Murry CJL. Estimating Permanent Income Using Indicator Variables 2003.
22. Chakrabarty N, Biswas S. A Statistical Approach to Adult Census Income Level Prediction.
23. Benin S, Randriamamonjy J. Estimating Household Income to Monitor and Evaluate Public Investment Programs in Sub-Saharan Africa. Washington D.C.: International Food Policy Research Institute; 2008.
24. The World Bank. World Bank Country and Lending Groups 2019 [August 13, 2020]. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups.
25. The Royal Government of Cambodia, Ministry of Labour and Vocational Training. Subdecree on Establishment of Social Security Scheme on Health Care for Persons Defined by the Provisions of Labour Law Royal Government. 2016.
26. The Kingdom of Cambodia, Ministry of Health. Health Equity Fund Operational Manual. Phnom Penh, Cambodia 2016.

27. National Social Security Fund. Achievement Over the Period of 10 years (2008-2017) and Direction for 2018. The NSSF 10-year Anniversary 2018.
28. The Royal Government of Cambodia. National Social Protection Policy Framework 2016-2025. Phnom Penh 2017.
29. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2010. 2011.
30. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2011. 2012.
31. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2012. 2013.
32. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2013. 2014.
33. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2014. 2015.
34. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2015. 2016.
35. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2016. 2017.
36. National Institute of Statistics, Ministry of Planning, the Kingdom of Cambodia. Cambodia Socio-Economic Survey 2017. 2018.
37. The World Bank Group, International Monetary Fund. Consumer price index (2010=100) - Cambodia, International Financial Statistics and data files 2020 [August 13, 2020]. Available from: https://data.worldbank.org/indicator/FP.CPI.TOTL?locations=KH.
38. Murry CJL, Knaul F, Musgrove P, Xu K, Kawabata K. Defining and measuring fairness in financial contribution to the health system. World Health Organization; 2000.
39. O'Donnell O, Doorslaer Ev, Wagstaff A, Lindelow M. Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation. Washington D.C.: The World Bank; 2008.
40. Wagstaff A. Catastrophic Medical Expenditures Reflections on Three Issues. Washington DC: The World Bank; 2018.
41. The International Bank for Reconstruction and Development, The World Bank. Guideline for Constructing Consumption Aggregates for Welfare Analysis. Washington D.C. 2002.
42. The Institute of Statistical Mathematics (ISM), Statistical Information Institute for Consulting and Analysis (SINFONICA). User's Manual for Handling Resampled Micro Data of Cambodia Socio-Economic Survey (CSES) CSES 2009 Version 2.0. 2016.
43. Santosa F, Symes WW. Linear Inversion of Band-Limited Reflection Seismograms. SIAM Journal on Scientific and Statistical Computing. 1986;7(4):1307–30.

44. Tibshirani R. Regression Shrinkage and Selection via the lasso. Journal of the Royal Statistical Society. 1996;58(1):267–88.
45. Greene WH. Economic Analysis. 8th Global ed. Harlow, England Pearson Education Limited; 2020.
46. De Myttenaerea A, Golden B, Le Grand B, Rossic F. Mean Absolute Percentage Error for Regression Models. 2017.
47. Nakamura H, Amimo F, Yi S, Tuot S, Yoshida T, Tobe M, et al. Implementing a sustainable health insurance system in Cambodia: a study protocol for developing and validating an efficient household income-level assessment model for equitable premium collection. International Journal for Equity in Health. 2020;19.

