What is Bayesian SSVS?

Stoachstic Search Variable Selection (SSVS)

G. Zens

Department of Economics
Vienna University of Economics and Business

June 2020
Advanced Macroeconometrics (Science Track)

Motivation

- often, it is not clear a priori which variables are relevant enough to be included in a model
- especially relevant in large data settings and settings with no theoretical guidance
- we would like to find a method to conduct **variable selection**, i.e. finding a subet of most relevant predictors
- similarly, including to many predictors makes models prone to overfitting
- here, we need to avoid noisy estimates and to regularize our models
- one method is to push coefficients of irrelevant predictors to 0, commonly refered to as **shrinking** the coefficients to 0

Variable selection

- suppose we have a regression model with 3 potential predictors x_1, x_2, x_3
- for each β_j we have a binary indicator δ_j indicating whether x_j should be included in the model or not

$$y_i = \delta_1 x_{1i}\beta_1 + \delta_2 x_{2i}\beta_2 + \delta_3 x_{3i}\beta_3 + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2) \tag{1}$$

- for δ = $(\delta_1, \delta_2, \delta_3)' = (1, 0, 1)'$, we obtain

$$y_i = x_{1i}\beta_1 + x_{3i}\beta_3 + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2) \tag{2}$$

- George and McCulloch (1993) specify a prior that allows you to estimate δ_j

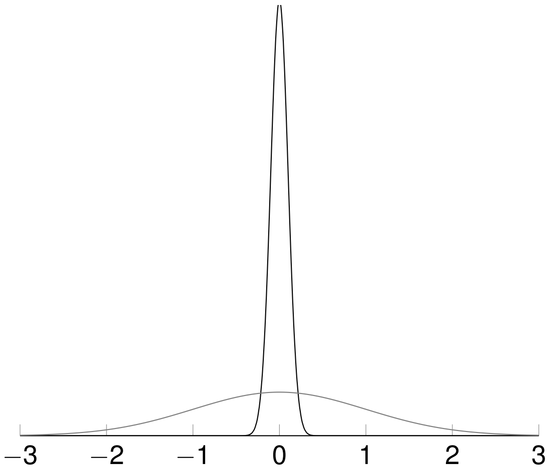
SSVS

- the stochastic search variable selection (SSVS) prior is specified as mixture of Gaussian distributions

$$\beta_j|\delta_j \sim \delta_j N(0, \sigma_{1j}^2) + (1 - \delta_j) N(0, \sigma_{0j}^2) \tag{3}$$

- $\sigma_{1j}^2 >> \sigma_{0j}^2$ leads to a so-called ”spike-and-slab prior”
- if a variable is included ($\delta_j = 1$), the prior variance is large
- if a variable is excluded ($\delta_j = 0$), the variance is small
- hence, excluded variables are pushed towards 0 through the prior (= shrinkage)

Stylized Spike and Slab Prior



SSVS Intuition

In each iteration of the Gibbs sampler, we will...

- check for each β_j whether it is more likely that it comes from the spike or from the slab component
- multiply this likelihoods with the prior probabilities of inclusion/exclusion (often 0.5)
- based on these measures, compute the posterior inclusion probability (PIP)
- with this PIP, we can sample the inclusion indicator δ_j

The posterior mean over all δ_j gives us the overall PIP for β_j. This tells you the probability that x_j is a relevant predictor of y_i (variable selection). At the same time, irrelevant coefficients are pushed into the spike component and will be estimated close to 0 to avoid noisy estimates (shrinkage).

SSVS Details

- note that the conditional prior on β_j can be written as

$$\beta_j|\delta_j \propto (\sigma_{1j}^{-1} exp\{-\beta_j^2/2\sigma_{1j}^2\})^{\delta_j}(\sigma_{0j}^{-1} exp\{-\beta_j^2/2\sigma_{0j}^2\})^{(1-\delta_j)} \tag{4}$$

- indicators $\delta = (\delta_1, \ldots, \delta_k)$ have a Bernoulli prior $\delta = \prod_{j=1}^{k} p_j^{\delta_j}(1-p_j)^{1-\delta_j}$ with success probability p_j
- combining the likelihood $p(\beta_j|\delta_j)$ with the prior $p(\delta_j)$ yields again a Bernoulli distribution

$$\begin{aligned} p(\delta_j = 1|\beta_j) &= Ber(u_{1j}/(u_{1j} + u_{0j})) \\ u_{1j} &= p_j \times \sigma_{1j}^{-1} exp\{-\beta_j^2/2\sigma_{1j}^2\} \\ u_{0j} &= (1-p_j) \times \sigma_{0j}^{-1} exp\{-\beta_j^2/2\sigma_{0j}^2\} \end{aligned} \tag{5}$$

Further notes on shrinkage and variable selection

Note that ...

- ... the SSVS is only one example out of a massive universe of shrinkage priors
- ... other combinations than Gaussian spike & Gaussian slab are available (Malsiner-Walli and Wagner 2018)
- ... there is a huge amount of continuous shrinkage priors around. These are not based on a spike and a slab, but try to approximate a spike and a slab using a single, continuous prior distribution
- ... in macroeconometrics, you are regularly going to encounter:
  - Bayesian LASSO (Park and Casella 2008)
  - Horseshoe prior (Carvalho et al. 2010)
  - Normal-Gamma prior (Griffin and Brown 2010)
  - Dirichlet-Laplace prior (Bhattacharya et al. 2015)
  - Triple-Gamma prior (brand-new, published two weeks ago! Cadonna et al. 2020)

Example: Continuous shrinkage prior (Normal-Gamma)

Fig. 1 Normal gamma prior with a variance of 1 for different values of ν_1, $\nu_1 = 0.5$ (*black dot-dashed line*), $\nu_1 = 1$ (*red dotted line*), $\nu_1 = 2$ (*blue long-dashed line*), and the standard normal density (*green solid line*), at zero (*left-hand side*) and the tails (*right-hand side*). (Color figure online)

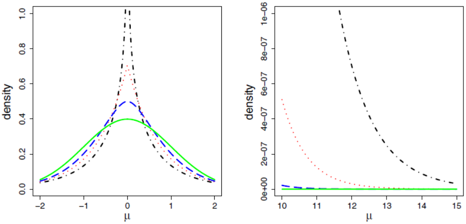


Figure: The Normal-Gamma shrinkage prior. Figure taken from Malsiner-Walli, Frühwirth-Schnatter, et al. (2016).

A continuous shrinkage prior should be able to simultaneously...

