Small-Area Multidimensional Poverty Estimates for Tonga 2016: Drawn from a Hierarchical Bayesian Estimator

Héctor E. Nájera Catalán [1,2,3] · Viliami Konifelenisi Fifita [4] · Winston Faingaanuku [4]

Abstract

Tonga has recently adopted the consensual approach to produce its official multidimensional poverty measure. This index is computed using data from the Household's Income and Expenditure Survey (HIES). The population of Tonga is scattered across 5 main groups of islands and high-quality spatial data is vital to inform policies. One limitation is that HIES data originate from a nationally representative survey that cannot produce reliable estimates for small areas such as constituencies, villages or blocks, and governments require highly disaggregated data to better inform policies, for example, with regard to natural disasters. This paper produces small-area estimates of poverty based on a hybrid hierarchical Bayesian (HHB) estimator, which draws on the standard hierarchical Bayes (HB) approach but uses a more efficient computation process – Hamiltonian (hybrid) Monte Carlo (HMC) – to produce the posterior distributions. The HHB estimator is then applied to Tonga's National Population Census (2016) to present estimates down to island, constituency and block level. The results suggest that the extent of poverty is lower in Tongatapu than in Eua, Vava'u, Ha'apai and Niuas, while its prevalence is very similar (around 35%) in Eua, Vava'u and Ha'apai. Constituencies in Tongatapu show lower poverty rates than in the rest of the islands, and block-level data show a clear spatial pattern of poverty distribution in the capital Tongatapu. These are the first small-area indirect poverty estimates based on a hierarchical Bayesian model and drawn from the consensual approach (CA).

Keywords Multidimensional poverty · Small-area estimation · Deprivation · Consensual Approach · Spatial correlation

✉ Héctor E. Nájera Catalán
  hecatalan@hotmail.com

Extended author information available on the last page of the article

🖄 Springer

Introduction

The Kingdom of Tonga has recently adopted the consensual approach (CA), which draws on Townsend's theory of relative deprivation and has been applied in both high income and low-income countries, to produce its official poverty estimates (Guio et al. 2012; Guio et al. 2017; Lau et al. 2015; Nandy and Pomati 2015; Pantazis et al. 2006; Townsend 1979). The CA follows the theory of relative deprivation in that it considers the socially perceived necessities of the population -what they think is necessary according to the living standards of the society they belong to- and identifies material and social deprivation based on the enforced lack of such needs -it distinguishes preferences from constraint- (Mack and Lansley 1985).

Tonga has been the first small island state in the South Pacific region to adopt an official multidimensional poverty measure to report the sustainable development goal (SDG) indicator 1.2 which aims to: "By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions". This is a major shift from the widely used calorie-based measure in that the CA is a multidimensional approach that focuses on people deprived of socially perceived necessities in the society to which they belong (see section 1) (Mack and Lansley 1985; Townsend 1979).

According to the official estimates, multidimensional poverty affects around 27% of the population in Tonga (Konifelenisi et al. Forthcoming). These estimates are computed using the Household and Income and Expenditure Survey (HIES 2015/2016). One of its limitations is that its sample design only produces representative estimates down to rural and urban areas, the capital Tongatapu and a group of four islands (Vavau', 'Eua, Ha'apai, Ongo Niua). That means that the HIES (2015/2016) is incapable of producing reliable multidimensional poverty estimates for each island, each constituency, village or block within island. This is a major disadvantage in that the location and concentration of poverty is vital in designing anti-poverty policies and responses to assist the most vulnerable and affected inhabitants when catastrophic events hit the islands, such as the recent Cyclone Gita in 2018.

Small-area estimation (SAE) is a field in statistics that comprises a series of methods to address the limitations of survey data to produce reliable estimates of poverty for different geographical locations (Pfeffermann 2013; Pratesi 2016; Rao and Molina 2015). SAE aims to accommodate the available ancillary data -Census data, for example- in the best possible way to estimate poverty for specific locations. This task involves exploiting between-area differences and within area similarities and the SAE literature has made both theoretical and computational breakthroughs in the modelling of outliers, dealing with categorical data, accounting by for the location of the unit of analysis, including both area and individual-level data and considering the sampling design of household surveys in the estimation (Rao and Molina 2015).

From an applied perspective, the small-area estimation of poverty is yet to be systematically applied for either unofficial or official measurement campaigns in developed countries and is seldom implemented in developing countries (EURAREA 2012; Haslett and Jones 2010; Pratesi 2016). Tonga has never had small-area estimates of poverty and in the developing world, poverty estimates have been mainly produced for income-poverty measures using the Elbers et al. (2003) or World Bank method (Haslett and Jones 2010). However, the World Bank method is not as good as others in

most circumstances. The experimental SAE literature has shown that his method is very sensible to between-area heterogeneity and that it tends to be outperformed by methods such as the Hierarchical Bayes (HB) and the Empirical Best Linear and non-linear Predictor (EBLUP) (Guadarrama et al. 2014; Haslett and Jones 2010; Rao and Molina 2015).

SAE requires fitting several hierarchical models in order to find a good predictive model. Modern reliable estimators such as the EBLUP rely on Maximum Likelihood (ML), which becomes very time consuming for increasingly complex models. Moreover, because multidimensional poverty measures are discrete-, computation becomes prohibitively time-consuming for estimators that rely on ML estimation such as the Empirical Bayes. This poses a challenge for the national statistical offices in that SAE requires fitting a series of often increasingly complex models that with ML become infeasible at some point (Guadarrama et al. 2014). The HB estimators are more efficient given that they do not rely on numerical integration -unlike ML- but its implementation is even more rare in official statistics in developing countries given that these methods are fairly new and there are few examples with real data in the literature (Nájera 2019).

The widespread implementation small-area estimation is necessary to cover the demand of more detailed geographical poverty data, but access to the most recent theoretical developments and both technical and computational difficulties hinder the systematic implementation of the best estimation approaches in developing countries (Haslett and Jones 2010). This is the case of Tonga and most of the islands in the South Pacific region where small-area estimation has not been implemented. The aims of the study are to compute for the first time small-area multidimensional poverty estimates for Tonga, advance the production of these small-area estimates for poverty measures that rely on the consensual approach and contribute to the illustration of the advantages of using a Hierarchical Bayesian estimator that relies on novel computational tools. This with the purpose to enhance the available data for policy makes in Tonga and to set a reference for countries with similar kinds of data.

The document is organised as follows. The first section briefly reviews the underlying theory behind the consensual approach and presents the characteristics of the official multidimensional poverty measure used in Tonga. The second section describes the strategy used to produce the indirect small-area estimates using the Hierarchical Bayes estimator. The kinds of models and the variables utilised for the SAE procedure are also described. The third section presents a descriptive analysis of income, deprivation and poverty in Tonga, using the HIES (2015/2016) data, and a brief discussion is included about how income differences between Tongatapu and the rest of the areas might have an impact on the estimates. The fourth section presents the results of the analysis, and these are presented in the following order: islands, constituencies and block-level estimates. Figures and maps are provided when relevant. The last section concludes the document and discusses pending tasks about the use of these estimates and the production of further analyses for specific population groups.

Multidimensional Poverty Measurement in Tonga

Poverty is a multidimensional phenomenon that can be defined as the lack of command of resources over time and material and social deprivation are its main consequences

(Gordon 2006). Townsend (Peter Townsend 1987) put forward the theory of relative deprivation which states that people are deprived when they lack the goods and services that are customary in the society to which they belong (Peter Townsend 1987). This theory is the core of the consensual approach (CA), which is a method with over 50 years of methodological developments (Mack et al. 2013; Pantazis et al. 2006). The CA considers the socially perceived necessities of the population and identifies material and social deprivation based on the enforced lack of such needs (Mack and Lansley 1985). The CA method has been applied successfully in the UK (Gordon 2018; Pantazis et al. 2006), is the official measure of the European Union (Guio et al. 2012; Guio et al. 2017; Guio et al. 2016), has been applied in many other countries (Halleröd 1995; Lau et al. 2015; Nandy and Pomati 2015; Perry 2002) and is now the official multidimensional poverty measure in Tonga (see description below) (Konifelenisi et al. Forthcoming).

Drawing upon Townsend's theory of relative deprivation, the consensual approach relies on a survey module -the only one explicitly designed to measure poverty- to identify the relevant deprivation indicators as follows: 1) it considers the necessities of the population, 2) identifies the households that cannot afford socially perceived needs and 3) draws on classic and contemporary statistical literature to derive a *suitable, reliable, valid and additive* multidimensional poverty index (Guio et al. 2017, Guio et al. 2016); and 4) combines income and deprivation to identify the poor population (Gordon 2006). Therefore, consistent with classic and contemporary measurement theory, the CA uses a direct approach to assess people's living standards (deprivation), which proves to be consistent (reliable) and measures what it is meant to measure (valid).

