Article

Sovereign Risk Indices and Bayesian Theory Averaging

Alex Lenkoski * and Fredrik L. Aanes

Norwegian Computing Center; Gaustadalleen 23a, Kristen Nygaards Hus, 0373 Oslo, Norway; fredrik@nr.no
* Correspondence: alex@nr.no; Tel.: +45-22-85-25-59

Abstract: In economic applications, model averaging has found principal use in examining the validity of various theories related to observed heterogeneity in outcomes such as growth, development, and trade. Though often easy to articulate, these theories are imperfectly captured quantitatively. A number of different proxies are often collected for a given theory and the uneven nature of this collection requires care when employing model averaging. Furthermore, if valid, these theories ought to be relevant outside of any single narrowly focused outcome equation. We propose a methodology which treats theories as represented by latent indices, these latent processes controlled by model averaging on the proxy level. To achieve generalizability of the theory index our framework assumes a collection of outcome equations. We accommodate a flexible set of generalized additive models, enabling non-Gaussian outcomes to be included. Furthermore, selection of relevant theories also occurs on the outcome level, allowing for theories to be differentially valid. Our focus is on creating a set of theory-based indices directed at understanding a country's potential risk of macroeconomic collapse. These Sovereign Risk Indices are calibrated across a set of different "collapse" criteria, including default on sovereign debt, heightened potential for high unemployment or inflation and dramatic swings in foreign exchange values. The goal of this exercise is to render a portable set of country/year theory indices which can find more general use in the research community.

Keywords: Bayesian model averaging; conditional Bayes factors; sovereign debt default; macroeconomic forecasting

1. Introduction

In economic applications, Bayesian Model Averaging (BMA) has proven a useful tool to assess theories related to the potentials and risks of economic expansion, see Steel (2019) for a comprehensive review. All economic theories are in some sense qualitative and no single empirical observation can encapsulate the theory's essence perfectly. To address this, a group of variables–self-evidently correlated–are often collected to proxy each theory. Not accounting for the uneven manner by which different variables may be available for each theory can lead to inappropriate conclusions regarding overall theory validity. Standard approaches to BMA can be modified, especially through the model prior, to account for these characteristics, but still consider the direct effect of the collected variables on the single response in question. One example is Chen et al. (2017), which consider the determinants of the 2008 crisis. They use a hierarchical formulation that allows for a simultaneous selection of both theories and relevant variables.

We propose an entirely separate approach to testing theories, both with regards to standard BMA and also to Chen et al. (2017), through a new model averaging approach. We assume each observation has a number of latent features encoding values for these theories. This requires the researcher to pre-specify which theory a given empirical observation is meant to proxy, a task which is often straightforward and frequently performed in practice. The outcome of this modeling exercise is a set of theory indices associated with each observation, as well as the model parameters necessary to derive

these indices for observations not included in training. Our second innovation is to link the embedding of empirical factors to theory indices across a number of correlated outcome variables. This is driven by a motivation for theory index consistency. Ideally, an index which assesses the strength of a government's institutions should be roughly the same when using the index to predict the potentials of economic growth and the susceptibility to economic collapse, for example. Indeed an ideal encoding would allow the theory index to be trained on one set of outcome variables and be immediately useful as a standalone feature in modeling separate but related economic activity. We therefore construct a framework by which theory-level modeling occurs on a latent level and is tuned to addressing a theory's role in explaining the variability of a number of economic outcome variables simultaneously. Brock et al. (2003) recommend considering both theory uncertainty (many theories can explain a phenomena) and variable (which empirical proxies should be used to explain each theory) uncertainty. Following this recommendation, model averaging in our Bayesian Theory Averaging (BTA) approach occurs on two separate levels. On the theory-level, a standard BMA formulation is used to determine which proxies for a given theory have the greatest relevance. Our modeling is across multiple different outcome variables and a given theory may only be relevant for a subset of these outcomes. Thus, we also perform theory averaging on the outcome level, allowing theories to selectively enter into each outcome under consideration.

Outcomes in economics can be quantified in a variety of manners and thus our framework is formulated to entertain a broader family of outcome sampling distributions than the Gaussian context to which most economic BMA applications have adhered (Steel 2019). Indeed, our framework is organized to accommodate all generalized additive models (GAMs) (see for example, Hastie and Tibshirani (1990) or Wood (2017)) and quantile GAMS (qgams) (Fasiolo et al. 2017). Operationally, the posterior model space is explored via Markov Chain Monte Carlo (MCMC), see for example, Gamerman and Lopes (2006) or Robert and Casella (2013), and model moves are efficiently performed via Conditional Bayes Factors (CBFs), (Karl and Lenkoski 2012), which have been shown to be highly useful in related model averaging exercises (Dyrrdal et al. 2015; Lenkoski 2013).

Our motivating example concerns developing useful theory-based indices for quantifying the potential for significant negative economic outcomes in macroeconomies, which we term Sovereign Risk Indices (SRIs). These outcomes range across default on Sovereign debt, the potential for high levels of inflation or unemployment, and heightened risks instability in foreign exchange. Useful introductions to sovereign default are found in Roubini and Manasse (2005) and Savona and Vezzoli (2015). Each of these outcomes have a number of theories which explain their variability. These theories encapsulate institutional and financial characteristics of each country and overall aspects of the global economy at the time and are proxied by a large number of potential variables. By modeling these outcomes jointly, we can construct a set of theory indices that are relevant for general research into macroeconomic extremes. Our goal is to create a broad database of SRIs that can then be made available to the general research population, where each index has a clearly defined construction and encodes a well-articulated theory regarding economic well-being. Our data combines the data in Savona and Vezzoli (2015) with new data sources, as explained in Section 3.1.

The structure of the article is as follows—Section 2 outlines BTA. The specifics of the algorithm that performs posterior inference for BTA is rather involved and relegated to Appendix A. Section 3 contains our analysis of the data which constructs the SRIs while Section 4 concludes.

2. Bayesian Theory Averaging

In this section we discuss our modeling framework. Our final modeling framework has multiple response variables and non-standard response likelihoods. However, the basic concepts behind BTA can be explained via a partitioning of a standard BMA problem and the addition of an intermediate random effects process. Therefore, Section 2.1 shows how standard BMA exercises can be grouped by similar variables, from which the indices that are our main focus naturally arise. Section 2.2 then develops the general joint modeling framework.

