|  |  |
| --- | --- |
| Comment | Response |
| I understand the last census corresponds to 1986. Was any effort made to incorporate district level covariates derived from the census? | We have reached out to SNBS colleagues to see if this is available. |
| When citing the work of van der Weide, you should point out that they find that geo referenced data are lacking. This is very relevant as a caveat to the work you’re presenting. You can also add Corral, Henderson and Segovia (2024) who also find that remotely-sensed data may not be ideal for poverty mapping due to their low predictive capacity. | The note will be updated with this information. |
| The document is missing information on the sample. I’d like to know how many areas benefit from EB and what’s their sample size. Related, I’d like to see the shrinkage factor plotted. | All this will be included as well. |
| I’ve veered away from using lasso for model selection in FH models since it ignores the actual assumed model and often yields poor results. Given the poor model fit, as noted by the very low R2, I recommend the team try another approach that should yield a better model. | We have started off by creating a correlation matrix and removing one of any pair of variables with correlation coefficient over 0.8. In addition, ranked the variables in descending order of correlation coefficient to the target area poverty rate. We applied stepwise regression as prescribed but only to the top 43 variables (there are 48 target areas). This has left us exactly 10 indicators which were picked using the BIC criteria. Please let us know what you think about this process. We will include in the next iteration of the note if you approve. |
| The regression table should indicate the number of observations. | This note will be updated with this information. |
| What exactly is the “average share of population within 2 km grids”? How are these constructed to fit the district? |  |
| When checking for normality of the residuals you could provide results from a Shapiro-Wilk normality test. | The note will be updated with this information. |
| I very much like figure 2, is the range of samples between 200 and 600 – or are there smaller and larger areas? |  |
| You note that benchmarking had been utilized after review of the initial since 3 districts were identified as having unrealistically low estimates of poverty. Before even considering this the model needs to be improved. I know that van der Weide et al (2024) and Corral et al (2024) indicate the problems of using GIS data, but in this context you should be getting a better model. You have a model with 2 covariates, one capturing tropical fruit production and the other capturing some form of population. Those two variables, alone, should not be enough to capture the spatial variation of welfare in the country. What was the sample size in these districts?  “In Burundi, we show that SAE generates poverty estimates that are sufficiently precise to report at the district level instead of the regional level.” Who is we? | We will apply the variable selection methods you recommend to improve the model and see if the questions asked here need to be answered. |
| I just want to confirm that the number of HH in the country is 13.6 million, while its population is 16.9 million. I was caught by surprise by the small number of household members. | This seems like a mistake. |
| I may be mistaken, but Corral et al. (2022) do not explicitly recommend implementing an FH with geospatial data. The Guidelines focused on methods, not data. First sentence of “The Methodology” (Corral et al. 2022) recommends implementing an area level Fay Herriot level model with geospatial indicators for poverty mapping. Also, you may want to go over typos. | In the absence of census data, my understanding from the guidelines is that the FH model is the next best alternative. In our case, the country team does not have access to administrative data and the last census is from 1986. The only other option we felt we had was to apply the FH model to geospatial data.  We will make this more explicit in the paper. |
| I’m assuming the estimated variances for the direct estimates followed the noted sampling structure. The team should make it clear in the document that these were considered when estimating the variances. The FH is heavily dependent on these being estimated correctly. Moreover, assuming SRS for these will tend to give too much weight to the direct estimates. | Noted, the team will make sure to reflect this in the next iteration of the note. |
| In the methods section, the document notes that are computed via weighted least squares, but afterwards note that the model is fit via restricted maximum likelihood. It would be good to clarify which method is used. | Noted, the team will fix this within the note. |
| EB stands for empirical best not Bayes, this was something JNK Rao has pointed out to us a few times. | Noted, the team will fix this in the note. |
| **Out-of-sample prediction**: As one of the key objectives of small area estimation is to provide reliable out-of-sample predictions, it would be good to provide in the methodological report an assessment of how well the chosen models do with out-of-sample predictions. This would particularly be useful in light of some concerns related to model predictions for some of targets expressed by the NSO. | I understand the need to show how well we predict out-of-sample, however, we only have 48 target areas. If we used the standard n-fold cross validation techniques, we would be splitting an already small sample into a smaller training and test set. One possibility is to use the leave-one-out CV method i.e. run the model on 47 areas and leave a different target area out each time and see well each area is predicted out of sample. What do you think about this method? |
| **Variance modelling**: I didn’t see in the report how variances associated with direct area estimates of poverty are treated in the model. These are commonly smoothed based on a separate model, and the smoothed variances then employed in the Fay-Herriot model. A plot of these variances against the shrinkage parameter from the FH model is typically instructive. | We will apply variance smoothing at the regional level and include the plots you have requested in the note. |
| **Model specification**: I worried as to the ability of an FH model with only two variables to explain meaningfully variation in poverty rates in Somalia. Moreover, in the model that removes from the sample the SNBS concern districts (Table S2), both of the variables from the baseline model disappear altogether. I think this concern is related to point (i) regarding out-of-sample predictions. Moreover, in the revised model in Table S2 there is only 1 variable in the model that can explain within-region variation in poverty rates, which is what the exercise is trying to model. I would explore avenues to obtain additional data that could be brought into the modelling. Perhaps something like the OpenStreetMap data that underlie the index of critical infrastructure: <https://www.nature.com/articles/s41597-022-01218-4>  It would also be helpful to have some more information about the explanatory variables that the modelling exercise relies on (the variables listed in Table 2). Some information about their distributions, pairwise correlations, spatial patterns etc. | As previously specified, we are in the process of applying the stepwise variable selection method. We will explore adding additional variables from OSM to the model if the selection process does not improve the model prediction. |
| **Model assumptions**:  **Normality**: The discussion related to Figure 1 suggests that normality assumptions are met by the model, but the residual plots in Figure 1 do not look normally distributed to me, particularly for the random effects. It would be helpful to present q-q plots, and formal tests.  **Outlier analysis**: It would be helpful to present a scatterplot of predictions with and without the outlier SNBS areas (it’s a bit hard to see the differences on the map version of this in Figure S1). It seems that estimates change quite a bit, which is a bit concerning that the estimates are so sensitive to model specification. | Once we have redone the variable selection model and re-estimated the model, we will include the skewness and kurtosis which were around 0 and 3 respectively but also the Shapiro-Wilks normality tests as well. In addition, we will perform the outlier analysis highlighted as well. |
| **Value added of district-level estimates**: It would be helpful to present a table of district-level predictions with their associated standard errors / confidence intervals, as well as how these relate to the direct regional poverty estimates that we have from the survey. Since the number of districts is relatively small, it would be helpful to highlight in this table the districts for which the predicted poverty rates are statistically different from the direct regional poverty rate for the region they belong to. This would showcase the additional information we’ve gained on account of poverty mapping that wasn’t available to us by just looking at direct estimates of regional poverty rates. | We will include these in the next iteration of the note. |