A total of 13 deprivation were identified to be *suitable, reliable, valid and additive* in Tonga using the HIES 2015/2016 (Konifelenisi et al. Forthcoming). The number of deprivations were counted (deprivation score) for everyone in the sample. Therefore, each person in the sample has a deprivation score which ranges between 0 and 13. The aggregation method follows Guio et al. (2017) approach, which relies on equal weighting. As it is now from numerical experiments, an index is self-weighting once reliability holds, i.e. the ordering of the population does not change under differential weighting for a reliable scale (Nájera 2018).

Multidimensional poverty is computed by combining the deprivation score with the household's income per capita following Townsend's prediction of the relationship between income and deprivation (Gordon 2006; Townsend 1979). This serves to identify the "truly or consistently" poor using Townsend's breaking point, i.e. the certain level of resources (approximated via income per capita) from which deprivation rises substantially (Gordon 2006; Townsend 1979). This prediction states that income should be non-linearly correlated negatively with the severity of deprivation (count of deprivations) and there should be an inflection point – level of income – at which people show similar levels of deprivation. This provides a split into two meaningful groups, namely 'the poor' and the 'not-poor'. In practice, drawing upon this theory, poverty is statistically identified via the optimal method (based on Generalized Linear Models), which finds the best possible split between the two groups, i.e. the level of income and deprivation count that maximises the differences between the two subpopulations.

For the analyses Konifelenisi et al. (Forthcoming), used unadjusted per capita income, adjusted per capita income by the square root of the number of household members and adjusted per capita income by the OCDE equivalence scales. They found

the same income deprivation cut-off using the three income measures: the mean income of the group with 3 or more deprivations. The poor are those people living with less than 677.6 Pa'anga per capita -adjusted by the square root of the total household members- per month and with three or more derivations. This binary variable is utilized for the small-area estimation procedure (see Table 1).

Data, Methods and Model Fitting Strategy

Data

This paper uses microdata from two main sources. First, microdata from the Household Income and Expenditure Survey (HIES 2015/2016), which is a nationally representative survey with a complex design, and, for the first time, the consensual module, to measure multidimensional poverty (Tonga Statistics Department 2016). Access to this data was obtained from the Tonga Statistics Department and is available upon request. The sample size of the HIES is 1805 households (10,319 individuals) and was collected from the five main islands: Tongatapu, Vavau', 'Eua, Ha'apai and Ongo Niua, totalling 165 villages and 579 blocks. Nonetheless, the sample is only representative of Tongatapu and the group of the four outer islands.

The Statistics Department of Tonga provided access to the full microdata from the Tongan Population and Housing Census ((Tonga Statistics Department 2017). The

Table 1 Predictors utilized in the hierarchical Bayesian models. Response variable: Multidimensional poverty (poor v not poor)

Predictor	Categories / Type
Household head education attainment	Ordinal (six categories)
	No education
	Primary
	Low secondary
	Secondary
	Technical education
	Tertiary
Household head type of employment of economic involvement	Nominal (four categories)
	Paid work or business owner
	Own-production and other
	Unpaid worker
	Other
Household size	Continuous
Floor deprivation	Binary
TV deprivation	Binary
Fridge deprivation	Binary
Rooms per people	Continuous
Rural location	Binary

population census was utilised for predictions at the small-area level (see details below) and collected information on the 22,021 households in Tonga (total 104,474 inhabitants) (Tonga Statistics Department 2017). The HIES and the Census have many common socio-economic and demographic variables that can be used to link the parameters obtained from the hierarchical model. The distribution of the variables was compared to assess its comparability, but variables exhibiting significant discrepancies (i.e. the Census value was not included in the confidence intervals of the HIES variable) were not used in the analysis (see the section on the model fitting strategy for the list of variables included in the model).

Methods

SAE approaches are divided into two main groups: *direct or design-based* and *indirect or model-based* estimators. A direct estimator presents point estimates (e.g. the poverty rate for an area) accompanied by a measure of uncertainty such as confidence intervals or the variance. However, when using survey data, for smaller areas the sample size is either very small or zero, making the resulting poverty estimates too inaccurate to be considered informative for policymaking.

Indirect estimation relies on different sorts of regression models that accommodate auxiliary data (i.e. area-level education rates, area-level averages of different socio-economic variables, etc) in the best possible way to account for the different sources of uncertainty and bias. There have been different proposals in the contemporary SAE literature that take into account available ancillary information in different ways (Chambers and Tzvidis 2006; Elbers et al. 2003; Gelman and Little 1997; Pratesi 2016; Rao and Molina 2015). Since information is available at different levels (i.e. area- and individual-level), hierarchical or multi-level models are suited to the task, as they permit to incorporate into one model household-level data, area-level information and an estimate of uncertainty for the model, i.e. the unaccounted variation by the variables included in the model (Gelman et al. 2014; Goldstein 1999).

Hierarchical modelling has permitted the expansion of indirect estimators (Pfeffermann 2013; Rao and Molina 2015), with the empirical best linear -or non-linear for Generalized Linear Models-unbiased predictor (EBLUP) probably being the most well-known approach in SAE, where the Elbers et al. (2003) method can be considered as especial case of a EBLUP estimator with random effects (Guadarrama et al. 2014; Lahiri and Rao 1995; Militino et al. 2007; You and Rao 2002). This approach fits a multi-level model (based on maximum likelihood) and then uses a pseudo-Bayesian method (empirical Bayes) to estimate area-level effects which in turn are added to the individual-level prediction. Other recent approaches include the M-quantile spatial estimator, which is suited especially to model outliers (Chambers et al. 2014; Chambers and Tzvidis 2006). The hierarchical Bayes (HB) estimator has very attractive properties such as high flexibility, suitability for complex models, it is not limited to continuous variables and it tends to do a better job in shrinking point estimates (Rao and Molina 2015). The possibilities of the HB have been boosted recently by the Hamiltonian, or hybrid, Monte Carlo (HMC) procedure, which is a recent breakthrough in Bayesian computation, in that it permits estimating HB models more efficiently and quickly (Betancourt 2017; Carpenter et al. 2016).

This paper uses an indirect small-area estimator to produce poverty rates down to island, constituency and block level in Tonga. The different kinds of indirect approaches in SAE are thoroughly reviewed in Rao and Molina (2015) and, for a less technical overview, Pratesi (2017). SAE approaches take a target index, which in this paper is a binary indicator- poor v not poor and model such a response variable with different predictors. Because there are different approaches to do this, the pursue of flexible, efficient and robust estimators for categorical data has led to several kinds of approaches: the Empirical Bayes estimator - maximum likelihood-, the M-quantile-based estimator and the Hierarchical Bayes (HB) estimators (Molina and Rao, 2015, Chambers et al., 2016). Small-area estimation often is a problem of high dimensions, i.e. complex data, several parameters and large samples. This poses a trade-off between model complexity and computational power. For applied researchers is vital to find the best strategy as SAE involves fitting several models ad exploring different alternatives. Maximum likelihood estimators like the EB are not feasible for complex problems are numerical integration is very computational intensive. In contrast, the HB is more flexible and tends to be more computationally feasible for increasingly complex models (Chen et al. 2014; Hernandez-Stumpfhauser et al. 2016; Park et al. 2004; Rao and Wu 2010).

This paper relies on a hierarchical Bayes (HB) estimator (for the detailed formulation see Molina et al., (2014). The HB requires prior information about the distribution and features of the parameters of interest (i.e. mean, variance and distributional form- normal, t-student, Cauchy, etc.) and uses Markov Chain Monte Carlo (MCMC) computation to derive the posterior distribution of such parameters (Rao and Molina 2015). The HB estimator takes the form (Rao and Molina 2015, eq. 10.5.1 and eq. 10.5.5). These posterior distributions are probability density functions for each parameter (slopes and the variances of the random effects of the model) with a mass and density. From the theoretical point of view, in the very long run, the MCMC based on the Metropolis-Hastings algorithm will converge and produce the posterior distribution for each parameter (Gelfan et al. 2010; Gelman et al. 2014). However, two main problems might emerge for applied researchers. First, several models need to be fitted to find a good predictive model, and sometimes the MCMC can take hours or days to converge. Second, related to this first problem, is the fact that for high-dimensional settings (i.e. large datasets with complex models involving many parameters) the MCMC might not be very efficient in exploring the typical set (i.e. solution space) (Betancourt 2017; Neal 2011). The consequence is very slow performance and obtaining very low numbers of effective posterior samples.

Drawing upon Duane et al. (1987), the Hamiltonian, or hybrid, Monte Carlo (HMC) has been put forward recently as an improvement in Bayesian computation for high-dimensional problems over the MCMC (Betancourt 2017; Hoffman and Gelman 2014; Neal 2011). The HMC uses Hamiltonian dynamics to guide exploration of the typical set, leading to a more efficient estimating process that is particularly useful when having to fit several complex models that are sometimes difficult to fit with the standard MCMC approach. Therefore, whereas this paper relies on the standard HB estimator, it uses a more efficient estimating algorithm that makes model building more efficient and feasible. This estimation procedure has been implemented, for example, to produce small-area estimates of stunting in Mexico (Nájera 2019).