2.1. BTA and Linear Gaussian Regression

We start with a standard Gaussian regression exercise. Let $\mathbf{Y}$ be a length n univariate response with $Y_i \in \mathbb{R}$ and $\mathbf{X}$ be a $n \times p$ matrix of covariates. Furthermore let $M \subset \{1, \ldots, p\}$ be a model over a subset of the p potential covariates and $\mathbf{X}_M$ the sub-matrix of columns associated with the model M. The standard BMA regression with known variance is then

$$\mathbf{Y} = \alpha + \mathbf{X}_M \boldsymbol{\beta}_M + \boldsymbol{\epsilon} \tag{1}$$

$$\epsilon_i \sim \mathcal{N}(0, 1). \tag{2}$$

We note that fixing the variance ϵ_i to 1 in (2) is done for expositional convenience. It is not important to the developments of Section 2.2 to consider the general case of unknown variance. Under the g-prior (Zellner 1962)

$$\boldsymbol{\beta}_M \sim \mathcal{N}(0, g(\mathbf{X}_M' \mathbf{X}_M)^{-1}),$$

we have that the integrated likelihood of this model is

$$pr(M | \mathbf{Y}, \mathbf{X}) \propto |\Xi_M|^{1/2} \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}_M' \Xi_M^{-1} \hat{\boldsymbol{\beta}}_M\right), \tag{3}$$

where

$$\Xi_M = \frac{g+1}{g} \mathbf{X}_M' \mathbf{X}_M$$

$$\hat{\boldsymbol{\beta}}_M = \Xi_M^{-1} \mathbf{X}_M' \mathbf{Y}.$$

Now suppose that there is a natural partition of the p covariates into two groups, that is, the first p_1 columns of $\mathbf{X}$ belong to group 1 and the final p_2 columns ($p_1 + p_2 = p$) belong to group 2. Then instead of considering a single model $M \subset \{1, \ldots, p\}$, we could imagine there is a collection (M_1, M_2) of models with $M_1 \subset \{1, \ldots, p_1\}$ and $M_2 \subset \{p_1 + 1, \ldots, p_1 + p_2\}$. In many cases in BMA-driven studies, such a partition is natural since various concepts are proxied by collecting several features which are meant to encapsulate a given concept quantitatively. We therefore find it natural to discuss the model M_1 as the "theory one" model and the model M_2 as the "theory two" model.

We note that at this point, the integrated likelihood of $pr(M_1, M_2 | \mathbf{Y}, \mathbf{X})$ can be evaluated jointly and efficiently by (3). However, while there is no reason to do so, one could instead elect to update the models M_1 and M_2 separately.

In particular suppose that M_1 and $\boldsymbol{\beta}_1$ are given. Then

$$pr(M_2 | \boldsymbol{\beta}_1, M_1, \mathbf{Y}, \mathbf{X}) \propto |\Xi_{M_2}|^{1/2} \exp\left(\frac{1}{2} \hat{\boldsymbol{\beta}}_{M_2|M_1}' \Xi_{M_2}^{-1} \hat{\boldsymbol{\beta}}_{M_2|M_1}\right), \tag{4}$$

where

$$\Xi_{M_2} = \frac{g+1}{g} \mathbf{X}_{M_2}' \mathbf{X}_{M_2}$$

$$\hat{\boldsymbol{\beta}}_{M_2|M_1} = \Xi_{M_2}^{-1} \mathbf{X}_{M_2}' \mathbf{E}_1 \tag{5}$$

$$\mathbf{E}_1 = \mathbf{Y} - \mathbf{I}_1$$

and

$$\mathbf{I}_1 = \mathbf{X}_{M_1} \boldsymbol{\beta}_1. \tag{6}$$

Thus, we have effectively "separated" the response Y from the update of M_2 by replacing it with the residual calculation E_1 given the theory 1 parameter set. This leads to the alternative representation

$$Y = \alpha + I_1 + I_2 + \epsilon.$$

Thus, again while there is no need to do so, an MCMC for the overall BMA exercise could be conducted by alternating between updating model M_1 and thereby I_1, then updating model M_2 and I_2. These two summary variables I_1 and I_2 can then be referred to as the theory one and two indices respectively.

     In the Bayesian paradigm is it often natural to now incorporate a notion of over-dispersion. In particular, we can imagine that while $X_{M_1}\beta_1$ represents the "mean" theory one index given the features X_{M_1}, a random process adds a source of randomness to this mean level. It is therefore common to replace (3) with

$$I_1 = X_1\beta_1 + \eta_1$$
$$\eta_{1i} \sim \mathcal{N}(0, \nu_1^{-1}),$$

where the overdispersion parameter ν_1 can then be given a prior distribution, for example $\Gamma(a_1/2, b_1/2)$. A similar formulation can be made for I_2. In the context of econometric BMA exercises we feel such a random effects representation is imminently sensible as it implicitly admits that the features X_1 can only ever be imperfect encapsulations of a theory's essence.

     At this junction, the joint marginal likelihood (3) is no longer directly applicable. However, the conditional strategy of alternating between models M_1 and M_2 using (4) can still be used with an important modification. In particular, we note that given β_2, I_1, ν_2 we have

$$pr(I_2|Y, I_1, \nu_2) \sim \mathcal{N}((1+\nu_t)^{-1}(E_1 + \nu_t X_2\beta_2), (1+\nu_t)^{-1}).$$

Furthermore, given I_2 we may replace (5) with

$$\hat{\beta}_{M_2|M_1} = \Xi_{M_2}^{-1} X'_{M_2} I_2.$$

Subsequent to the sampling of the latent factors I_2 we may resample the random effects precision parameters ν_t via a standard Gibbs step.

     Indeed, we could then consider one final embellishment where

$$Y = \alpha + \gamma_1 I_1 + \gamma_2 I_2 + \epsilon,$$

with $\gamma_t \in \{0, 1\}$ with for example, prior probability that $\gamma_t = 1$ set to $1/2$ (or any other value in $(0,1)$). Then when $\gamma_2 = 0$ the update of I_2 would simply be

$$pr(I_2|Y, I_1, \nu_2, \gamma_2 = 0) \sim \mathcal{N}(X_2\beta_2, \nu_t^{-1}),$$

that is, a sample from the prior conditional on β_2. Updating the parameter γ_2 conditional on all other factors would then involve a straightforward Metropolis-Hastings step. If the models M_1 and M_2 indicate which variables are included in the theory one and theory two models, the γ_1 and γ_2 act as a wholesale inclusion parameter which dictates the overall relevance of the respective theory.

     This partitioning and random effects strategy forms the basis of our development in Section 2.2. We note that the inclusion of the random effects component has the effect of keeping model evaluations conditionally Gaussian, which enables the use of conditional Bayes factors to efficiently resample model parameters.