- ... put the vast majority of the prior mass extremely close to zero (this overrules weak likelihood information and pushes irrelevant coefficient towards zero)
- ... exhibit ”fat tails” such that the information in the likelihood can still overrule the prior information, if strong enough

References I

- Bhattacharya, Anirban, Debdeep Pati, Natesh S Pillai, and David B Dunson (2015). 'Dirichlet–Laplace priors for optimal shrinkage'. In: *Journal of the American Statistical Association* 110.512, pp. 1479–1490.
- Cadonna, Annalisa, Sylvia Frühwirth-Schnatter, and Peter Knaus (2020). 'Triple the Gamma—A Unifying Shrinkage Prior for Variance and Variable Selection in Sparse State Space and TVP Models'. In: *Econometrics* 8.2, p. 20.
- Carvalho, Carlos M, Nicholas G Polson, and James G Scott (2010). 'The horseshoe estimator for sparse signals'. In: *Biometrika* 97.2, pp. 465–480.
- George, Edward I and Robert E McCulloch (1993). 'Variable selection via Gibbs sampling'. In: *Journal of the American Statistical Association* 88.423, pp. 881–889.
- Griffin, Jim E. and Philip J. Brown (2010). 'Inference with normal-gamma prior distributions in regression problems'. In: *Bayesian Analysis* 5.1, pp. 171–188.

References II

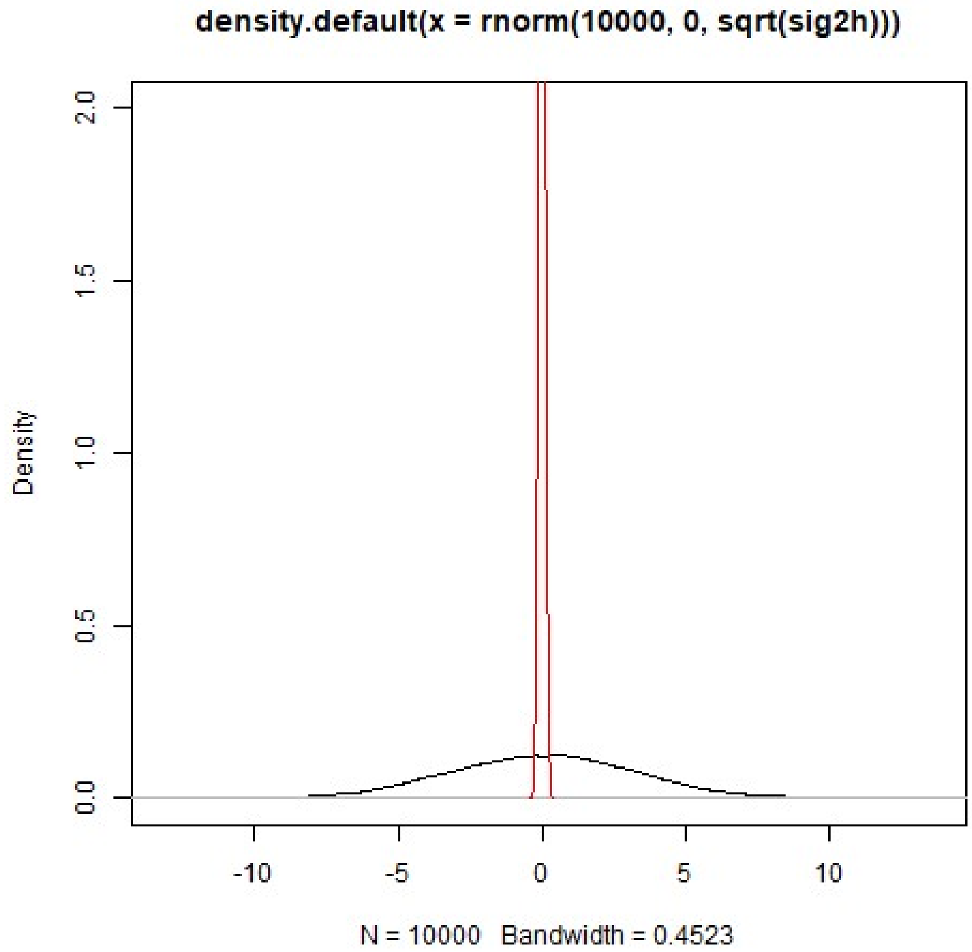
Malsiner-Walli, Gertraud, Sylvia Frühwirth-Schnatter, and Bettina Grün (2016). 'Model-based clustering based on sparse finite Gaussian mixtures'. In: *Statistics and computing* 26.1-2, pp. 303–324.

Malsiner-Walli, Gertraud and Helga Wagner (2018). 'Comparing spike and slab priors for Bayesian variable selection'. In: *arXiv preprint arXiv:1812.07259*.

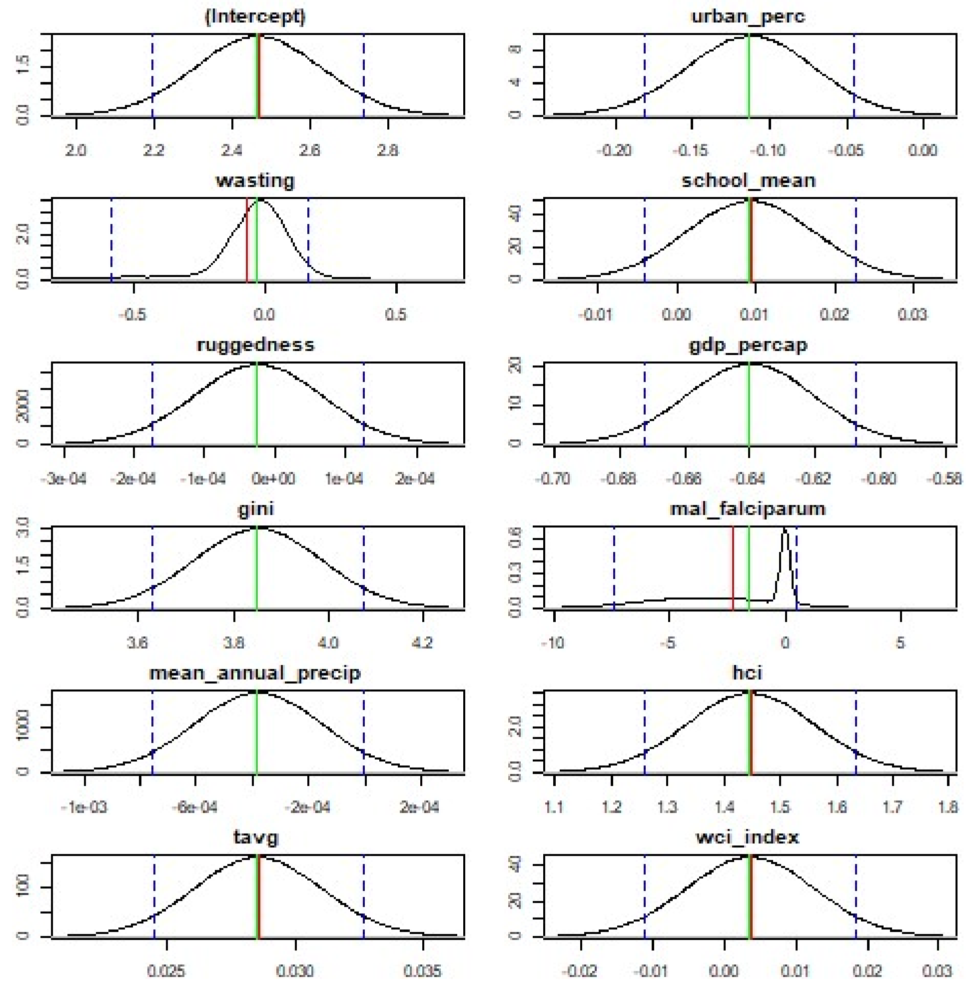
Park, Trevor and George Casella (2008). 'The bayesian lasso'. In: *Journal of the American Statistical Association* 103.482, pp. 681–686.