The HMC can be implemented easily thanks to the STAN project, which has produced a C++ function that can be run on standard statistical programmes such as R, Phyton, Stata, etc. (Carpenter et al. 2016). For this paper, we rely on the R-package "brms," which translates R-code into STAN, includes the possibility of including weights in the estimation (see further below) and offers many options for prediction using the same or an alternative sample (Bürkner 2017). We also use the "survey" R-package to produce the direct estimates (Lumley 2011, 2016), "ggplot2" for the plots and "ggmap" for the maps (Kahle and Wickham 2013; Wickham 2009).

Model-Fitting Strategy

The goal in small-area indirect estimation is to predict poverty using available information in the best possible way. The steps involved in SAE are summarized as follows and explained with more detail in the following paragraphs:

1. **Common variables:** The first task is to find a series of common potential predictors that are available in both the income and expenditure survey and the population census.
2. **Predict poverty with the survey data:** Using the preliminary list of predictors the second task consists in finding the best predictive model (see Table 1). In this case poverty is a binary variable. Therefore, it is necessary to fit several models (with Bernoulli distribution) and find a good predictive hierarchical model using the survey microdata. As discussed above, one could use maximum likelihood (logit) or Marcov-chain Monte Carlo (with Bernoulli distribution but these two approaches are equivalent under uninformative priors). This paper uses the Hamiltonian Monte Carlo because it is more flexible and faster than ML and more efficient than MCMC (see above).
3. **Model predictive accuracy:** A crucial step is to check that the best predictive model reproduces the direct estimates from the survey.
4. **Prediction with the population census data:** Apply the coefficients of the best predictive model to the microdata of the Census. Each unit in the Census will therefore have a probability of being "poor".
5. **Estimation of area-level poverty rates:** The prevalence poverty rate is calculated based on Rao and Molina (2015). They have proved that the mean probability at area-level, for example, blocks is equal to the "proportion of poor people".

From a statistical perspective the main goal in SAE is finding a parsimonious model that predicts poverty accurately, using individual and contextual auxiliary information. This model is based on the set of predictors that is common across both sources (the survey and the population census). This is essential in that the coefficients of the best model using the survey data are applied to each people in the population census. The second step in this study consisted in finding a series of individual-level, reliable predictors of poverty that were available and comparable in both the HIES and the population Census. This involved choosing socio-demographic and economic indicators and some deprivation indicators that had the same distribution and prevalence in the HIES and in the Census.

Then, the final list of predictors was selected based on a series of logit models using the whole sample (individual-level) of the HIES, where the response variable was multidimensional poverty according to the consensual approach (CA) consisting of two-categories: multidimensionally poor and not poor. These predictors were almost always significant and had high odds ratios. The list of predictors is presented in Table 1[1]:

Prior to the estimation of the hierarchical models (step 2 below), the survey weights of the HIES survey were rescaled following Carle (2009), because within-islands samples do not necessarily capture the distribution of poverty for each island, and so this adjustment can help reduce any bias (Carle 2009). As previously noted, Bayesian models have the advantage of using information about the parameters of interest for the estimation (Gelman et al. 2014). In this case, we did not have very strong information about the distribution, mean and variance of the parameters of the hierarchical models. Therefore, all the different models were fitted using both weakly informative (mean zero, large variance) and slightly stronger priors (mean zero, smaller variance) as means to assess the impact of the priors upon the results. There was no evidence of such an impact, meaning the models were highly stable.

After a set of stable and reliable level-1 predictors was found, a series of increasingly complex hierarchical models (random intercepts for blocks) were estimated to take into account the between-level information and improve the accuracy of the model. A summary of the main variants of the models is displayed in Table 1. The simplest model was M1, which included only the household-level predictors listed in the table; the rest of the models were hierarchical in the sense that they included island-level and block-level intercepts and random slopes. The best possible model among the alternatives was chosen based on three main criteria: a) the WAIC (widely applicable information criterion) statistic of fit, b) assessing the quality of the posterior distributions and effective samples and c) looking at its predictive accuracy based on the direct survey estimates, i.e. the prevalence of poverty for Tongatapu and for the cluster of four islands. This latter criterion meant applying the estimated coefficients to the characteristics of the population in both the HIES sample and the population Census (i.e. predicting poverty) and then computing the mean prevalence of poverty for Tongatapu and the four other islands.

The predictive power of almost all hierarchical models, using the HIES data, was very good in the sense that they replicated the prevalence of poverty for Tongatapu and the other four islands, using the HIES data. Differences increased when using the population Census for prediction, but in all cases, the mean prevalence rate was within the design-based estimate. However, in some instances, the estimate when considering the population Census was located on the lower or upper bound. Among all the models, M3 showed the best statistical properties (WAIC), it was simpler than models M4 and M5 and the estimated poverty rate was within the lower and upper bounds (see Table 2). M3 included random slopes for the employment categories and fridge deprivation

---

[1] The hierarchical models included some area-level fixed effects such as percentage of people without a flushing toilet and so forth, but none of these was significant, meaning that the level-1 information accounted for a large part of the variation.

Table 2 Model fit statistics and predictive capacity

Model	Reproduced direct estimate?	HIES WAIC
M1. Fixed Effects	YES	7873
M2. two-level (Blocks RI)	YES	6233
M3. two-level (Block RI and RS)	YES	4537
M5. three-level (Block and Island RI and Block RS)	NO	4711
M4. three-level (Block and Island RI and Island RS)	YES	6199

RI = Random intercept

RS = Random slopes (education, employment and fridge deprivation)

WAIC = Widely applicable information criterion. The lower, the better

(which has a high effect on the odds of being poor), meaning that these two variables had a specific coefficient for within each block.

The final step involved applying the coefficients from M3 to the population Census and, following Rao and Molina (2015) approach, obtaining the mean prevalence rate for the areas of interest, namely islands, constituencies and blocks. This procedure had the advantage that it could be applied for any geographical grouping variable included in the population Census.

Results

Descriptive Analysis

This section presents a descriptive analysis that contextualises small-area estimation challenges in the Kingdom of Tonga. The Bristol optimal method identifies poverty by using both income and deprivation (Pantazis et al. 2006; Gordon, 2010). The household per capita income and the deprivation score are considered to find an optimal national poverty line, i.e. the level of income and the number of deprivations, to identify poverty. The official poverty rate in Tonga, individual level using the consensual approach, is 25%.

One of the main features of Tonga and the islands in the south pacific is that households rely on different kinds of resources aside from income (see below). This is important to note, because income differences will be exacerbated by the kind of prevailing economy in each island. Figure 1 shows the mean income per capita according to the two main island groups – Tongatapu and then the rest of the islands, namely Vavau', 'Eua, Ha'apai and Ongo Niua. The mean income per capita in Tongatapu is much higher relative to the other islands, which explains partly why poverty is much higher outside the capital (see National Report). Figure 2 plots the distribution of the deprivation score (household items and adult items): the pattern is the same, but material deprivation is much higher in the cluster of the four islands relative to Tongatapu. These important differences have important implications for SAE, as they indicate that the prevalence of poverty is likely to fluctuate a lot between and within islands.
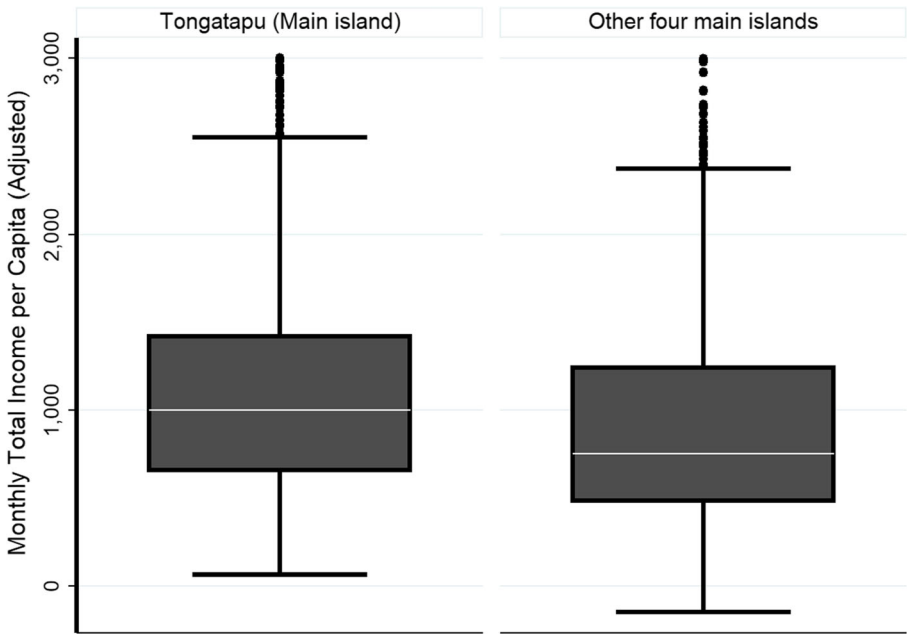
Fig. 1 Income per capita (adjusted by square root of the total household members) by island group, 2015
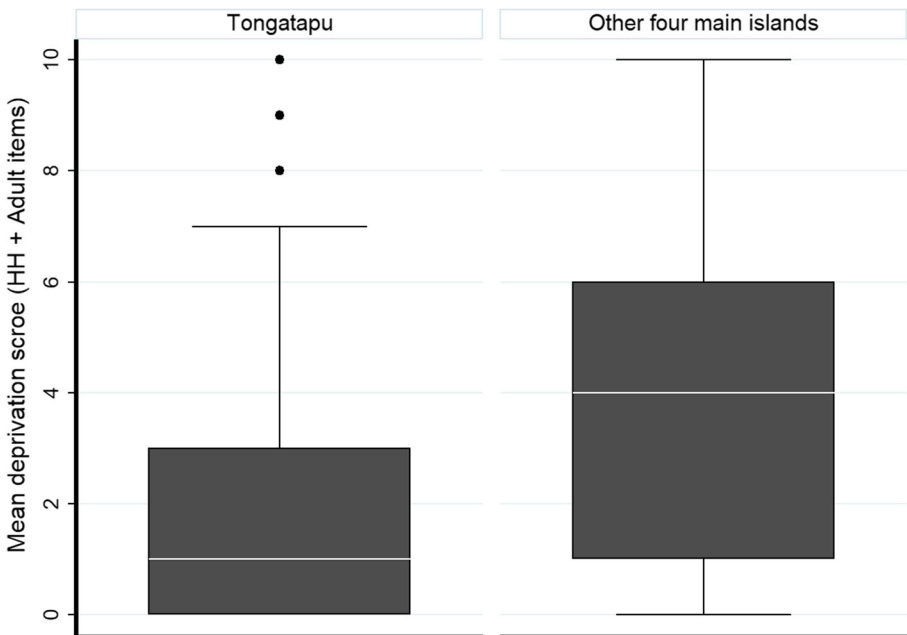