2.2. Multivariate BTA and Generalized Regression Models

We now generalize to the case where we have R responses from a general response family. Let $\mathbf{Y}_i$ be an R dimensional response vector for observation i and $\mathcal{D} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_n\}$ be a collection of n such observations. Each variate Y_{ir} in the vector $\mathbf{Y}_i$ is assumed to belong to a general field $\mathcal{F}_r$. In this paper we consider examples where $\mathcal{F}_r$ is $\{0, 1\}$, $\mathbb{R}$ and $\mathbb{R}_+$, though others such as $\mathbb{N}$, could easily be entertained. We associate Y_{ir} with an outcome distribution as

$$Y_{ir} \sim g_r(\boldsymbol{\alpha}_r, \mu_{ir}),$$

where g_r is a general probability density or mass function, $\boldsymbol{\alpha}_r$ is a set of global parameters and μ_{ir} is an observation i dependent mean value. We note that the assumption that only the mean parameter μ_{ir} varies according to the observation i could be relaxed in future work.

The parameter μ_{ir} is then assumed to have the form

$$\mu_{ir} = \sum_{t=1}^{T} \gamma_{rt} I_{it}.$$

In the above formulation γ_{rt} can either be 0 or $\gamma_{rt} \in \mathbb{R}$. We assign a prior probability of $1/2$ to these two possibilities, clearly other prior probabilities could be entertained. By convention if several γ_{rt} are non-zero for a given index t then one of these non-zero γ_{rt} is set to 1 to avoid issues related to identification. This matter is discussed subsequently.

The variable I_{it} is then referred to as the theory-t index for observation i. We further assume that the I_{it} depends on a set of p_t theory proxies $\mathbf{X}_{it}$ according to the linear model

$$I_{it} = \mathbf{X}'_{it} \boldsymbol{\beta}_t + \epsilon_{it},$$

where $\epsilon_{it} \sim \mathcal{N}(0, \nu_t^{-1})$ independently. The precision term ν_t is assigned a $\Gamma(a_t/2, b_t/2)$ prior. We note that this prior actually is forced to adapt throughout the procedure (by adjusting the a_t, b_t parameters) to control for issues of identification, we discuss this aspect below. We typically begin the inference procedure setting a_t and b_t to 1.

Associated with the parameter $\boldsymbol{\beta}_t$ is a model $M_t \subset \{1, \ldots, p_t\}$ such that $\beta_{it} = 0$ when $i \notin M_t$, a standard BMA formulation. As the "null" model can be controlled by the γ_{rt} parameter, we exclude $M_t = \varnothing$ from our consideration, see [Kourtellos et al. (2019)](#) for a motivation of this structure. Writing $\boldsymbol{\beta}_{M_t}$ to represent the subvector of $\boldsymbol{\beta}_t$ not constrained to zero we assume

$$\boldsymbol{\beta}_{M_t} \sim \mathcal{N}(0, \nu_t^{-1} g_t (X'_{M_t} X_{M_t})^{-1}),$$

where p_{M_t} is the size of model M_t and independent across t. As with the prior parameters a_t, b_t, the g-prior parameter g_t adapts throughout the procedure, we begin with $g_t = 1/n$. Alternative priors for this model could have been considered, see our discussion in the Conclusions section.

Finally, the model M_t can have a number of priors, see [Ley and Steel (2009)](#) for an overview of potential issues to consider when selecting this prior. For the time being we choose the uniform prior

$$pr(M_t) \propto 1.$$

When $\gamma_{rt} \in \mathbb{R}$ we assign the prior probability $\gamma_{rt} \sim \mathcal{N}(0, 1)$. This has the effect of imposing a uniform model prior on the inclusion of theories in the outcome equation. Alternatively, joint priors for the γ factors could be considered which would control for the size of the included theories. However, since the number of theories is meant to be modest (roughly five to ten), we have avoided such aspects in the current framework.

The system outlined above then serves as the core latent process which drives the subsequent outcome variables. Thus we see that the models M_t investigate which proxies best encode a theory quantitatively while also accounting for the obvious model uncertainty in this formulation and incorporating a notion of over-dispersion. The γ_{rt} terms serve two purposes. First, by examining their non-zero elements we see for which response equations a given theory is relevant. Secondly, by requiring the first non-zero γ_{rt} to be equal to 1 and all others to be in $\mathbb{R}$ the γ_r term scales the latent indices to allow them to enter into model parameters differentially and indeed in opposite directions.

Finally, the latent theory indices I_{it} are potentially of greatest interest, as they are meant to encapsulate the way that the theory proxies affect the outcome equations of interest. Again, as outlined in the Appendix, these terms suffer from potential identification issues when combined with the restrictions placed on a given γ_r. The hyperparameters a_t, b_t ultimately control this aspect and therefore, final interest focuses on the scale-free term $\tilde{I}_{it} = (a_t/b_t)I_{it}$.

This concern regarding identification requires a modicum of bookkeeping when conducting posterior inference. If, for example all non-zero γ values were allowed to be in $\mathbb{R}$ then the final outcome equation could have a variety of γ_{rt} and $\boldsymbol{\beta}_t$ combinations that would yield the same posterior probability. This is the justification for our restriction that the γ_{rt} with the smallest r be constrained to 1.

However, this constraint yields its own issues, primarily due to its effects on the priors for the β and ν parameters. If, for example, $\gamma_{11} = 1$ and $\gamma_{21} = 0.5$ and our chain sets γ_{11} to 0, γ_{21} will suddenly double. This would imply that $\gamma_{21}I_1$ will suddenly have twice the effect on the mean value of outcome Equation (2). The obvious answer is to simultaneously halve I_1, or equivalently, halve $\boldsymbol{\beta}_1$. However, it would no longer be appropriate to keep the priors for $\boldsymbol{\beta}_t$ and ν_t fixed and therefore their priors are also adjusted by this factor. Technical details are given in Appendix A.

To review, our full modeling framework therefore takes the form

$$Y_r \sim g_r(\boldsymbol{\alpha}_r, \mu_{ir})$$
$$\mu_{ir} = \sum_{t=1}^{T} \gamma_{rt} I_{it}$$
$$pr(\gamma_{rt} = 0) = \frac{1}{2}$$
$$pr(\gamma_{rt}|\gamma_{rt} \neq 0) \sim \mathcal{N}(0,1) \text{ unless } \gamma_{rt} = 1$$
$$I_{it} \sim \mathcal{N}(\boldsymbol{X}_{it}'\boldsymbol{\beta}_{M_t}, \nu_t^{-1})$$
$$\boldsymbol{\beta}_{M_t} \sim \mathcal{N}(0, g_t(\boldsymbol{X}_{M_t}'\boldsymbol{X}_{M_t})^{-1})$$
$$pr(M_t) \propto 1$$
$$pr(\nu_t) \sim \Gamma(a_t/2, b_t/2).$$

Choices for families g_r that control the outcome variables are considerable. In our application, we focus on three models. The first is logistic regression. In this case $Y_{ir} \in \{0, 1\}$, α_r is univariate and

$$pr(Y_{ir} = 1) = \frac{\exp(\alpha_r + \mu_{ir})}{1 + \exp(\alpha_r + \mu_{ir})}.$$

We use this logistic regression to model the probability that a country will default on its sovereign debt based on theory-indices.