Figures

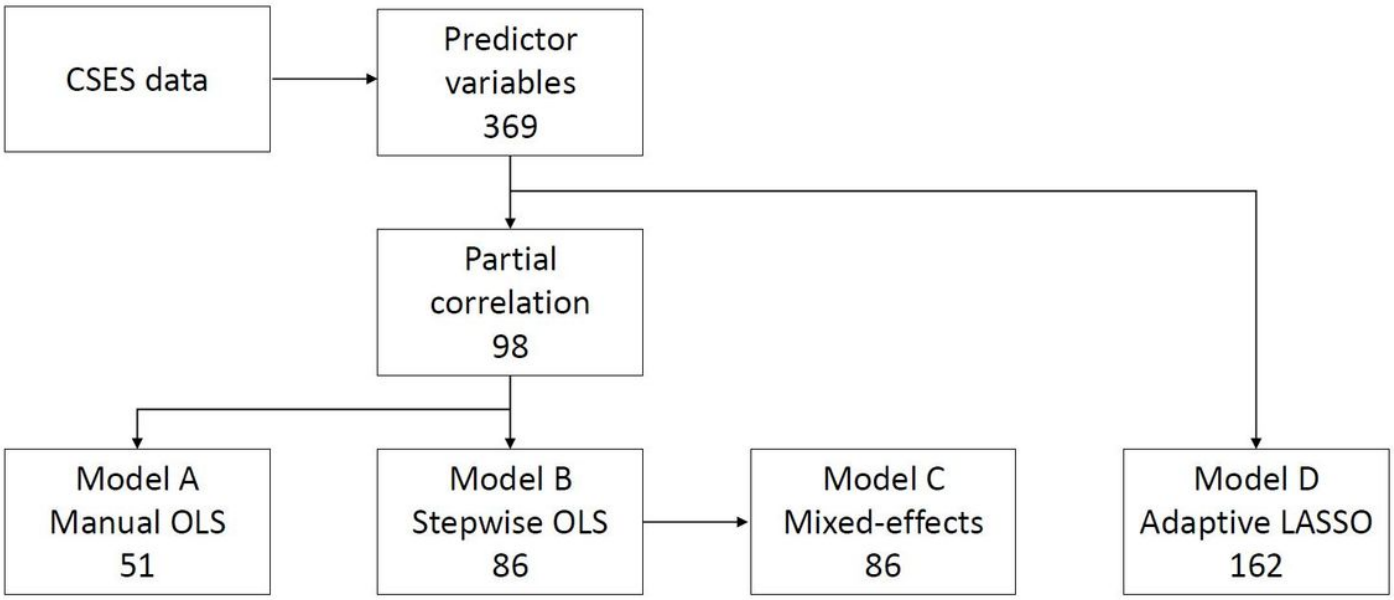

Figure 1

Conceptual framework of predictor variable selection

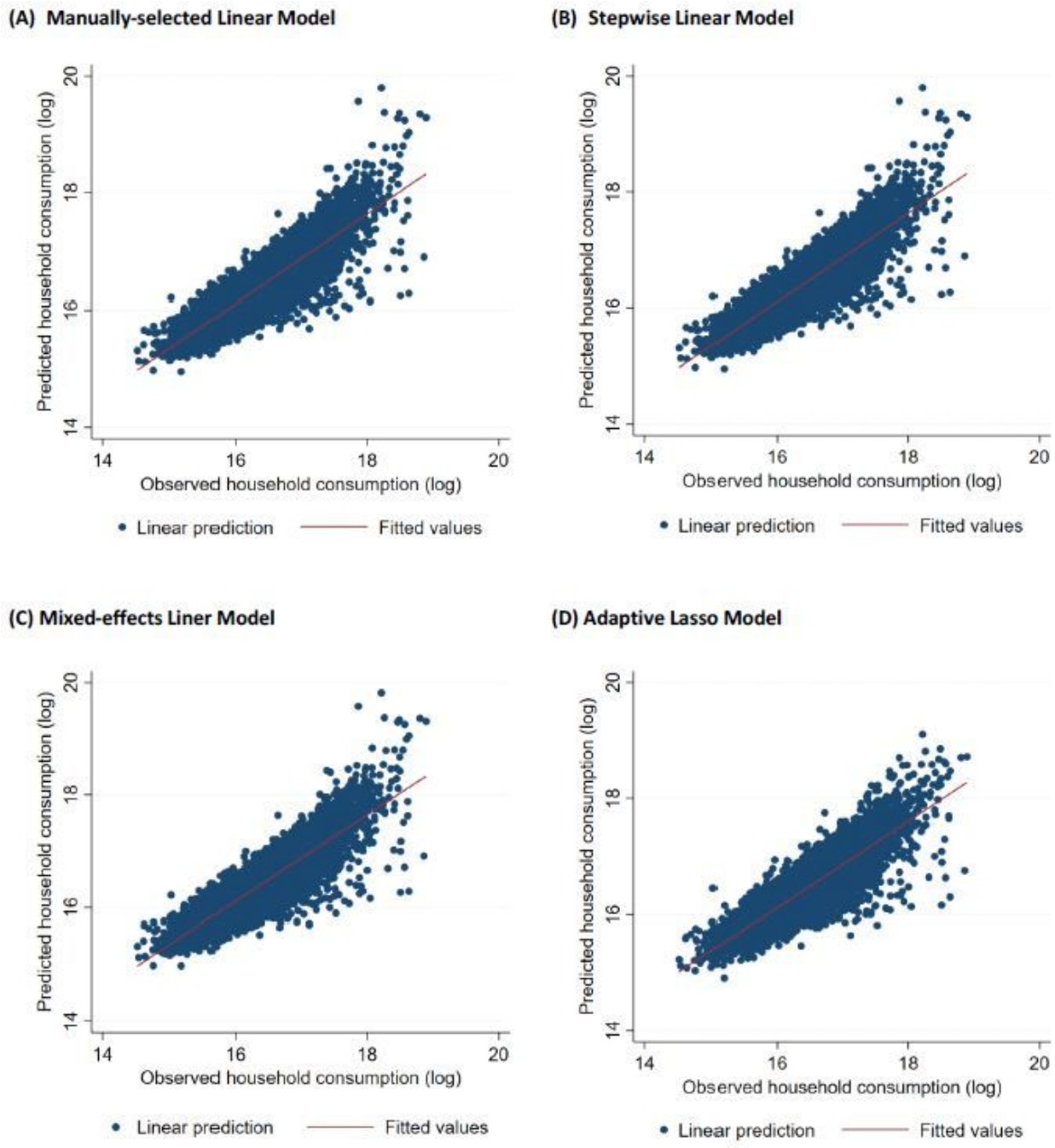

Figure 2. Observed vs. predicted household annual consumption in Cambodia in 2010-2017

Figure 2

Observed vs. predicted household annual consumption in Cambodia in 2010-2017

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytable1Nakamura.pdf