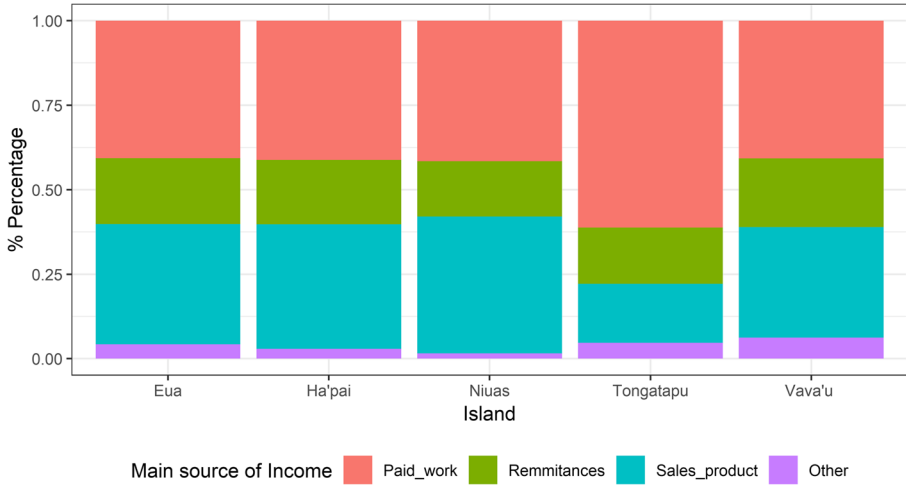**Fig. 2** Mean deprivation score by island group, 2015

Fig. 3 Main source of household income by island. Tonga. Census data

Figure 3 plots the main source of income as a means of assessing between-island heterogeneity and gaining a better idea of what can be expected from the SAE estimates. In Tongatapu, around 60% of income comes from paid employment, whereas in the rest of the islands this figure is around 40%. The other main difference in income composition is the proportion of income from sales of own-produced products. While in the other islands this kind of income accounts by for 35% of the total, in the capital it is less than 25%.
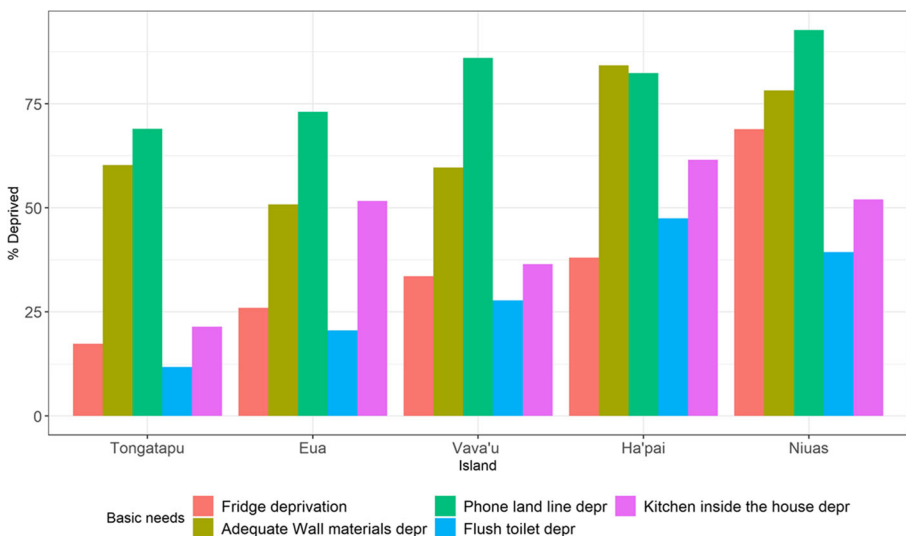
Fig. 4 Percentage of people deprived of different basic needs. Tonga. Census data

Table 3 Individual-level poverty rates by island

	Tongatapu		Rest (four islands)	
	Mean %	95% CI	Mean %	95% CI
Survey estimate	18	[17–19]	39	[37–41]
HHB survey data*	18	[17–19]	39	[38–40]
HHB Census data*	17	[16–18]	38	[37–39]

*Credible intervals draw from the posterior distribution of the model. HHB=Hierarchical Hybrid Bayes

The differences in the total income per capita and the income composition between islands, particularly between the capital and the rest, raise the question about whether this is reflected in the standard of living of the population between islands. As an approximation of the possible differences in standards of living between islands, a series of material deprivation indicators were computed, to gain a better idea of a possible pattern across islands. Figure 4 plots the deprivation rate for five indicators: fridge deprivation, living in a house with walls made of natural materials, phone line deprivation, lack of a flushing toilet and living in a house with an outside kitchen. In contrast with income, there is more variability across islands, which is consistent with the idea that income is not sufficient to assess the standard of living in Tonga. Tongatapu, Eua and Vava'u show lower deprivation rates for each item relative to the other two islands, which is an indication that the HIES sample produces direct estimates for each island that are highly unreliable, given that they seem to overestimate poverty in Eua and underestimate it Ha'apai.

Island-Level Estimates

The accuracy of the prediction of Model M3 is presented in Table 3. The survey (direct design- based) estimate is the prevalence of poverty after applying weights and considering the complex design of the HIES data. This is just the common reliable estimate of poverty for the capital and the other islands (see National Report). The second row in Table 3 shows the hybrid Bayes estimate, which is the model-based

Table 4 Individual-level poverty rates by island

	Survey (Unreliable)		HHB Census	
	Mean %	95% CI	Mean %	95% CrI
Tongatapu	16	[13–20]	17	[16–18]
Vava'u	47	[45–50]	37	[34–40]
Ha'apai	20	[17–22]	32	[29–35]
Eua	53	[49–58]	36	[33–39]
Niuas	79	[72–87]	54	[51–57]