The second family considered corresponds to the non-central asymmetric Laplace variates. In this

case $\boldsymbol{\alpha}_r$ is two dimensional with α_{r1} denoting the intercept and α_{r2} the log-precision parameter. In particular, we write

$$pr(Y_{ir}|\boldsymbol{\alpha}_r, \mu_{ir}\tau) = \tau(1-\tau)\exp\left\{\alpha_{r2} - e^{\alpha_{r2}}\rho_\tau(Y_{ir} - \alpha_{r1} - \mu_{ir})\right\}$$
$$\rho_\tau(x) = x(\tau - \mathbf{1}\{x < 0\}),$$

where τ is the quantile under consideration. This model is often referred to as the Bayesian Quantile Regression since its posterior mode is related to the quantile regression estimate under the so-called *pin-ball* loss ρ_τ. We employ this model for two separate variates, the inflation and unemployment rates and set $\tau = 0.9$ for both, thus focusing on the 90th percentile of the respective distributions.

Finally, we consider the Generalized Extreme Value (GEV) model with $\boldsymbol{\alpha}_r = (\alpha_{r1}, \alpha_{r2}, \alpha_{r3})$ parameterized by

$$pr(Y_{ir}|\boldsymbol{\alpha}_r, \mu_{ir}) = e^{\alpha_{r2}}h(Y_{ir})^{-(\alpha_{r3}+1)/\alpha_{r3}}\exp\left(-h(Y_{ir})^{-\alpha_{r3}^{-1}}\right),$$

for $h(Y_{ir}) > 0$ with

$$h(Y_{ir}) = 1 + \alpha_{r3}e^{\alpha_{r2}}(Y_{ir} - \alpha_{r1} - \mu_{ir}).$$

The GEV model is used to model block maxima and hence understand the nature of extreme behavior. In our case we use it to model the largest daily percentage jump in a country's exchange rate (relative to USD) seen over the course of a year. The global parameters α_{r2} and α_{r3} are the log precision and shape respectively while α_{r1} again serves as the global intercept.

Based on $\mathcal{D}$ we then conduct posterior interference on the full parameter set, which includes global parameters $\boldsymbol{\alpha}_r$, the theory-level models M_t, theory-inclusion and scaling parameters γ_{rt} and linear model parameters $\boldsymbol{\beta}_t$ as well as the latent indices $\boldsymbol{I}_t$ and their random effect variances ν_t. Posterior inference is performed via Markov Chain Monte Carlo (MCMC). Given the involved and nested nature of the MCMC, several different approaches are employed at different stages of the hierarchy and the full details are provided in Appendix A.

However, the main themes of the MCMC involve conditional Bayes Factors (CBFs) to change models M_t and update proxy regression parameters $\boldsymbol{\beta}_t$. Standard block Metropolis-Hastings proposals using local Laplacian calculations of the log posterior density are used to update latent theory indices I_{it} as well as any global parameters in $\boldsymbol{\theta}_r$. Finally, reversible jump methods (Green 1995) alternate γ_{rt} between being 0 or in $\mathbb{R}$, with a modicum of book keeping to ensure that at least one γ_{rt} is set to 1 when theory t is represented in more than one dependent equation r, again to ensure identification of the system. When conducting this bookkeeping exercise, prior distributions are adjusted accordingly to ensure that log-posterior density values are not affected by mere changes in variate representations.

3. Using BTA to Construct Sovereign Risk Indices

3.1. Data Outline

Our dataset for constructing SRIs originated from the dataset in Savona and Vezzoli (2015). Savona and Vezzoli (2015) track 70 countries between the years of 1975 and 2010. We have extended these original data to 2018 and are primarily focused on whether a country defaults on its sovereign debt in a given year. Some country year combinations are not present and thus $n = 2032$. To model this default probability, Savona and Vezzoli (2015) collect 27 covariates. These covariates are meant to proxy 5 different theories related to sovereign debt default. In particular, they entertain the concepts of Insolvency, Illiquidity, Macroeconomic Factors, Political Factors and Global Systemic factors. Table 1 provides an overview of the 27 covariates considered and to which theory Savona and Vezzoli (2015) associate them. For a given year, most covariates are "lagged"(except for contagion, dummy for oil and dummy for international capital markets), in that these values would be available at the start of a given

year, as opposed to co-occurring with the default event. Covariate missingness is prevalent, we derive imputed values using the semi-parametric Gaussian copula of Hoff (2007) within each theory-group.

Table 1. Overview of variables considered in the SRI dataset.

Theory	Short Variable Name	Description
Insolvency	MAC	Market access to capital markets, dummy
Insolvency	IMF	IMF lending dummy
Insolvency	CAY	Current account balance, in % of GDP
Insolvency	ResG	Reserves growth/change in %
Insolvency	XG	Export growth in %
Insolvency	WX	Export in USD billions
Insolvency	TEDX	Total external debt to exports, in %
Insolvency	MG	Import growth, in %
Insolvency	FDIY	Foreign direct investment to GDP, in %
Insolvency	FDIG	Change in % of foreign direct investment inflows
Insolvency	TEDY	Total external debt to GDP, in %
Insolvency	SEDY	Short term external debt to GDP, in %
Insolvency	PEDY	Public external debt to GDP, in %
Insolvency	OPEN	Exports and imports over GDP, in %
Illiquidity	STDR	Short term debt to reserves
Illiquidity	M2R	M2 to reserves
Illiquidity	DSER	Debt service on long term debt to reserves
Macroeconomic	DOil	Oil producing dummy
Macroeconomic	RGRWT	Real (inflation adjusted) GDP change in %
Macroeconomic	OVER	Exchange rate residual over liner trend
Macroeconomic	UST	US treasury bill
Political	PR	Index of political rights, 1 (most free) to 7 (least free)
Political	History	Number of past defaults
Systemic	Cont_tot	Number of defaults in the world
Systemic	Cont_area	Number of defaults in the region the country is part of

Savona and Vezzoli (2015) is concerned with predictive models of sovereign default and therefore solely focus on this binary outcome. We augment the default binary with three other measures that can also indicate a macroeconomy in a state of collapse. First, the country's lagged (i.e., one-year behind) inflation rate was originally included in the Macroeconomic factors group of covariates in Savona and Vezzoli (2015). We instead treat (non-lagged) inflation as another dependent variable and note that doing so has no effect on the Default outcome; a run of BTA solely on Default with inflation included in the Macroeconomic factors gave this variable 0 inclusion probability. In addition, we collected unemployment data from the IMF website. These data were only available for a subset (897 country/year pairs) of the total data. We note that this dependent variable missingness poses no substantive problem in terms of the derivation of SRIs. The BTA approach simply ignores the missing likelihood contributions necessary to update the associated latent theory indices.