Bayesian SSVS for Hunger modelling

Spike and slab prior

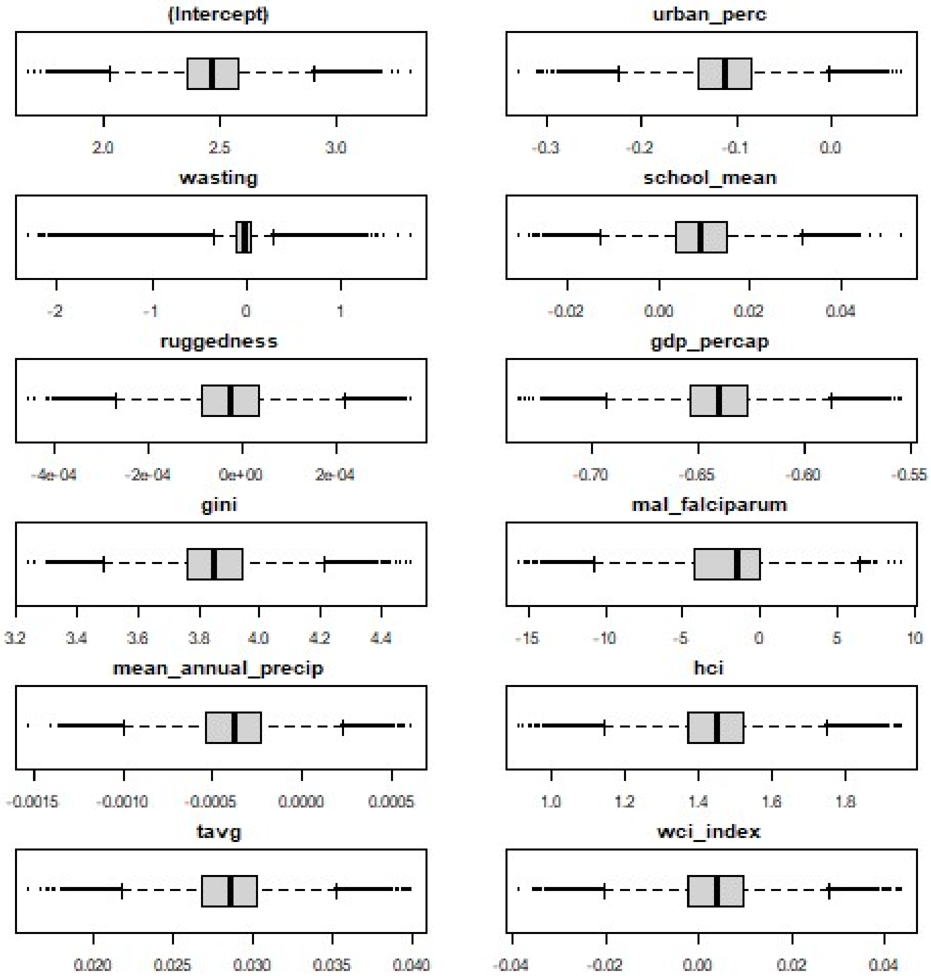


Posterior distributions



mean
median
credible interval: 0.05/0.95

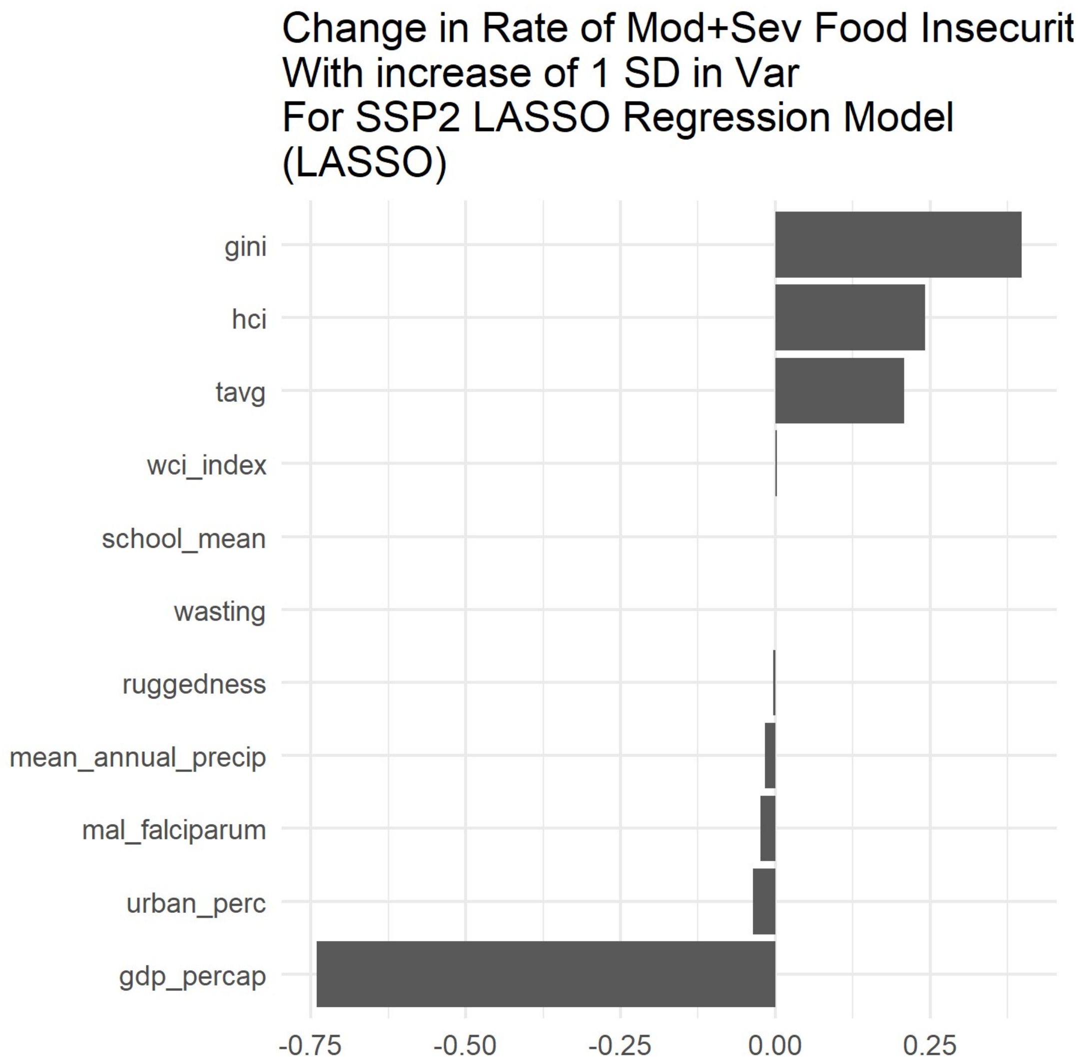
Posterior boxplots

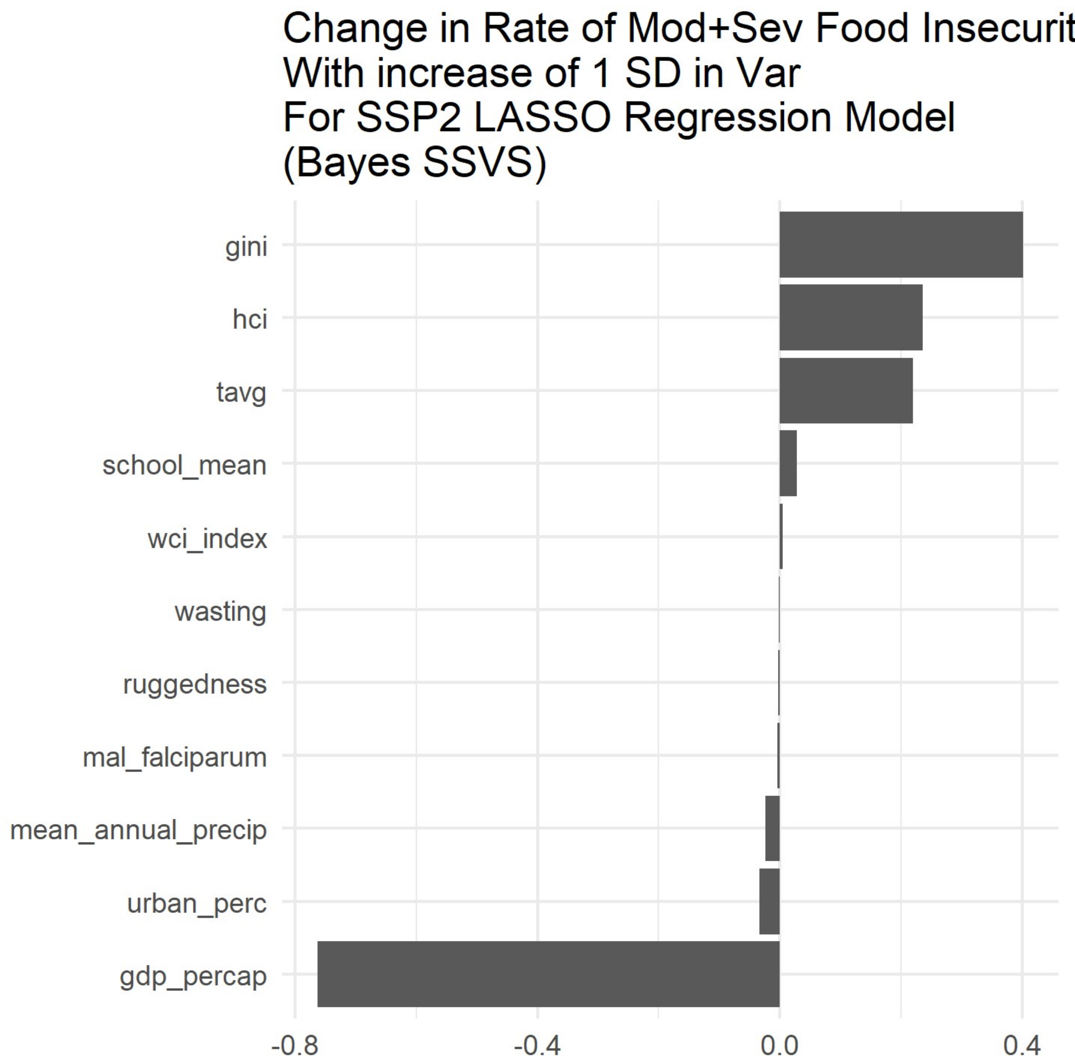


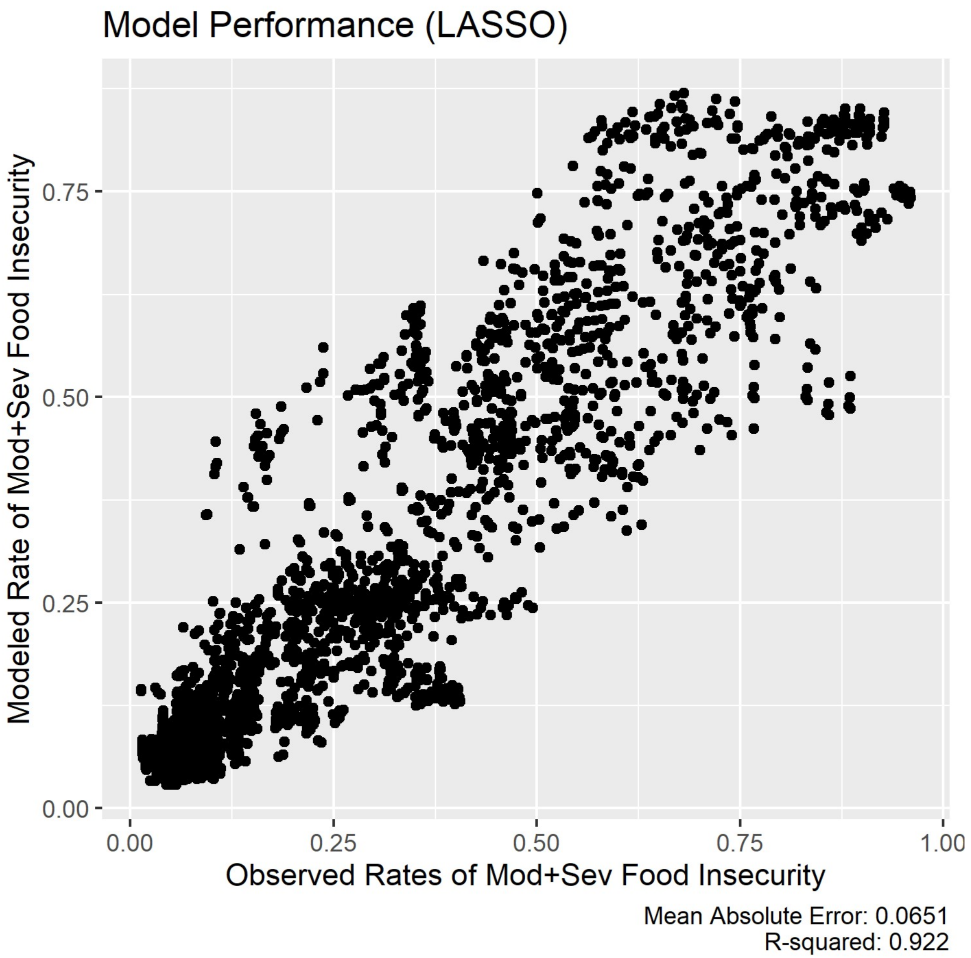
Posterior dataframe

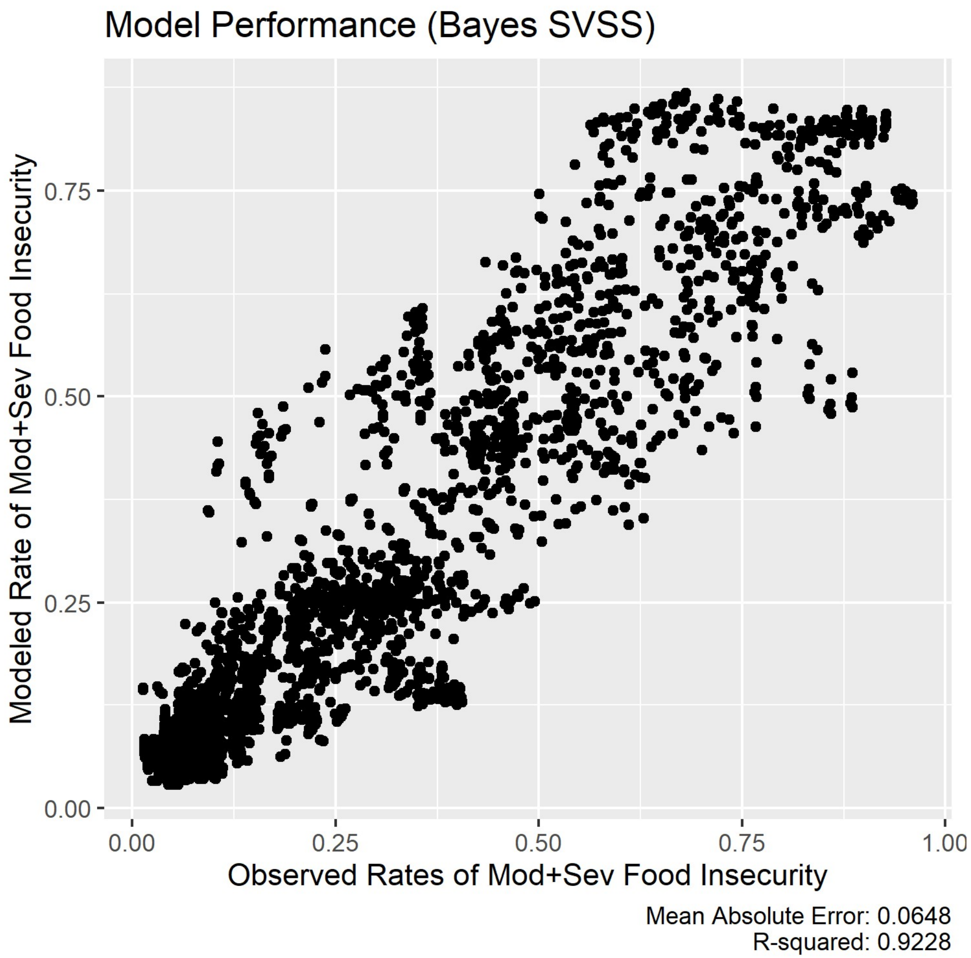
term	mean	pip	median	sd	cred0.05	cred0.95	cred0.01	cred0.99	scaled
(Intercept)	2.465024e+00	1.00000000	2.464592e+00	1.642944e-01	2.1951347336	2.735597e+00	2.0826847113	2.8494416749	*NA*
urban_perc	-1.126987e-01	0.06716222	-1.124820e-01	4.137246e-02	-0.1811215110	-4.496794e-02	-0.2100676490	-0.0168353152	-0.0340147792