*CrI. Credible intervals

estimate obtained after fitting the two-level hierarchical model and using the HIES sample for the mean prediction. Therefore, the HHB survey figure shows the capacity of the model to reproduce the survey-based rate using the same sample. This is an indication that the estimated parameters for each variable do a good job in reproducing the pattern of poverty in Tonga. The HHB (Census data) estimate (third row) shows the prevalence rate obtained after applying the coefficients to the population Census. Table 3 shows that the model can reproduce the results even after considering areas beyond the sample, thereby indicating the capacity of the model for reproducing the
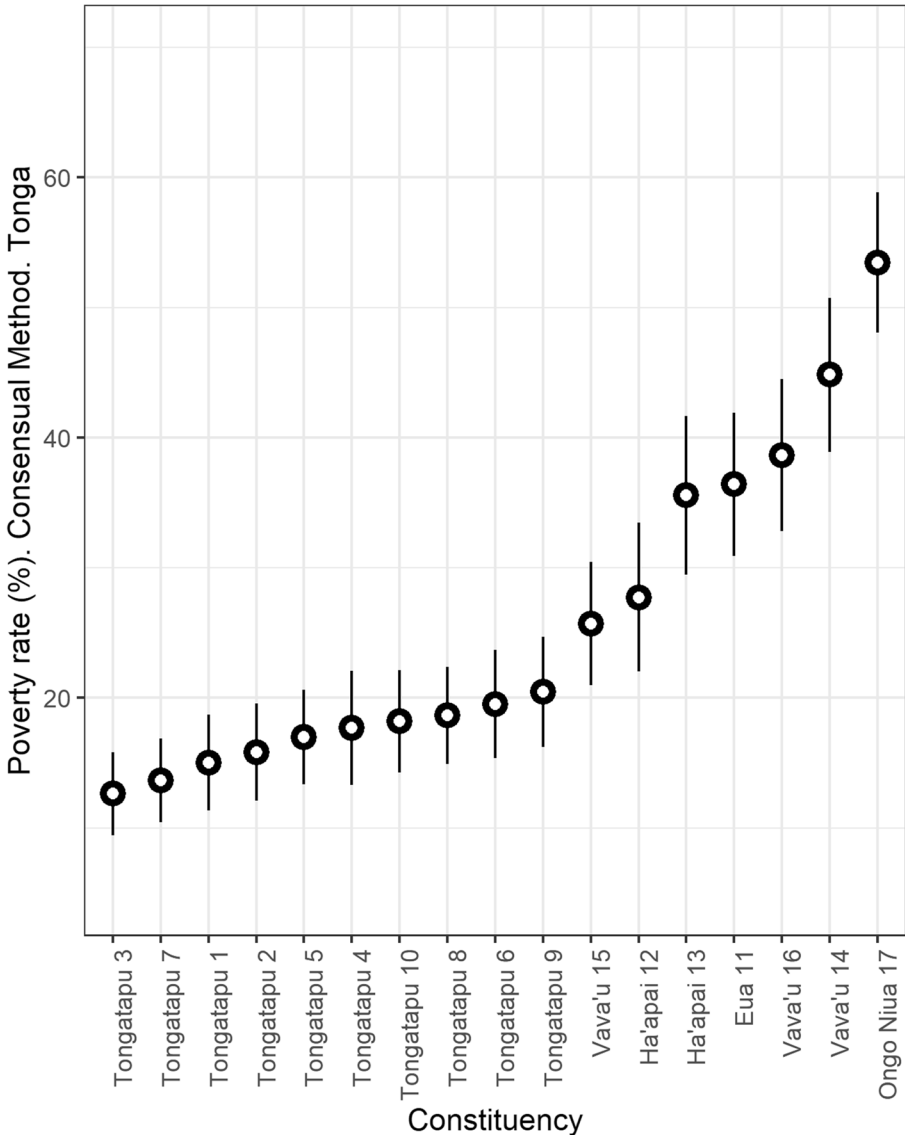

Fig. 5 Poverty estimates 2015/2016 based on the HHB estimator. Constituencies, Tonga

observed prevalence rates in the survey. This is the desired result and is considered as one of the minimum criteria to validate an SAE model (Rao and Molina 2015). In other words, it is an indication that the parameters from the hierarchical model lead to a consistent estimation of poverty with different data.

Once the HHB estimator has been validated, it is possible to produce the island-level estimates of poverty for each of the five main islands. Table 4 compares the direct estimates for each island from the survey (which are unreliable) with the indirect estimates from the HHB. It was noted in the descriptive analyses that there were some differences in living standards between Tongatapu and the rest. The values in the first column are the estimates obtained directly from the HIES for each island, but these are unreliable and biased, given the random variation in sampling (i.e. islands are not representative). These estimates, for example, suggest that Ha'apai is less poor than Eua, which is hardly believable in view of the fact that Eua does not exhibit a very low standard of living. The HHB estimator, on the other hand, provides sensible estimates. The prevalence rate in Ha'apai and Eua suffer a dramatic adjustment, with an increase
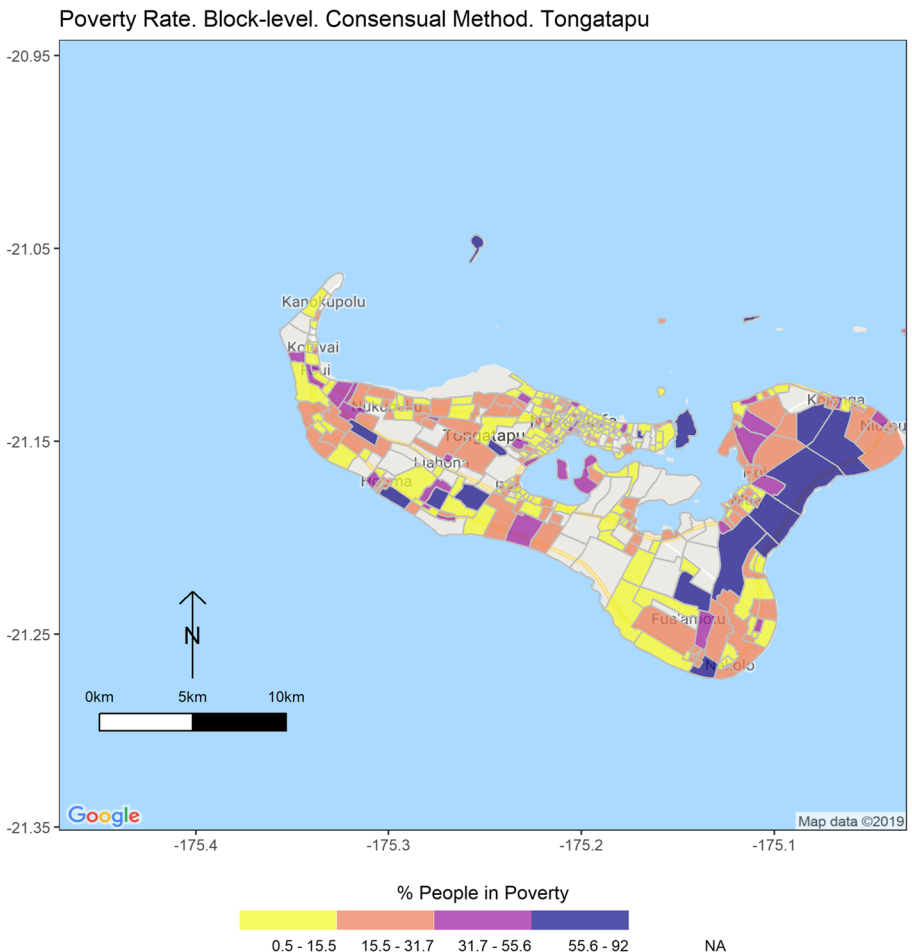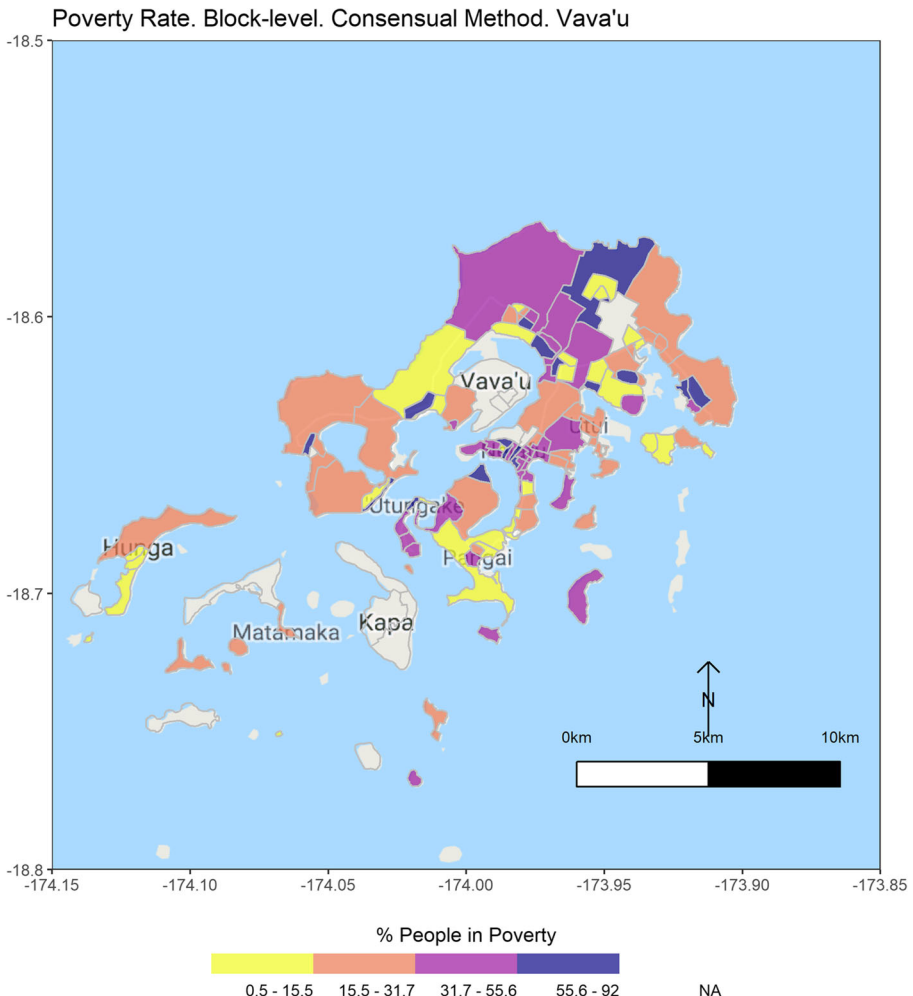

Map 1 Poverty rate. Consensual method. Block-level estimates. Tongatapu 2016

in 12% and a decrease in 17%, respectively, therefore suggesting that the HHB did a good job in adjusting the poverty rate via the predictors included in the model. Poverty is also reduced dramatically in Niuas by 25%. Table 4, below, suggests that poverty is lower in Tongatapu relative to the other four islands, it has the same prevalence in Vava'u, Ha'apai and Eua and it is higher in Niuas.