Finally, we collected foreign exchange rate data from the website of the IMF. For each country/year pair, we first computed the log rate change relative to the US dollar and then used the annual maximum of these log changes. This variable therefore shows the largest single-day weakening of a currency relative to the US dollar in the course of a year. We avoided commercial sources of foreign exchange data and therefore only had these values for 272 country/year pairs. See our discussion in the conclusions section regarding expanding these data.

A country is in default if it is classified so by Standard & Poor's (SP) or if it receives a large nonconcessional loan by the IMF. A nonconcessional loan is a loan that has the standard IMF's market-related interest rate, while a concessional loan has a lower interest rate. The type of loans we consider from IMF must be in excess of 100 percent of quota. Each member country has a quota, where the initial quota is set when a country joins IMF. The quota determines, by other things, the country's access to IMF loans and for instance its voting power. By augmenting the data from

SP with data from IMF, we also capture near-defaults or debt restructurings avoided through loan packages from IMF. We consider Stand By Agreements (SBA) and Extended Fund Facility (EFF).

Our posterior inference is performed after running the BTA algorithm for 400 thousand iterations over these data. In order to verify convergence, 30 separate runs of the algorithm were run simultaneously and the resulting output was inspected to verify posterior inference for each individual chain was nearly identical. Runtime on a 32 core machine (dual 8-core 3.4 GHz AMD Ryzen processors with hyperthreading capabilities) with 128 GB of RAM was roughly 7.5 h. Runtime of a single chain (as opposed to all 30) on a Macbook with 8 GB RAM and a 2.4 GHz dual-core processor is roughly similar, indicating that specialized hardware is not necessary.

3.2. Results

We begin our discussion of the SRI results by investigating outcomes on the theory level. Table 2 shows the theory inclusion probability (i.e., the proportion of time that γ_{rt} was not constrained to zero in the chain) for each theory, across the four outcome variables. Given that the original dataset was constructed to model the default variable, it is unsurprising that all theories achieve inclusion probabilities of one for this outcome. We note, however, that this does not indicate that all theories are equally strong in explaining default, simply that none of them can be considered irrelevant. The inflation outcome is interesting in that it suggests that proxies measuring a country's political stability, macroeconomic and system factors have the best explanation for the upper tails of the inflation distribution. Insolvency and Illiquidity theories are also relevant to inflation, achieving probabilities between 0.64 and 0.367. The results for unemployment in Table 2 finds little inclusion for the Insolvency index, while all others achieve inclusion of 1. Finally, we achieve relatively low inclusion probabilities for all theories for the devaluation outcome. This is likely due in part to the relatively small amount of data that was available using public sources, see our discussion in Section 4. However, we feel this result highlights a useful feature of BTA, namely that including this outcome variable and having the system set theory-inclusion probabilities to zero meant there was no subsequent effect on the calculation of theory indices.

Table 2. Theory Inclusion Probabilities by Dependent Variable.

	Default	Inflation	Unemployment	Devaluation
Insolvency	1	0.643	0.006	0.036
Illiquidity	1	0.367	1	0.061
Macroeconomic	1	1	1	0.114
Political	1	1	1	0.052
Systemic	1	0.996	1	0.036

Table 3 shows the mean value (conditional on inclusion) of the parameter γ_{rt} for each theory and outcome pair. Since Default was ordered first in our system and achieves inclusion probabilities of 1 for all theories, this system serves to orientate the rest of the outcomes. In particular, a positive γ_{rt} for the outcomes indicates that the directionality of this theory on the outcome is similar to that of default. The conditional means even more clearly the importance of the Macroeconomic, Political and Systemic theories relative to the remaining two. The value of their condition means (between 1.4 and 2.5) is substantially higher than 0.451 and 0.124 for Insolvency and Illiquidity respectively. Since these three theories achieved substantially higher inclusion probabilities in Table 2 this implies that their unconditional effect is the main driver of the upper tail of inflation.

Recalling again from Table 2 that the unemployment outcome was driven by the Illiqudity, Macroeconomic, Political factors and Systemic Factors, the results in Table 3 are interesting. They show that Macroeconomic and Systemic have a strong, positive effect on the unemployment outcome. We note that "positive" is in the sense of working in the same direction as Default. The Illiquidity and Political factors then balance this behavior; they are negatively orientated to the impact these factors

have on Default. Finally, as noted in Table 2 there appears to be negligible effect of the theory indices on the devaluation outcome.

Table 3. Mean value of γ_{rt} conditional on inclusion.

	Default	Inflation	Unemployment	Devaluation
Insolvency	1	0.451	0	0
Illiquidity	1	0.124	−0.598	−0.002
Macroeconomic	1	2.24	1.775	0.007
Political	1	2.499	−1.644	−0.002
Systemic	1	1.422	0.946	0

Table 4 shows the inclusion probabilities and conditional posterior mean for each proxy contained in the Insolvency theory group. Five factors achieve probabilities above 0.98. These include one factor that measures the strength of the country's balance sheet (ResG), two factors describing the country's trade balance dynamics (XG and MG) and finally two factors describing the state of foreign direct investment (FDIY and FDIG). Interestingly, features that assess the country's debt load are included in the posterior to an appreciable degree.

Table 4. Proxy Level Results for the Insolvency Theory.

	Probability	Conditional Mean
IMF	0.037	−0.024
CAY	0.04	0.025
ResG	0.991	0.082
XG	0.985	0.099
WX	0.023	0.003
TEDX	0.084	−0.043
MG	1	−0.397
FDIY	0.994	−0.103
FDIG	1	−0.249
TEDY	0.087	−0.052
SEDY	0.1	0.06
PEDY	0.084	−0.039
OPEN	0.268	−0.054

Table 5 shows the inclusion probabilities for the Illiquidity theory. In contrast to the balanced view offered in the Insolvency results of Table 4, Table 5 puts almost all weight on a single feature, a measure of a country's shot-term cash and cash equivalents to reserves (M2R).

Table 5. Proxy Level Results for the Illiquidity Theory.