wasting	-6.958151e-02	0.16598889	-2.666910e-02	2.458187e-01	-0.5931115389	1.663797e-01	-1.0775209430	0.3917311855	-0.0009936429
school_mean	9.311932e-03	0.03078000	9.315781e-03	8.123169e-03	-0.0040758013	2.266904e-02	-0.0095645672	0.0281954583	0.0278637661
ruggedness	-2.436385e-05	0.03060667	-2.442272e-05	9.077163e-05	-0.0001733941	1.249918e-04	-0.0002356860	0.0001872051	-0.0031281407
gdp_percap	-6.400185e-01	1.00000000	-6.400748e-01	1.962266e-02	-0.6722517703	-6.077633e-01	-0.6858136894	-0.5943289349	-0.7624603979
gini	3.852238e+00	1.00000000	3.852134e+00	1.354230e-01	3.6299234067	4.074648e+00	3.5370489027	4.1668965663	0.4007571369
mal_falciparum	-2.285057e+00	0.65942889	-1.526777e+00	2.776749e+00	-7.4232958219	4.903141e-01	-9.4268444876	2.5075981367	-0.0035734208
mean_annual_precip	-3.810528e-04	0.03053111	-3.810742e-04	2.275194e-04	-0.0007554246	-7.092767e-06	-0.0009124053	0.0001497397	-0.0233679512
hci	1.447094e+00	1.00000000	1.446677e+00	1.138138e-01	1.2603226575	1.634836e+00	1.1832105444	1.7134376970	0.2354817557
tavg	2.855834e-02	0.03259333	2.854794e-02	2.486625e-03	0.0244858518	3.266546e-02	0.0228167265	0.0344224579	0.2199976218
wci_index	3.626797e-03	0.03083111	3.621381e-03	8.965913e-03	-0.0110773143	1.838806e-02	-0.0172357625	0.0244937766	0.0047999961

