One more – can you check this statement against the new IPC data?  (on page 56 of the document, in the SI)

Nearly 77% of Malawi’s IPC assessments between 2009 and 2016 were classified as phase 1, or “minimal.” Slightly over one-fifth (21%) were assessed at phase 2, “stressed,” and only 3% were phase 3, “crisis.” The periods of greatest phase 3 food insecurity occurred in 2012 (11% of assessments for that year), and 2015 (6% of assessments for that year).

Should be :

Nearly 78% of Malawi’s IPC assessments between 2009 and 2016 were classified as phase 1, or “minimal.” About one-fifth (19%) were assessed at phase 2, “stressed,” and only 2.5% were phase 3, “crisis.” The periods of greatest phase 3 food insecurity occurred in 2015 (7% of assessments for that year) and 13 IPC zones were evaluated as in crisis at the end of year 2015 .

Boring details below:

1. Table of IPC value total

1 2 3 88 99

1232 301 40 27 20

Portion of Phase 1: 1232/(1232 + 301 +40) = 0.7832

Portion of Phase 2: 301/(1232 + 301 +40) = 0.1913

Portion of Phase 3: 40/(1232 + 301 +40) = 0.025

1. Table of year with IPC value==3 occurrence (ipczone and quarter)

2010 2012 2013 2014 2015 2016

3 6 1 5 17 8

1. Periods/number of ipc zones that have phase 3,

year = 2010

value

key 3

CS20100701 3

CS20121001 0

CS20130101 0

CS20141001 0

CS20150101 0

CS20151001 0

CS20160601 0

, , year = 2012

value

key 3

CS20100701 0

CS20121001 6

CS20130101 0

CS20141001 0

CS20150101 0

CS20151001 0

CS20160601 0

, , year = 2013

value

key 3

CS20100701 0

CS20121001 0

CS20130101 1

CS20141001 0

CS20150101 0

CS20151001 0

CS20160601 0

, , year = 2014

value

key 3

CS20100701 0

CS20121001 0

CS20130101 0

CS20141001 5

CS20150101 0

CS20151001 0

CS20160601 0

, , year = 2015

value

key 3

CS20100701 0

CS20121001 0

CS20130101 0

CS20141001 0

CS20150101 4

CS20151001 13

CS20160601 0

, , year = 2016

value

key 3

CS20100701 0

CS20121001 0

CS20130101 0

CS20141001 0

CS20150101 0

CS20151001 0

CS20160601 8

so, 6 month for 2015 and 3 month for 2010/2012/2013/2014/2016

1. Table of all the IPC values in 2015

1 2 3 88

144 75 17 4

Portion of Phase 3 in 2015 : 17/(144 + 75 + 17 ) = 0.072

1. Table of all the IPC values in 2012

1 2 3 88 99

129 38 6 3 4

Portion of Phase 3 in 2012 : 6/(129 + 38 + 6 ) = 0.034

Hi Yujun – one more thing.  Here’s the paragraph in the text that we need to update.  It’s about figure 3 in the main text – specifically about the amount of over and underprediction rates for each measure (as well as the type I vs II errors).  Do you mind tackling that?

For the HDDS in Fig. 3A vertical dotted lines at 3 and 6 indicate food security category cutoffs of low and medium dietary diversity (proxies for low and medium food security, respectively). Most cluster averages lie between these two categories. While our model tends to under-predict food security in the middle-range of the distribution (we predict 88% have lower than actual HDDS values and 12% have higher values than their true level), we correctly classify the food security category of 87% we predict 79% have lower than actual HDDS values and 21% have higher values than their true level), we correctly classify the food security category of 88% of the sample, outperforming the simple assumption of all households being in the largest category (medium dietary diversity), which would correctly categorize 78%. categorize 76%. Similarly, we observe more type I than type II errors (13% versus than less than 1%); 11% versus than less than 1%); thus, our model misses very few clusters that are truly food insecure.

For FCS, shown in Fig. 3B, the model slightly under-predicts food security for the majority of the sample (we predict 70% have lower FCS values than their true level and 30% 82% have lower FCS values than their true level and 18% have higher). The model performs especially well for households in the lowest food security category in the 2013 data (“borderline”), correctly classifying 77% of the most food insecure clusters. Overall, we correctly predict 62% of the cluster categories, under-estimating the food security category of 31% of the sample (area IV), and over-estimating only 6% (area I). 83% of the most food insecure clusters. Overall, we correctly predict 76% of the cluster categories, under-estimating the food security category of 23% of the sample (area IV), and over-estimating only 1% This result is much better than random, which would correctly categorize only one third of households. On the other hand, due to the fact that the cut-off levels used by humanitarian agencies are designed to capture relatively rare events, our prediction performs slightly worse than if we simply assumed that each cluster was food secure, on average, which would correctly categorize 74% of categorize 91% of (Note: removing the ipc 12 lag adds in more food secure clusters)the clusters.