	Probability	Conditional Mean
STDR	0.045	−0.026
M2R	0.971	0.082
DSER	0.054	0.04

Table 6 shows the inclusion probabilities for the proxies in the Macroeconomic grouping. We see that measures related to inflation dynamics (RGRWT) and the rate on US treasuries (UST) are given inclusion probabilities of one while the other factors are given low inclusion probabilities.

Table 6. Proxy Level Results for the Macroeconomic Theory.

	Probability	Conditional Mean
DOil	0.05	0.028
RGRWT	1	−0.466
OVERN	0.022	0.011
UST	1	0.175

Tables 7 and 8 show the inclusion probabilities for proxies of the Political and Systemic theories respectively. In each theory there are only two features and all four receive inclusion probabilities of 1. We see that the Political theory is thus a blend of the Political rights index (PR) and a measure of past susceptibility to default (History). Likewise, a measure of global contagion (Cont_tot) as well as local factors (Cont_area) combine to form the Systemic theory.

Table 7. Proxy Level Results for the Political Theory.

	Probability	Conditional Mean
PR	1	0.095
History	1	0.213

Table 8. Proxy Level Results for the Systemic Theory.

	Probability	Conditional Mean
Cont_tot	1	0.529
Cont_area	1	0.215

Our results echo many of the main themes in Savona and Vezzoli (2015), in that variables from the Illiqudity, Insolvency and Systemic theories are included in both cases. However, Savona and Vezzoli (2015) find no inclusion of variables from the macroeconomic or political theories, while these theories are included with probability one in our results. We note that this is likely to partially due to overall model size; Savona and Vezzoli (2015) only include 6 of the 26 variables in their tree-based approach. In contrast, the average total model size using BTA was 12.8 variables, with all iterations having between 11 and 17 variables included.

We conclude by investigating detailed results for two of the theories, namely Insolvency and Illiquidity. Table 9 shows the country/year pairs with the five lowest and five highest posterior mean values of I_{it} for the insolvency theory. The lowest five country/year pairs listed represent the countries whose Insolvency index indicates the lowest probabilities of default. Interestingly, Gabon is represented twice amongst these five countries (for the years 1981 and 1995), which is unsurprising given the country's oil wealth and relative aggregate prosperity amongst African nations. Amongst the five country/year pairs with the highest Insolvency index scores we see a mix of African (Tunisia 1988; Niger 1983) Caribbean (Trinidad and Tobago 1987; Haiti 1979) and Southeast Asian (Sri Lanka 2009) countries. Two of the five (i.e., 40%) of these pairs experience a default, which is substantially higher than the 6% average over all the data, showing the degree to which this feature is positively associated with default.

Table 10 shows the five highest and lowest country year pairs according to the Illiquidity index. Burundi in 1991 (i.e., two years before the start of the civil war that ran between 1993 and 2005) receives the lowest Illiqudity index, otherwise followed by countries in South Asia. On the highest end, we see both Jamaica and Lesotho represented twice. In addition, Gabon in 2002 is present, a year in which the country defaulted on its sovereign debt. This contrast to Table 9 is illuminating, as it shows the trade off between potential for insolvency and risks of illiquidity in precipitating sovereign default. We note again that two of the top five country year pairs record a default, similar to the results of Table 9. However, when inspecting the unemployment result, we also see high levels of unemployment for four of the five top countries (and a missing value for Gabon 2002, the remaining country). Simultaneously

the countries with the lowest illiquidity indices have negligible unemployment rates. This lines up with the results presented in Table 4, where the insolvency index had a large, positive effect on the unemployment outcome equation.

Table 9. Highest and Lowest Five Country/Year Pairs for the Insolvency Theory.

	Year	Insolvency	Default	Inflation	Unemployment	Devaluation
Gabon	1995	−13.683	0	36.116	NA	NA
Moldova	1994	−6.704	0	35.749	NA	NA
Korea, Rep.	2009	−6.157	0	4.704	3.2	0.071
Portugal	1995	−5.664	0	5.214	6.713	0.018
Gabon	1981	−5.543	0	12.34	NA	NA
Trinidad and Tobago	1987	4.407	0	7.694	9.37	NA
Tunisia	1988	4.437	1	8.226	NA	NA
Sri Lanka	2009	4.448	0	22.564	5.22	0.012
Niger	1983	4.703	1	11.642	NA	NA
Haiti	1979	4.864	0	−2.674	NA	NA

Table 10. Highest and Lowest Five Country/Year Pairs for the Illiquidity Theory.

	Year	Illiquidity	Default	Inflation	Unemployment	Devaluation
Burundi	1991	−5.391	0	7.002	0.48	NA
Pakistan	1976	−5.223	0	20.905	1.7	NA
Bangladesh	1997	−5.154	0	2.377	2.51	NA
Malaysia	2007	−5.079	0	3.609	3.32	0.012
Indonesia	1977	−5.07	0	19.859	1.92	NA
Jamaica	1986	7.266	0	25.673	33.39	NA
Lesotho	2009	7.311	0	10.721	35.46	NA
Lesotho	1998	7.329	0	−100	37.94	NA
Gabon	2002	8.162	1	2.138	NA	NA
Jamaica	1981	9.208	1	27.308	35.51	NA

Finally, we address an issue related to theory index portability. In the theoretical construction of these indices we specified an independence structure between indices. However, there has been nothing enforcing this condition in posterior estimation. If theory indices were correlated in the posterior, this would be acceptable, however it would imply that these indices would need to be included as a set when attempting to model other phenomena. Table 11 suggests such considerations are likely unnecessary. In Table 11 we show the posterior correlation matrix of the theory indices, estimated over all samples and country/year pairs. We see in general a low degree of correlation (the entry −0.161 between the Macroeconomic and Political theories being the highest in absolute value). This feature is desirable, since it suggests that the theory indices can be used on an individual basis for subsequent modeling of other issues related to economic collapse.

Table 11. Posterior correlation matrix of theory indices. This table mainly shows that the indices have the desirable property of low dependence between one another.

	Insolvency	Illiquidity	Macroeconomic	Political	Systemic
Insolvency	1	−0.015	0.024	−0.012	0.002
Illiquidity	−0.015	1	−0.061	0.066	−0.021
Macroeconomic	0.024	−0.061	1	−0.161	0.081
Political	−0.012	0.066	−0.161	1	−0.095
Systemic	0.002	−0.021	0.081	−0.095	1

3.3. Investigating the Multiple Response Framework

One of the innovations of the BTA framework has been the use of multiple responses to jointly determine the parameters β_t and implicitly the theory indices I_i. This section investigates the advantages of this joint modeling in the context of an out of sample prediction exercise. In this study we conduct a leave-one-country-out cross validation. For each country we create a training dataset which excludes all observations from that country. We then fit five models. The first is the full specification described above with all four responses contained in the framework. We then fit versions of BTA including each of the responses individually.