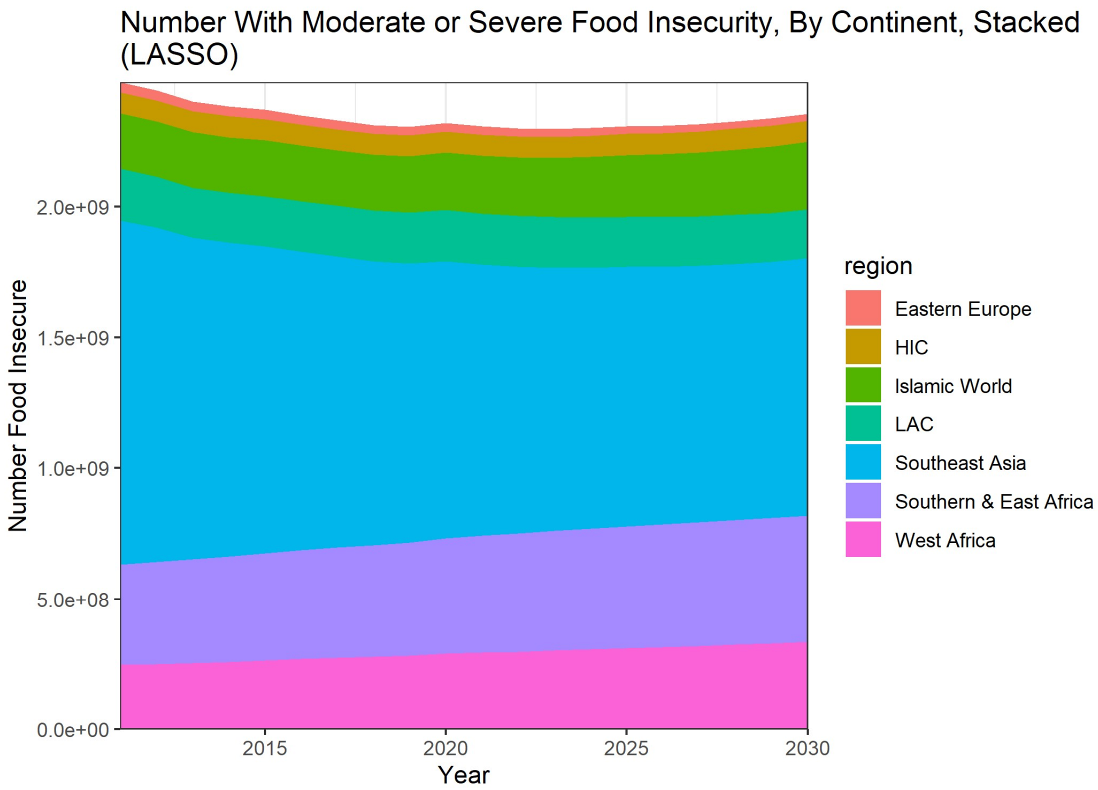
Results compared to LASSO


Change in Rate of Mod+Sev Food Insecurit
With increase of 1 SD in Var
For SSP2 LASSO Regression Model
(LASSO)
gini
hci
tavg
wci_index
school_mean
wasting
ruggedness
mean_annual_precip
mal_falciparum
urban_perc
gdp_percap
-0.75
-0.50
-0.25
0.00
0.25

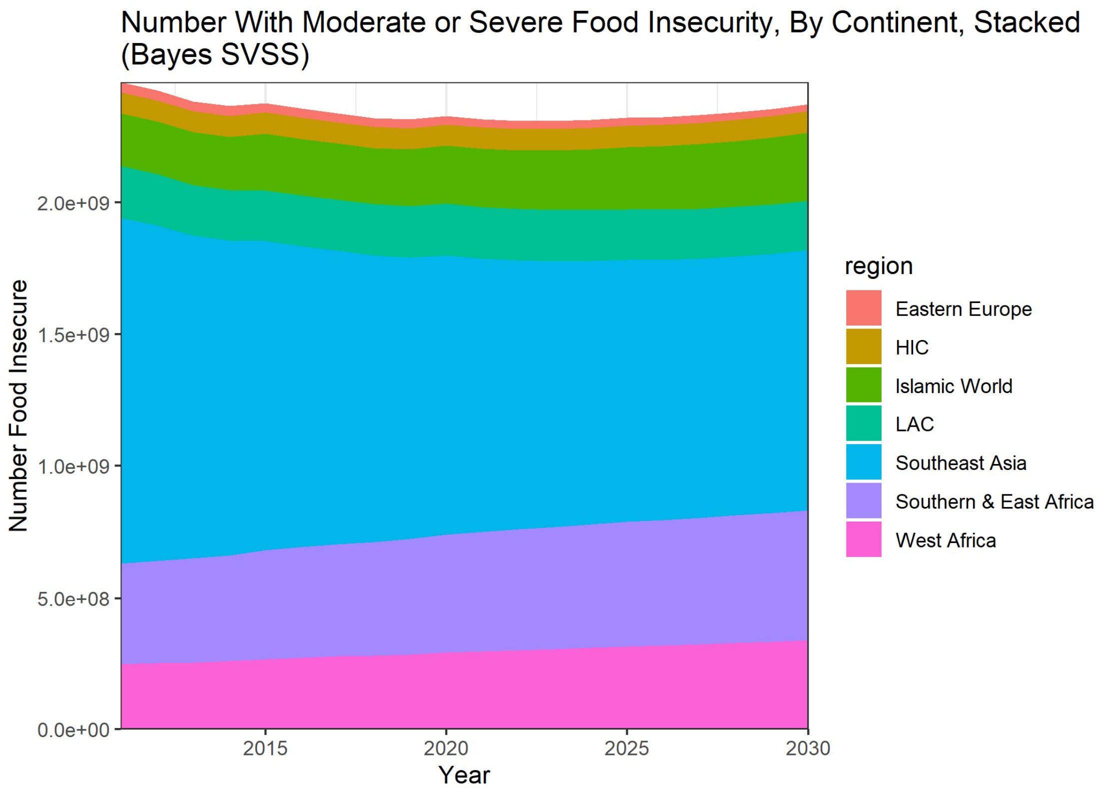

Change in Rate of Mod+Sev Food Insecurit
With increase of 1 SD in Var
For SSP2 LASSO Regression Model
(Bayes SSVS)
gini
hci
tavg
school_mean
wci_index
wasting
ruggedness
mal_falciparum
mean_annual_precip
urban_perc
gdp_percap
-0.8
-0.4
0.0
0.4


Model Performance (LASSO)
Modeled Rate of Mod+Sev Food Insecurity
0.00
0.25
0.50
0.75
1.00
Observed Rates of Mod+Sev Food Insecurity
Mean Absolute Error: 0.0651
R-squared: 0.922