## Paul Andres Corral Rodas

Dear Alejandro, Ify, and team,

Thank you for inviting me to review the small area estimation report for Somalia. The report is well written and offers a solid overview of the team's work. However, while we often describe these SAE exercises as strictly predictive, they also call for additional intuition and thoughtful consideration from our side. I would not expect that only 2 covariates suffice to explain the welfare variation across districts in Somalia, and the model’s R2 suggests the same. I’m sharing below some comments and suggestions which I hope can help the team improve upon their estimates.

* I understand the last census corresponds to 1986. Was any effort made to incorporate district level covariates derived from the census?
* When citing the work of van der Weide, you should point out that they find that geo referenced data are lacking. This is very relevant as a caveat to the work you’re presenting.
  + You can also add Corral, Henderson and Segovia (2024) who also find that remotely-sensed data may not be ideal for poverty mapping due to their low predictive capacity.
    - Corral, P., Henderson, H., & Segovia, S. (2024). Poverty mapping in the age of machine learning. *Journal of Development Economics*, 103377.
* The document is missing information on the sample.
  + I’d like to know how many areas benefit from EB and what’s their sample size
  + Related, I’d like to see the shrinkage factor plotted
* I’ve veered away from using lasso for model selection in FH models since it ignores the actual assumed model and often yields poor results. Given the poor model fit, as noted by the very low R2, I recommend the team try another approach that should yield a better model
  + I’m sharing an example of the code I use, which I’m sure you can convert to R and even improve upon. It is similar to a stepwise but uses the assumed model to do the selection. This is what was illustrated in the Guidelines, but I’ve made it a bit faster.
* The regression table should indicate the number of observations
  + What exactly is the “average share of population within 2 km grids”?
    - How are these constructed to fit the district?
* When checking for normality of the residuals you could provide results from a Shapiro-Wilk normality test
* I very much like figure 2, is the range of samples between 200 and 600 – or are there smaller and larger areas?
* You note that benchmarking had been utilized after review of the initial since 3 districts were identified as having unrealistically low estimates of poverty
  + Before even considering this the model needs to be improved. I know that van der Weide et al (2024) and Corral et al (2024) indicate the problems of using GIS data, but in this context you should be getting a better model.
    - You have a model with 2 covariates, one capturing tropical fruit production and the other capturing some form of population. Those two variables, alone, should not be enough to capture the spatial variation of welfare in the country.
  + What was the sample size in these districts?
* I’m sharing the SAE checklist we developed for the global team as guidance as to what teams should be including in their reports. It is mostly geared to unit level models, but much is also applicable to area level models, particularly the outputs and result tables.