For the country that has been left out, we then derive the fitted values, based on their features for each of the four responses from the joint model as well as each model individually. These fitted values are then scored relative to the observed value by appropriate proper scoring rules, namely the brier score for the binary response (default), the quantile score for the two quantile regressions (unemployment, inflation) and the likelihood score for the GEV regression. A permutation test (see e.g., a similar procedure in Möller et al. (2013)) assesses the significance of these results.

Table 12 shows the mean scores across all countries in the joint and single model cases for each of the four response variables. The results for the default variable are nearly identical. While multiple variables did not aid predictive performance, this result is still positive as it indicates that performance was not hampered by their inclusion. However, for the two quantile regressions we see a substantial improvement in predictive performance when using the model output from the joint model with the GEV regression on devaluation substantially improved when using the joint model. We feel the results of Table 12 provide a strong indication of the usefulness of jointly modeling several response variables together.

Table 12. Comparison of out of sample scores for each dependent variable. This table shows that, with the exception the default variable, there is significant reduction in predictive loss when using a joint modeling framework. Asterisks indicate greater than 99% significance of differences based on a permutation test.

Variable	Joint Model	Single Model	Ratio Joint to Single
Default	0.046	0.046	1.003
Unemployment	19.992	22.765	0.878 *
Inflation	2.362	2.658	0.889 *
Devaluation	3.528	5.464	0.646 *

4. Conclusions

We have constructed a system whose purpose is to create indices representing various theories which are believed to drive heterogeneity in economic outcomes. When constructing an index, interpretability is an important feature to retain. This is primarily because through interpretability additional proxies can be found when deficiencies become apparent, and specific results can be explained directly. Our BTA approach then forms a natural means of incorporating and resolving the obvious model uncertainty present in such a specification. Furthermore, our focus on modeling multiple outcomes coupled with the ability to entertain a broad set of outcome sampling distributions lends our system both generalizability and flexibility.

There is considerable additional work to be done, both on the technical, algorithmic sides of BTA and also related to the specific goal of modeling an economy's potential for collapse. One key point has been the assumption that the multiple outcome variables are conditionally independent from one another given the indices. In practice, this did not seem to be overly critical, as seen by the fact that the inflation outcome was not present in the posterior when BTA was run on default using this feature alongside the others shown above. However, incorporating outcome variable dependence should be relatively straightforward using the Gaussian copula approach of Hoff (2007). Indeed uncertainty over

these conditional independence assumptions could also be model averaged using the copula Gaussian graphical model approach of Dobra and Lenkoski (2011).

Another matter that was avoided was country and year effects. Initial investigations using country-level fixed effects suggested little residual country-level correlation once other features were accounted for. Furthermore, since our goal is ultimately the use of indices for forecasting, it is desirable that latent factors such as random effects (which would not be internally estimated for countries or years not in the dataset) do not need to be supplied when forecasting. It our view that evidence of result clustering along year or country lines is primarily an indication of feature inadequacy. As we continue to build the SRI dataset we will monitor for clustering in results that are not captured in the feature set and use these to continue building out our collected features.

In this current system, outcome equations had a linear dependence on theory indices. While it will always be necessary to orientate the indices for reasons of identification (i.e., the assumption that $\gamma_{rt} = 1$ for at least one non-zero r), expanded linear forms such as spline models (Wood (2017)) are entirely feasible. Indeed a third layer of model selection would be to test between linearity and the expanded linearity offered by spline modeling.

The MCMC algorithm necessary to resolve BTA was neither trivial nor the most complex. As outlined as early as Rue (2001), block updates of parameters in hierarchical generalized models is often advantageous. We have in general avoided block updates at present, but such a sampling regime could speed up convergence and also algorithm run-time.

One difficulty we experience when implementing the quantile regression was the null second derivative in the asymmetric Laplace distribution. This in turn, makes intelligent updates of parameters for this distribution somewhat harder, since there is less information regarding posterior curvature and thus proposals have a tendency to move too far along the posterior density surface. This feature has already been investigated in some detail in related contexts. One potential for improved mixing would be to follow Fasiolo et al. (2017), who propose a smooth version of the pinball loss to aid the fitting of qgam models.

Finally, our reversible jump proposals were in some sense the least inspired part of the current system. Though mixing appeared acceptable, more focused jumps could have been constructed, by following much the same Laplacian formulations as the other model parameters.

With the onset of a global pandemic in the form of the COVID-19 virus, the great expansionary period following the global financial crisis appears to have finally halted. It is clear that we can expect to enter a retractionary phase of the global business cycle. Our applied interest has been to begin building a monitoring, forecasting and inferential toolset that can prepare us for this period. While we believe the current version of the SRI estimation system is encouraging, considerable work remains to be done.

Fresh data will be paramount to this effort. We intend to continue building this system to include all available years. We are broadly happy with the proxies collected to model insolvency and illiquidity in an economy. Macroeconomic and Systemic features could likely be expanded in a number of obvious ways. For instance including information on global financial markets or personal or industrial bankruptcy information could expand the Systemic theory proxies.

However, we are convinced that the Political risk proxies can be expanded in several important manners. Aspects related to political regimes are likely to affect potential for economic collapse. Merging our data with the regime change dataset of Reich (2002) could be one avenue to account for the effect of differing regimes and overall regime uncertainty.

Finally, it has been our hope to use only publicly available data sources to aid in the reproducability of our index construction. While we are convinced that devaluation matters should be included in our set of outcome equations, the necessary currency data has been hard to find publicly. We will continue to investigate open and public sources of currency exchange data to increase the coverage of this variable. In doing so, we hope the relative inconclusivity related to theories and their effect on sudden devaluations can be resolved.