Constituency-Level Estimates

The Kingdom of Tonga has 17 constituencies distributed across the five main islands. Nine out of the 17 are in Tongatapu, three in Vava'u, two in Ha'apai, and Eua and Ongo Niua have only one. Figure 5 shows the distribution of poverty by constituency along with its respective credible intervals. All of the constituencies in Tongatapu show low and similar poverty rates. The small-area estimates suggest that there are important
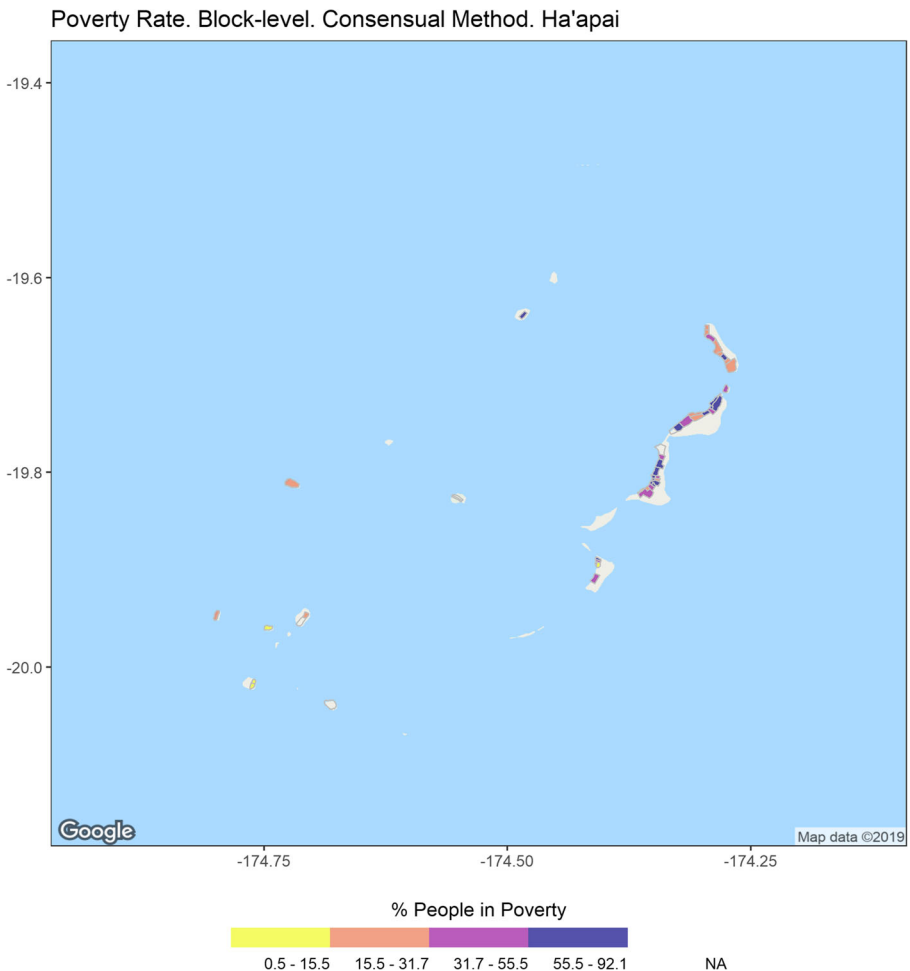

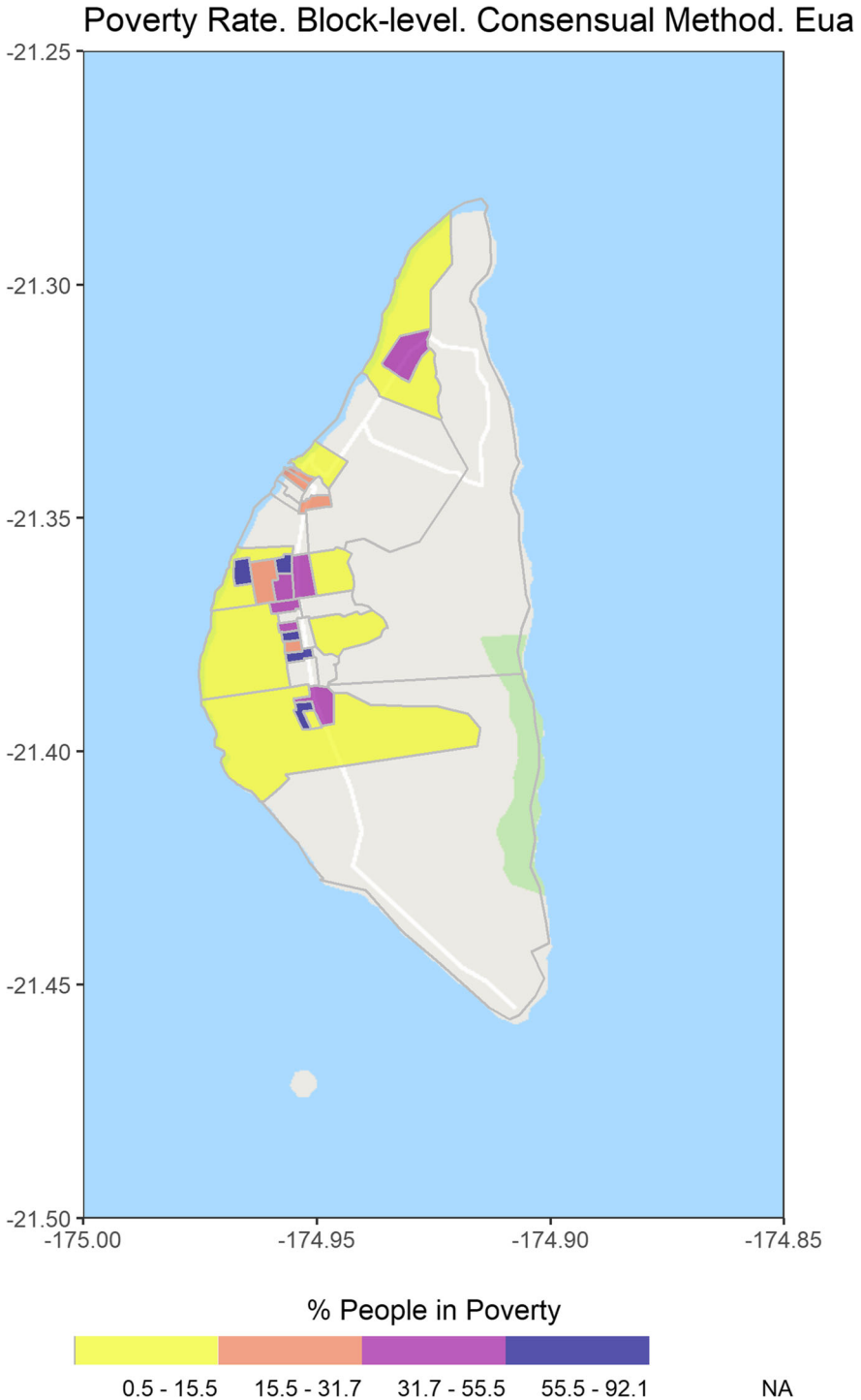Map 2 Poverty rate. Consensual method. Block-level estimates. Vava'u, 2016