Other comments:

* The document lacks some references:
  + “In Burundi, we show that SAE generates poverty estimates that are sufficiently precise to report at the district level instead of the regional level.”
    - Who is we?
* I just want to confirm that the number of HH in the country is 13.6 million, while its population is 16.9 million. I was caught by surprise by the small number of household members.
* I may be mistaken, but Corral et al. (2022) do not explicitly recommend implementing an FH with geospatial data. The Guidelines focused on methods, not data.
  + First sentence of “The Methodology”
    - (Corral et al. 2022) recommends implementing an area level Fay Herriot level model with geospatial indicators for poverty mapping.
      * Also, you may want to go over typos.
* I’m assuming the estimated variances for the direct estimates followed the noted sampling structure. The team should make it clear in the document that these were considered when estimating the variances.
  + The FH is heavily dependent on these being estimated correctly. Moreover, assuming SRS for these will tend to give too much weight to the direct estimates.
* In the methods section, the document notes that are computed via weighted least squares, but afterwards note that the model is fit via restricted maximum likelihood. It would be good to clarify which method is used.
* EB stands for empirical best not Bayes, this was something JNK Rao has pointed out to us a few times.

## Alexandru Cojocaru

Dear colleagues,

Thank you for the chance to review the Somalia poverty mapping methodological report. Please find below a few comments for the team’s consideration:

1. **Out-of-sample prediction**: As one of the key objectives of small area estimation is to provide reliable out-of-sample predictions, it would be good to provide in the methodological report an assessment of how well the chosen models do with out-of-sample predictions. This would particularly be useful in light of some concerns related to model predictions for some of targets expressed by the NSO.
2. **Variance modelling**: I didn’t see in the report how variances associated with direct area estimates of poverty are treated in the model. These are commonly smoothed based on a separate model, and the smoothed variances then employed in the Fay-Herriot model. A plot of these variances against the shrinkage parameter from the FH model is typically instructive.
3. **Model specification**: I worried as to the ability of an FH model with only two variables to explain meaningfully variation in poverty rates in Somalia. Moreover, in the model that removes from the sample the SNBS concern districts (Table S2), both of the variables from the baseline model disappear altogether. I think this concern is related to point (i) regarding out-of-sample predictions. Moreover, in the revised model in Table S2 there is only 1 variable in the model that can explain within-region variation in poverty rates, which is what the exercise is trying to model. I would explore avenues to obtain additional data that could be brought into the modelling. Perhaps something like the OpenStreetMap data that underlie the index of critical infrastructure: <https://www.nature.com/articles/s41597-022-01218-4>
   1. It would also be helpful to have some more information about the explanatory variables that the modelling exercise relies on (the variables listed in Table 2). Some information about their distributions, pairwise correlations, spatial patterns etc.
4. **Model assumptions**:
   1. **Normality**: The discussion related to Figure 1 suggests that normality assumptions are met by the model, but the residual plots in Figure 1 do not look normally distributed to me, particularly for the random effects. It would be helpful to present q-q plots, and formal tests.
   2. **Outlier analysis**: It would be helpful to present a scatterplot of predictions with and without the outlier SNBS areas (it’s a bit hard to see the differences on the map version of this in Figure S1). It seems that estimates change quite a bit, which is a bit concerning that the estimates are so sensitive to model specification.
5. **Value added of district-level estimates**: It would be helpful to present a table of district-level predictions with their associated standard errors / confidence intervals, as well as how these relate to the direct regional poverty estimates that we have from the survey. Since the number of districts is relatively small, it would be helpful to highlight in this table the districts for which the predicted poverty rates are statistically different from the direct regional poverty rate for the region they belong to. This would showcase the additional information we’ve gained on account of poverty mapping that wasn’t available to us by just looking at direct estimates of regional poverty rates.

Warm regards,

Sandu