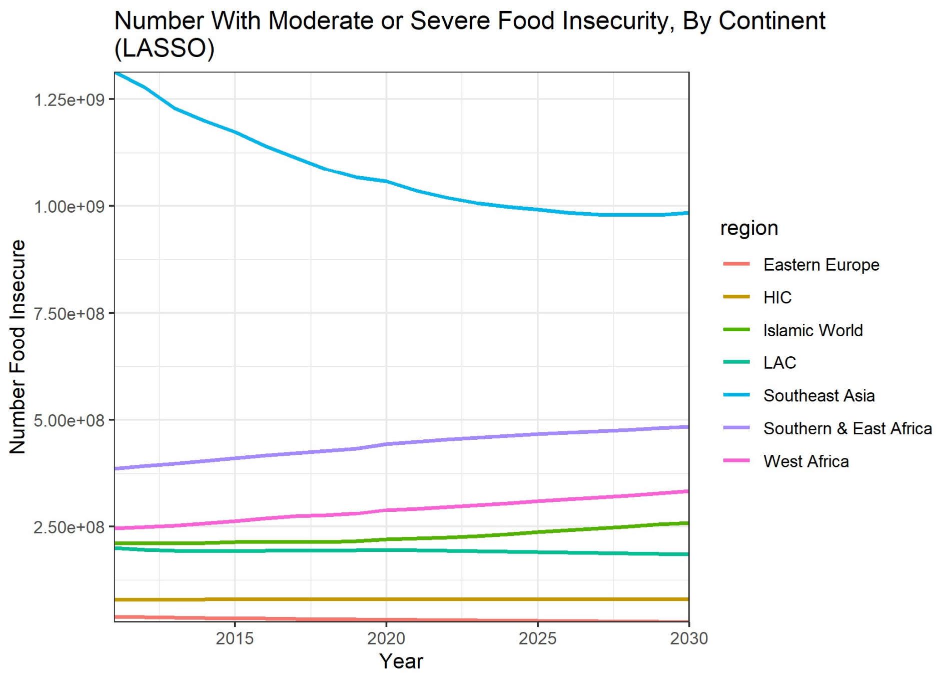

Model Performance (Bayes SVSS)
Modeled Rate of Mod+Sev Food Insecurity
0.00
0.25
0.50
0.75
1.00
Observed Rates of Mod+Sev Food Insecurity
Mean Absolute Error: 0.0648
R-squared: 0.9228

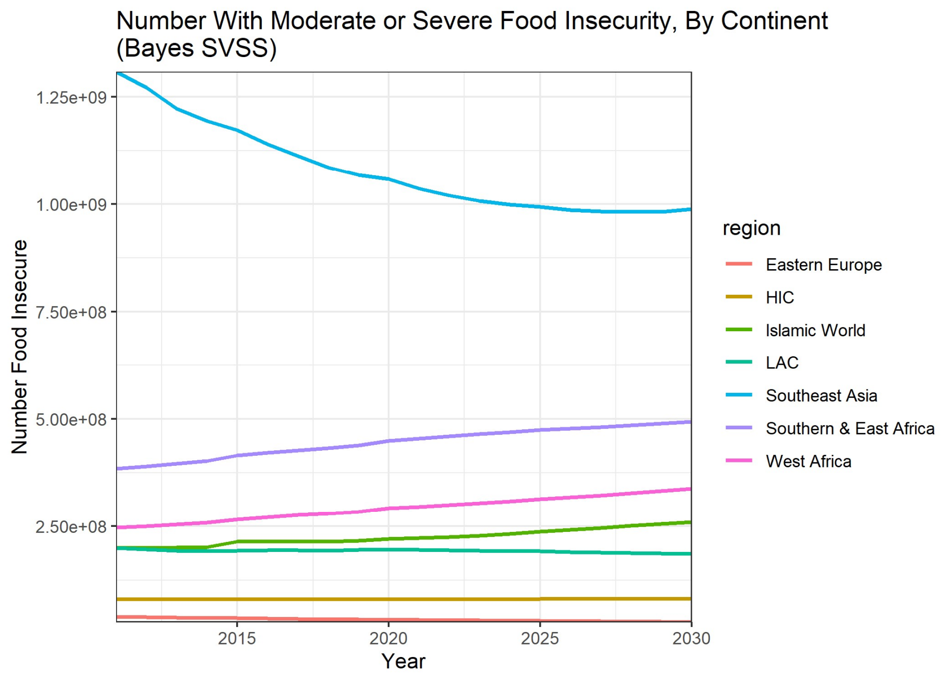


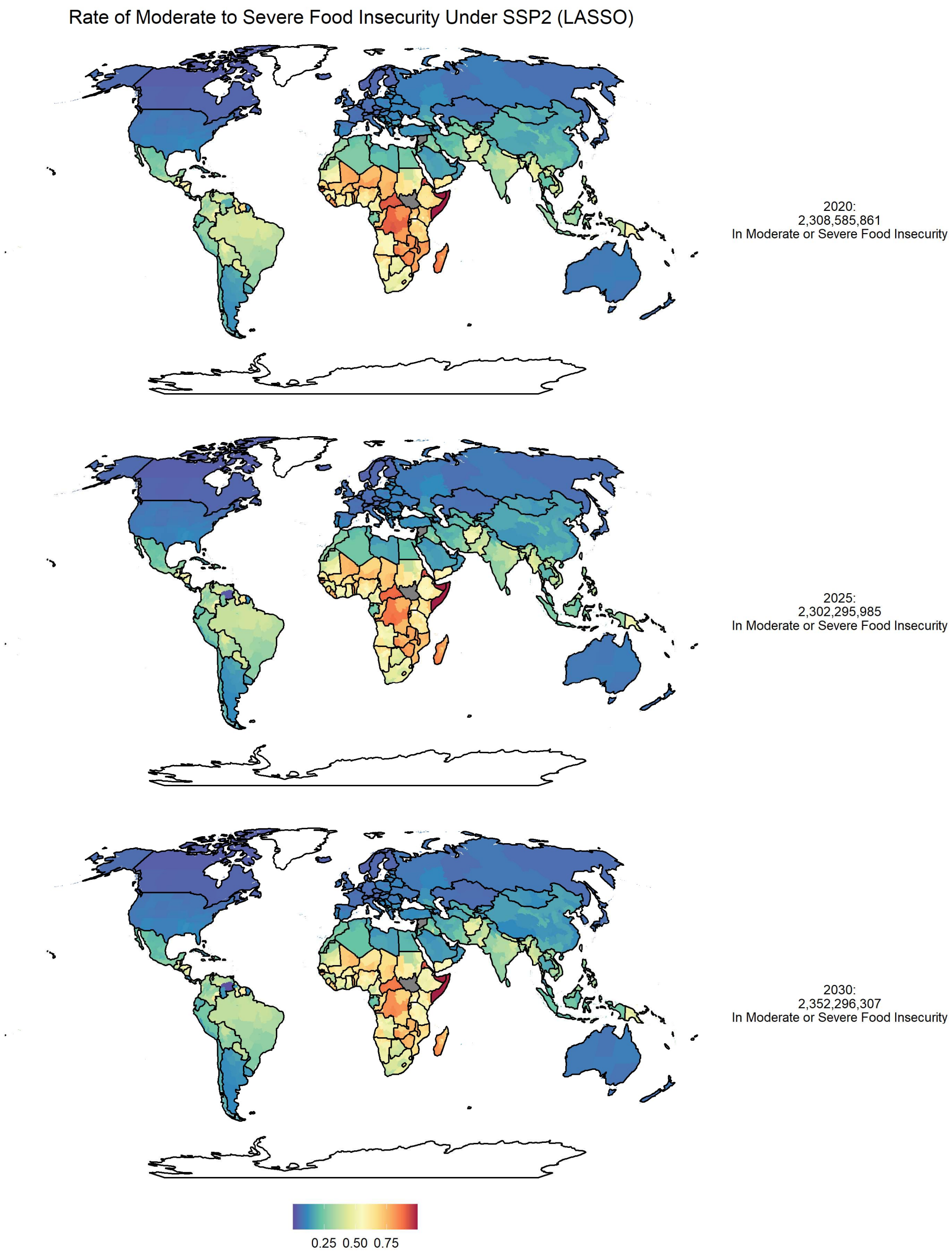
Number With Moderate or Severe Food Insecurity, By Continent, Stacked
(LASSO)
Number Food Insecure
0.0e+00
5.0e+08
1.0e+09
1.5e+09
2.0e+09
2015
2020
2025
2030
Year
region
Eastern Europe
HIC
Islamic World
LAC
Southeast Asia
Southern & East Africa
West Africa