For rCSI, shown in Fig. 3C, our best model obtains 65% accuracy, under-predicting the food security status of 56% of households, and over-predicting food security for 44%. 65% accuracy, under-predicting the food security status of 61% of households, and over-predicting food security for 39%As with the HDDS, our best model does better than the naïve prediction that all clusters are food secure on average, which would correctly classify 56%, classify 54%, and considerably better than random allocation, which would only accurately predict the food security category of one quarter of the clusters.

Hi all,

I’m trying to work through revising these numbers now, and I’ve confused myself on two, related issues. I’ve got Skype on, if a call is quicker.

1. The accuracy numbers don’t match between the attached results for Figure 3, and Table 3 (old 5 from Yujun). Which categorical accuracy numbers should we be using when discussing Fig 3?

**Yujun: (A)** Our model correctly classifies 88%88% of the household dietary diversity scores. **(B)** Our model correctly predicts 89% 76% of the log food consumption score categories. **(C)** Our model correctly predicts 63% 65% of reduced coping strategies index categories.

**Erin: (A)** Our model correctly classifies 88% of the household dietary diversity scores. **(B)** Our model correctly predicts 76% of the log food consumption score categories. **(C )** Our model correctly predicts 65% of reduced coping strategies index categories.

Erin is right!!!

1. For the text below, I’m finding the details about predicting lower and higher than true values complicated (it is probably me). I’m not sure where those numbers come from. I think we want accuracy + over prediction + underprediction = 100%, which would suggest not using Table 3, but rather using the attached data for Fig 3. Below, I’ve left highlighted the numbers I don’t understand (some might not have changed – e.g., correct classification). The unhighlighted numbers should be correct (Based on the attached excel sheet).

88% have lower than actual HDDS values and 12% have higher values than their true level (This number is basically comparing prediction with the 45 degree line, regardless of the catogories.)

The other numbers have been updated above

While our model tends to under-predict food security in the middle-range of the distribution (we predict 88% have lower than actual HDDS values and 12% have higher values than their true level), we correctly classify the food security category of 88% of the sample, outperforming the simple assumption of all households being in the largest category (medium dietary diversity), which would correctly categorize 78%. Similarly, we observe more type I than type II errors (11% versus than less than 1%); thus, our model misses very few clusters that are truly food insecure.

For FCS, shown in Fig. 3B, the model slightly under-predicts food security for the majority of the sample (we predict 70% have lower FCS values than their true level and 30% have higher). The model performs especially well for households in the lowest food security category in the 2013 data (“borderline”), correctly classifying 77% of the most food insecure clusters. Overall, we correctly predict 88% of the cluster categories, under-estimating the food security category of 11% of the sample (area IV), and over-estimating in less than 1% (area I). This result is much better than random, which would correctly categorize only one third of households. On the other hand, due to the fact that the cut-off levels used by humanitarian agencies are designed to capture relatively rare events, our prediction performs slightly worse than if we simply assumed that each cluster was food secure, on average, which would correctly categorize 74% of the clusters.

For rCSI, shown in Fig. 3C, our best model obtains 65% accuracy, under-predicting the food security status of 29% of households, and over-predicting food security for 6%. As with the HDDS, our best model does better than the naïve prediction that all clusters are food secure on average, which would correctly classify 56%, and considerably better than random allocation, which would only accurately predict the food security category of one quarter of the clusters.

Corrected Table 5

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **R squared** | | | **Sensitivity of food** | | | **Categorical** | | | |
| **Insecure Category** | | | **accuracy** | | | |
|  | logFCS | | HDDS | rCSI | logFCS | HDDS | rCSI | logFCS | HDDS | rCSI |  |
| **Split by year** | | | | | | | | | | | |
| **Main model** | 0.649 | | 0.643 | 0.119 | 0.833 | 0.994 | 0.860 | 0.759 | 0.882 | 0.650 |  |
| **Main model**  **– cluster price**  **variables** | 0.638 | | 0.661 | 0.119 | 0.222 | 0.968 | 0.258 | 0.916 | 0.882 | 0.591 |  |
| **Main Model**  **+ region fixed effect**  **- cluster price variables** | | 0.557 | 0.554 | 0.073 | 0.278 | 0.955 | 0.290 | 0.916 | 0.862 | 0.616 |  |
| **Main Model**  **+ GIEWS price**  **- cluster price variables** | 0.634 | | 0.661 | 0.036 | 0.722 | 0.994 | 0.978 | 0.793 | 0.892 | 0.468 |  |
| **LASSO** | 0.640 | | 0.611 | 0.061 | 0.833 | 1 | 1 | 0.665 | 0.857 | 0.463 |  |