within-island differences when poverty is measured at the constituency level; in Vava'u, for example, poverty fluctuates considerably between the three constituencies.
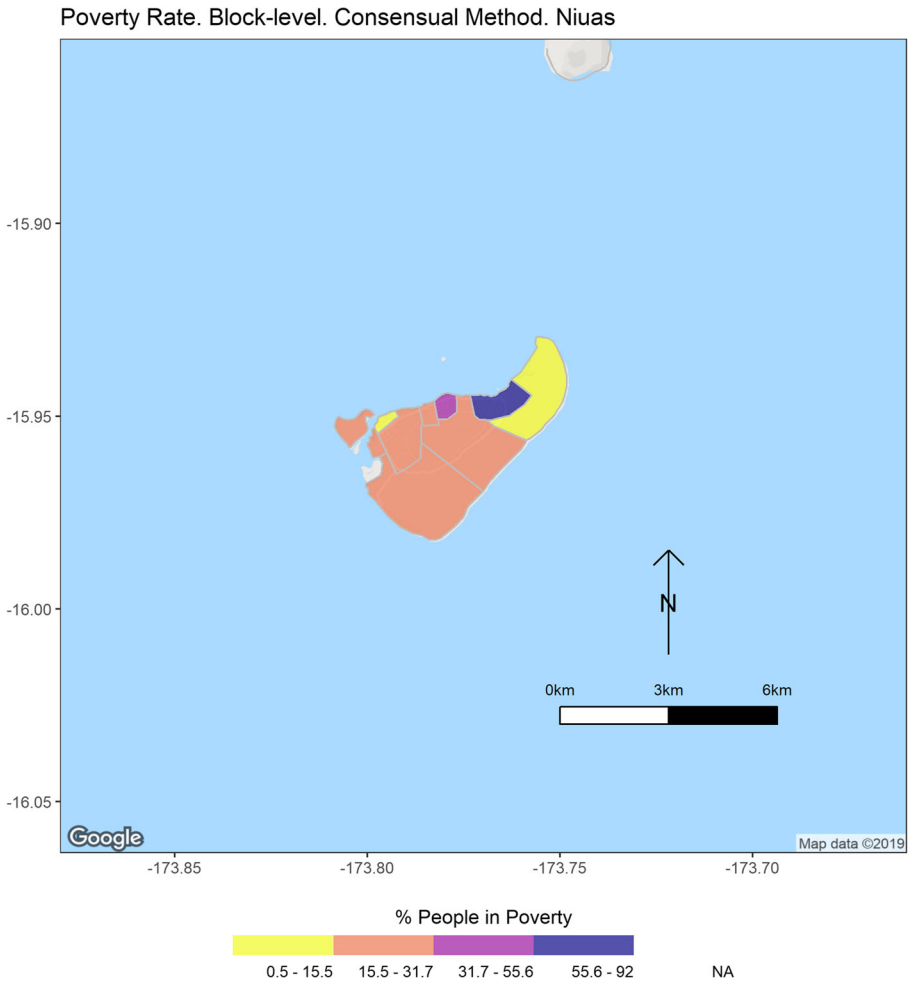
Block-Level Estimates

This section presents the within-island distribution of poverty, according to the HHB estimator, using 'blocks' (small areas within islands) as units of analysis. A map is produced for each of the five main islands and ordered as follows: Tongatapu, Vava'u, Ha'apai, Eua and Ongo Niua (Niuas). The appendix displays a reference map with the location of each island. Map 1 suggests that poverty in Tongatapu's city (North) centre is lower relative to the north-west, south-west and south. For the Vava'u group archipelago, Map 2 does not suggest a very clear clustering pattern, while Neiafu (Vava'u downtown) shows both moderate and high poverty rates. Ha'apai (Map 3) shows a rather

Map 3 Poverty rate. Consensual method. Block-level estimates. Ha'apai, 2016

Map 4 Poverty rate. Consensual method. Block-level estimates. Eua, 2016

Map 5 Poverty rate. Consensual method. Block-level estimates. Niuas, 2016

homogeneous distribution of poverty and not many variations in its mean prevalence rate, and Eua and Niuas do not show a clear pattern. It seems that the lack of any clear pattern can be attributed to population dispersion within the islands.

Conclusion

The main objective of this article is to produce small-area poverty estimates, based on the consensual approach (CA) at island, constituency and block levels in the Kingdom of Tonga, through data from the HIES (2015/2016) and the population Census (2016). These estimates were based on an indirect SAE estimator, namely the hybrid hierarchical Bayes (HB) estimator, which, unlike the standard HB, uses Hamiltonian Monte Carlo (HMC) instead of Marcov Chain Monte Carlo (MCMC) samplers (Betancourt

2017; Carpenter et al. 2016). This estimator has been recently applied by Nájera (2019) to produce municipal-level estimates of stunting in Mexico.

The findings suggest that, at the island level, Tongatapu has the lowest poverty rates, followed by Vava'u, Eua and Ha'apai, which have similar but higher prevalence rates of poverty than the capital. Niuas is the island with the highest poverty rate in the Kingdom of Tonga, and this pattern is reflected across constituencies, in that all constituencies in Tongatapu have lower poverty rates than those located in the other four main islands. However, poverty in Tonga – leaving aside the constituency in Niuas and constituencies 14 and 16 in Vava'u – fluctuates moderately across constituencies (20% and 35% range). The HHB also was useful in producing block-level estimates. In Tongatapu, poverty is highly spatially autocorrelated, with a high rate on the outskirts and a lower rate near the downtown area. In the rest of the islands, there were no strong signs of spatial autocorrelation, most likely due to significant population dispersion and low poverty variability in some of the poorer islands like Niuas.

The paper contributes to the current literature in SAE by producing the first small-area estimates for Tonga and for a country in the South Pacific, using a Bayesian hierarchical model. The article is also one of the very few undertakings to conduct an SAE exercise based on a poverty measure derived from the consensual approach (Dorling et al., 2007). Whereas Dorling and colleagues produced small-area estimates for the UK, their estimates were based on a GREG estimator that does not make the most of the contextual information and is likely to be less accurate. Furthermore, although there have been some SAE implications using the HB estimator (Rao and Wu 2010; You and Rao 2002; Chen et al. 2014), there are very few applications for poverty research – and even fewer using the HMC approach (Hernandez-Stumpfhauser et al. 2016).

The HMC approach allowed for estimating several kinds of hybrid hierarchical Bayesian models, not only rather quickly (10–15 min) but with a very high success rate in terms of prediction accuracy and the effective number of sample sizes. Although the Tonga data were not particularly big, these features highlight some of the potential HHB advantages available for small-area estimation, given that the SAE often leads with high-dimensional problems, due to the large number of parameters, sample sizes and numbers of models that need to be estimated. From the perspective of the SAE literature, computer simulations are required to assess further in which circumstances the HHB is necessary, as in some settings the standard HB is easier and quicker.

The methodology employed in this report can be reproduced in other settings with different poverty measures and with similar kinds of data, i.e. a survey plus the Census data. Furthermore, the methodology permits computing poverty for specific population groups, provided there is a way to identify them using variables found in the population Census.

Compliance with Ethical Standards

Conflict of Interest   Dr. Viliami Konifelenisi Fifita and Mr. Winston Faingaanuku are members of the Tonga Statistics Department and were involved in the data collection process. However, the authors declare that they have no conflict of interest.

Annex

Table 5 Poverty rate. Constituencies. Tonga 2016

Constituency	Number	Constituency poverty rate	Standard error (Bayesian Sd)
Tongatapu	1	16.4	2.7
Tongatapu	2	18.3	2.7
Tongatapu	3	16.4	2.2
Tongatapu	4	19.8	2.5
Tongatapu	5	20.9	1.7
Tongatapu	6	21.8	2.4
Tongatapu	7	16.4	2.2
Tongatapu	8	21.0	1.9
Tongatapu	9	23.3	3.4
Tongatapu	10	20.0	2.0
Eua	11	37.0	7.1
Ha'apai	12	30.5	6.3
Ha'apai	13	37.3	5.2
Vava'u	14	48.6	6.4
Vava'u	15	31.4	5.6
Vava'u	16	43.6	6.2
Ongo	17	56.7	6.1
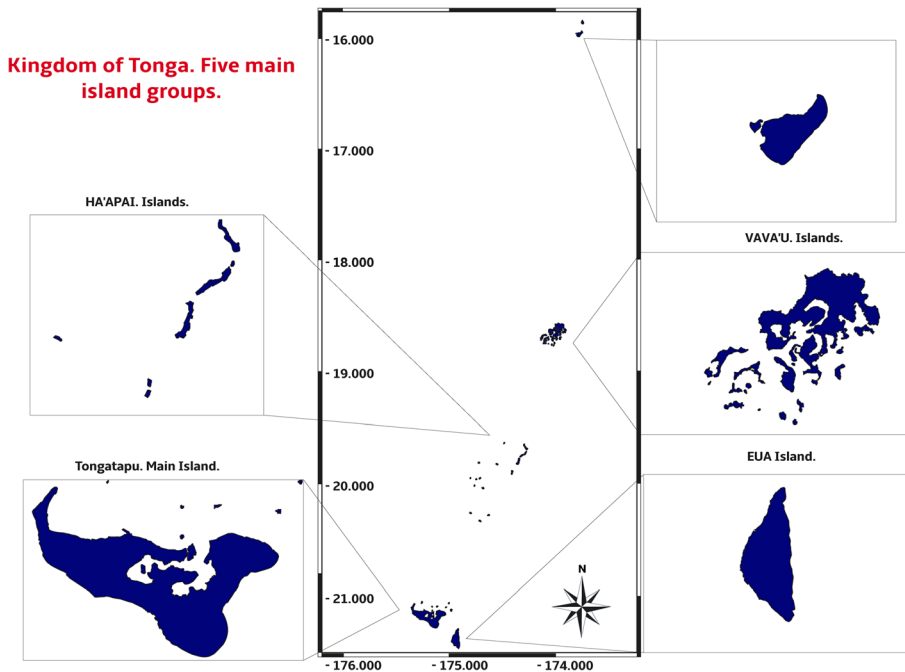
Map 6 Tonga reference map. Main 5 islands. Source: Tonga Statistical Department (2017)

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. ArXiv E-Prints.