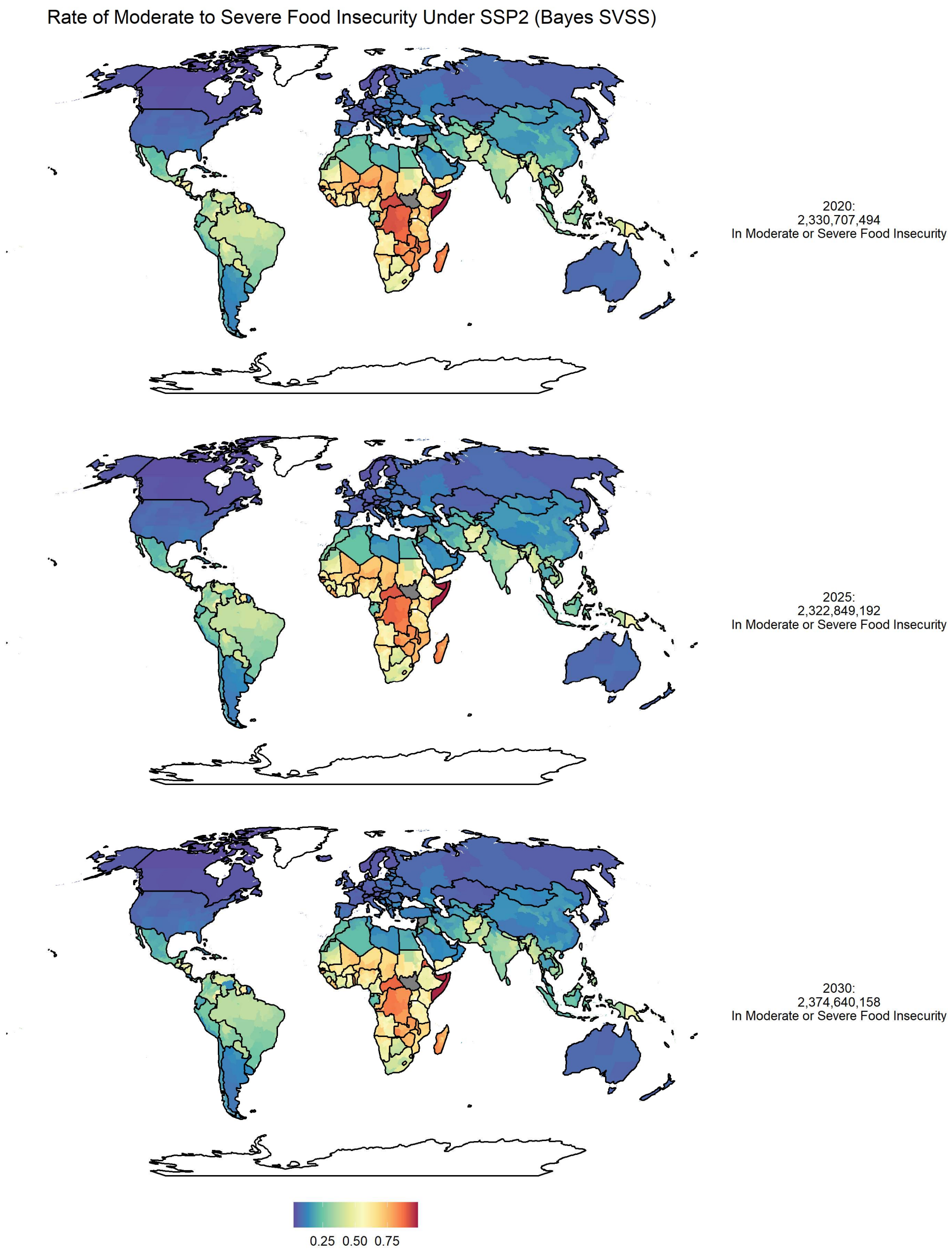


Number With Moderate or Severe Food Insecurity, By Continent, Stacked
(Bayes SVSS)
Number Food Insecure
0.0e+00
5.0e+08
1.0e+09
1.5e+09
2.0e+09
2015
2020
2025
2030
Year
region
Eastern Europe
HIC
Islamic World
LAC
Southeast Asia
Southern & East Africa
West Africa


Number With Moderate or Severe Food Insecurity, By Continent
(LASSO)
Number Food Insecure
1.25e+09
1.00e+09
7.50e+08
5.00e+08
2.50e+08
2015
2020
2025
2030
Year
region
Eastern Europe
HIC
Islamic World
LAC
Southeast Asia
Southern & East Africa
West Africa


Number With Moderate or Severe Food Insecurity, By Continent
(Bayes SVSS)
Number Food Insecure
1.25e+09
1.00e+09
7.50e+08
5.00e+08
2.50e+08
2015
2020
2025
2030
Year
region
Eastern Europe
HIC
Islamic World
LAC
Southeast Asia
Southern & East Africa
West Africa


Rate of Moderate to Severe Food Insecurity Under SSP2 (LASSO)
2020:
2,308,585,861
In Moderate or Severe Food Insecurity
2025:
2,302,295,985
In Moderate or Severe Food Insecurity
2030:
2,352,296,307
In Moderate or Severe Food Insecurity
0.25 0.50 0.75


Rate of Moderate to Severe Food Insecurity Under SSP2 (Bayes SVSS)
2020:
2,330,707,494
In Moderate or Severe Food Insecurity
2025:
2,322,849,192
In Moderate or Severe Food Insecurity
2030:
2,374,640,158
In Moderate or Severe Food Insecurity
0.25 0.50 0.75