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

Carle, A. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology, 9*(1), 49. https://doi.org/10.1186/1471-2288-9-49.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software, 20*, 1–37.

Chambers, R., & Tzvidis, N. (2006). M-quantile models for small area estimation. *Biometrika, 93*(2), 255–268. https://doi.org/10.1093/biomet/93.2.255.

Chambers, R., Chandra, H., Salvati, N., & Tzvidis, N. (2014). Outlier robust small area estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(1), 47–69. https://doi.org/10.1111/rssb.12019.

Chen, C., Wakefield, J., & Lumely, T. (2014). The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spatial and Spatio-Temporal Epidemiology, 11*(11(Supplement C), 33–43. https://doi.org/10.1016/j.sste.2014.07.002.

Dorling, D., Rigby, J., Wheeler, B., Ballas, D., Thomas, B., Fahmy, E., Gordon, D & Lupton, R. (2007). Poverty, wealth and place in Britain, 1968 to 2005. The Policy Press for the Joseph Rowntree Foundation.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B, 195*(2), 216–222.

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

EURAREA. (2012). EURAREA small area estimation. European Comission.

Gelfan, A., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics (Handbooks of Modern Statistical Methods)* (Gelfan A., P. Diggle, P. Guttorp, & M. Fuentes, Eds.). Chapman & Hall/CRC.

Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology, 23*(2), 127–135.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL: CRC press.

Goldstein, H. (1999). *Multilevel Statistical Models* (H. Goldstein, Ed.). London Institute of Education.

Gordon, D. (2006). The concept and measurement of poverty. In C. Pantazis, D. Gordon, & R. Levitas (Eds.), *Poverty and social exclusion in Britain: The millennium survey* (pp. 29–69). Bristol Policy Press.

Gordon, D. (2010), Metodología para la Medición Multidimensional de la Pobreza para México a Partir del Concepto de Privación Relativa, en Minor Mora (coord.), Medición Multidimensional de la Pobreza en México, El Colegio de México-coneval, México, (pp. 401-498).

Gordon, D. (2018). Measuring poverty in the UK. In E. Dermott & G. Main (Eds.), *Poverty and Social Exclusion in the UK*.

Guadarrama, M., Isabel, M., & Rao, J. N. K. (2014). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition New Series, 17*(2450–0291), 41–66. https://doi.org/10.21307/stattrans-2016-005.

Guio, A., Gordon, D., & Marlier, E. (2012). *MEASURING MATERIAL DEPRIVATION IN THE EU: Indicators for THE whole population and child-specific indicators*. EUROSTAT.

Guio, A.-C., Marlier, E., Gordon, D., Fahmy, E., Nandy, S., & Pomati, M. (2016). Improving the measurement of material deprivation at the European Union level. *Journal of European Social Policy, 26*(3), 219–333. https://doi.org/10.1177/0958928716642947.

Guio, A.-C., Gordon, D., Najera, H., & Pomati, M. (2017). *Revising the EU material deprivation variables*. EUROSTAT.

Halleröd, B. (1995). The truly poor: Direct and indirect consensual measurement of poverty in Sweden. *Journal of European Social Policy, 5*(2), 111–129. https://doi.org/10.1177/095892879500500203.

Haslett, S. J., & Jones, G. (2010). Small-area estimation of poverty: The aid industry standard and its alternatives. *Australian & New Zealand Journal of Statistics, 52*(4), 341–362. https://doi.org/10.1111/j.1467-842X.2010.00588.x.

Hernandez-Stumpfhauser, D., Breidt, F. J., & Opsomer, J. D. (2016). Hierarchical Bayesian small area estimation for circular data. *Canadian Journal of Statistics, 44*(4), 416–430. https://doi.org/10.1002/cjs.11303.

Hoffman, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.

Kahle, D., & Wickham, H. (2013). Ggmap: Spatial visualization with ggplot2. *The R Journal, 5*(1), 144–161.

Konifelenisi, V., Lanau, A., Nájera, H., & Gordon, D. (Forthcoming). *Assessing progress towards the eradication of poverty in the Kingdom of Tonga*. Tonga Statistics Department.

Lahiri, P., & Rao, J. (1995). Robust estimation of mean squared error of small area estimators. *Journal of the American Statistical Association, 90*(430), 758–766.

Lau, M., Pantazis, C., Gordon, D., Lai, L., & Sutton, E. (2015). Poverty in Hong Kong. *The China Review, 15*(2), 23–58.

Lumley, T. (2011). *Complex surveys: A guide to analysis using R* (Vol. 565). John Wiley & Sons.

Lumley, T. (2016). Survey: Analysis of complex survey samples.

Mack, J., & Lansley, S. (1985). *Poor Britain (J. Mack & S. Lansley, Eds.)*. London: George Allen & Unwin.

Mack, J., Lansey, S., Nandy, S., & Pantazis, C. (2013). *Attitudes to necessities in the PSE 2012 survey: Are minimum standards becoming less generous?* Bristol University.

Militino, A., Ugarte, M., & Goicoa, T. (2007). A BLUP synthetic versus an EBLUP estimator: An empirical study of a small area estimation problem. *Journal of Applied Statistics, 34*(2), 153–165.

Molina, I., Nandram, B., & Rao, J. N. K. (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics, 8*(2), 852–885. https://doi.org/10.1214/13-AOAS702.

Nájera, H. (2018). Reliability, population classification and weighting in multidimensional poverty measurement: A Monte Carlo study. *Social Indicators Research, 142*, 887–910. https://doi.org/10.1007/s11205-018-1950-z.

Nájera, H. (2019). Small-area estimates of stunting. Mexico 2010: Based on a hierarchical Bayesian estimator. *Spatial and Spatio-Temporal Epidemiology, 29*, 1–11. https://doi.org/10.1016/j.sste.2019.01.001.

Nandy, S., & Pomati, M. (2015). Applying the consensual method of estimating poverty in a low income African setting. *Social Indicators Research, 124*(3), 693–726. https://doi.org/10.1007/s11205-014-0819-z.

Neal, R. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo, 2*(11).

Pantazis, C., Gordon, D., & Levitas, R. (2006). *Poverty and Social Exclusion in Britain: The Millennium Survey* (C. Pantazis, D. Gordon, & R. Levitas, Eds.). Retrieved from https://books.google.com/books?id=o-H0J4BMWS8C

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis, 12*(4), 375–385.

Perry, B. (2002). The mismatch between income measures and direct outcome measures of poverty. *Social Policy Journal of New Zealand*, 101–127.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science, 28*(1), 40–68. https://doi.org/10.1214/12-STS395.

Pratesi, M. (2016). *Analysis of poverty data by small area estimation*. John Wiley & Sons.

Rao, J. N., & Molina, I. (2015). *Small-Area Estimation*. Wiley Online Library.

Rao, J. N. K., & Wu, C. (2010). Bayesian pseudo-empirical-likelihood intervals for complex surveys. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(4), 533–544. https://doi.org/10.1111/j.1467-9868.2010.00747.x.

Tonga Statistics Department (2016). *Tonga Household Income and Expenditure Survey*. Retrieved from https://sdd.spc.int/en/news/latest-news/148-tonga-household-income-and-expenditure-survey-2015-2016-report

Tonga Statistics Department (2017). *Tonga 2016. CENSUS OF POPULATION AND HOUSING. Volume 1: Basic tables and administrative report*. Retrieved from Tonga statistics department website: http://tonga.popgis.spc.int/#c=home

Townsend, P. (1979). *Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living* (P. Townsend, Ed.). University of California.

Townsend, P. (1987). Deprivation. *Journal of Social Policy, 16*(02), 125–146. https://doi.org/10.1017/S0047279400020341.

Wickham, H. (2009). ggplot2:s Elegant graphics for data analysis. Retrieved from http://ggplot2.org

You, Y., & Rao, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *Canadian Journal of Statistics, 30*(3), 431–439. https://doi.org/10.2307/3316146.

Affiliations

Héctor E. Nájera Catalán [1,2,3] · Viliami Konifelenisi Fifita [4] · Winston Faingaanuku [4]

Viliami Konifelenisi Fifita
vfifita@stats.gov.to

Winston Faingaanuku
wfaingaanuku@yahoo.com

[1]   Universidad Nacional Autónoma de México (UNAM), Ciudad de México, CDMX, Mexico

[2]   University of Bristol, Bristol, UK

[3]   Programa de Estudios de Desarrollo, Ciudad de México, CDMX, Mexico

[4]   Tonga Statistics Department, Nuku'alofa, Tonga