Appendix A. Full Algorithm Details

Based on data $\mathcal{D}$ we use MCMC to obtain a sample $\{\varsigma^{[1]}, \ldots, \varsigma^{[S]}\}$ of the posterior distribution, where each $\varsigma^{[i]}$ contains

- $M_1, \ldots, M_T$ the models associated with theories 1 through T
- $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_T$, the coefficient vectors associated with each theory. Note that by construction $\beta_{jt} = 0$ when $j \notin M_t$
- $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R$ the theory-scaling vectors for each outcome equation r. A γ_{tr} can be set to zero, indicating that theory-t is not currently relevant for outcome equation r. For purposes of identification if multiple γ_{tr} are non-zero for a given t, we set $\gamma_{tr} = 1$ for whichever r is smallest.
- $\boldsymbol{I}_1, \ldots, \boldsymbol{I}_T$ the latent theory index vectors (each of length n) where I_{it} is the current state of the theory t index for observation i. By convention if $\gamma_{tr} = 0$ for all r then $I_{it} = 0$ for all i.
- $\nu_1, \ldots, \nu_T$, the random effect precision terms
- Global parameters $\boldsymbol{\alpha}_r$ in the R outcome equations

When moving from $\varsigma^{[s]}$ to $\varsigma^{[s+1]}$ we utilize four different MCMC strategies, all of which are now relatively standard in the MCMC literature. These are

- Gibbs sampling, relevant for updating $\boldsymbol{\beta}_t$ and ν_t
- Conditional Bayes Factors, which are used to update the theory-level models M_t
- Metropolis-Hastings via Laplacian calculations of the log posterior density which are used, in turn, to update theory indices $\boldsymbol{I}_t$, global parameters $\boldsymbol{\alpha}_r$ and those theory-scaling parameters γ_{tr} which are neither constrained to zero or one.
- Reversible Jump Methods for alternating γ_{tr} between being 0 or in $\mathbb{R}$. Note that the moves here become especially detailed–though primarily in the sense of bookkeeping–when γ_{tr} is currently set to 1, or if γ_{tr} is currently zero and r is smaller than all other non-zero $\gamma_{tr'}$. Finally, this becomes a joint reversible jump move when the model move will either turn-on or shut-off the theory entirely, as both γ_{tr} and $\boldsymbol{I}_r$ will be affected.

The sections below detail each of these approaches individually.

Appendix A.1. Gibbs Sampling Updates

To resample $\boldsymbol{\beta}_t$ we note that its posterior distribution

$$pr(\boldsymbol{\beta}_t | \cdot) \propto pr(\boldsymbol{\beta}_t | M_t, \boldsymbol{I}_t, \boldsymbol{X}_t, \nu_t)$$

where $\boldsymbol{\beta}_{M_t}$ and $\boldsymbol{X}_{M_t}$ indicate the restriction to those elements and columns of $\boldsymbol{\beta}_t$ and $\boldsymbol{X}_t$, respectively, associated with the variables in model M_t. We then have that

$$pr(\boldsymbol{\beta}_{M_t} | \boldsymbol{I}_t, \boldsymbol{X}_{M_t}, \nu_t) \propto pr(\boldsymbol{I}_t | \boldsymbol{\beta}_{M_t}, \boldsymbol{X}_{M_t}, \nu_t) pr(\boldsymbol{\beta}_{M_t} | \nu_t).$$

Via standard results this yields

$$pr(\boldsymbol{\beta}_{M_t}|\boldsymbol{I}_t, \boldsymbol{X}_{M_t}, \nu_t) = \mathcal{N}(\hat{\boldsymbol{\beta}}_{M_t}, \boldsymbol{\Xi}_{M_t}^{-1})$$

$$\boldsymbol{\Xi}_{M_t} = \nu_t \frac{g+1}{g} \boldsymbol{X}'_{M_t} \boldsymbol{X}_{M_t}$$

$$\hat{\boldsymbol{\beta}}_{M_t} = \boldsymbol{\Xi}_{M_t}^{-1} \boldsymbol{X}'_{M_t} \boldsymbol{I}_t.$$

Given $\boldsymbol{I}_t, \boldsymbol{\beta}_t$ we have that

$$\nu_t \sim \Gamma\left(\frac{a_t + n}{2}, \frac{b_t + S_t}{2}\right)$$

$$S_t = (\boldsymbol{I}_t - \boldsymbol{X}_t \boldsymbol{\beta}_t)'(\boldsymbol{I}_t - \boldsymbol{X}_t \boldsymbol{\beta}_t)$$

Appendix A.2. Conditional Bayes Factors to Update M_t

Conditional Bayes Factors compare integrated likelihoods for models M_t and a new proposal model M_t', conditioning on the latent indices $\boldsymbol{I}_t$. This conditioning then separates the Gaussian regression components on which the models operate from the larger non-Gaussian components in the response equations, leading to an efficient sampling regime. This efficiency is present both in the availability of closed form calculations to compare models and the relative parsimony of the approach's exposition.

In particular, note that

$$pr(M_t|\mathcal{D}, \cdot) \propto pr(\boldsymbol{I}_t|M_t)pr(M_t)$$

where it is implicit that we have conditioned on the fixed regressors. This implies that the latent theory indices $\boldsymbol{I}_t$ separate the conditional posterior of the model M_t from the data $\mathcal{D}$ and the associated non-integrable likelihoods. This term can the be represented by

$$pr(\boldsymbol{I}_t|M_t)pr(M_t) = \int_{\boldsymbol{\beta}_{M_r}} pr(\boldsymbol{I}_t|\boldsymbol{\beta}_{M_r})pr(\boldsymbol{\beta}_{M_r}|M_r)d\boldsymbol{\beta}_{M_r} pr(M_t)$$

The integrand above is then

$$\int_{\boldsymbol{\beta}_{M_r}} pr(\boldsymbol{I}_t|\boldsymbol{\beta}_{M_r})pr(\boldsymbol{\beta}_{M_r}|M_r)d\boldsymbol{\beta}_{M_r} \propto |\boldsymbol{\Xi}_{M_t}|^{1/2} \exp\left(\frac{1}{2}\hat{\boldsymbol{\beta}}'_{M_t}\boldsymbol{\Xi}_{M_t}\hat{\boldsymbol{\beta}}_{M_t}\right)$$

where $\hat{\boldsymbol{\beta}}_{M_t}$ and $\boldsymbol{\Xi}_{M_t}$ are defined as above. Similar to the classic MC3 algorithm, models M_t and M_t' are compared via Metropolis-Hastings.

Appendix A.3. Metropolis-Hastings Updates via Laplacian Expansions

The two sections above dealt with parameters that could effectively be "conditioned" away from the sampling model of the dependent variables, in both cases by conditioning on the latent variables $\boldsymbol{I}_t$. This, in turn, led to updates that were straightforward to calculate as in both cases they relied on well-known results for integrals over the Gaussian distribution. However, when conditional posterior distributions do not have a form amenable to integration or Gibbs sampling, Metropolis-Hastings algorithms provide an obvious alternative. This section therefore details all proposal distributions and acceptance ratios necessary to update these parameters.

In all cases, we follow a standard approach to creating Gaussian proposals which requires no pre-specified tuning parameters and instead adapts proposals to the local curvature of the log posterior density, see for example, chp. 4 of Rue and Held (2005) for a detailed discussion of this approach

and Dyrrdal et al. (2015) for a similar algorithmic design. More involved methods, such as Hamiltonian MCMC, Manifold MCMC and so forth, which build on these concepts could have been entertained but mixing was already sufficiently acceptable that these more sophisticated methodologies seemed unnecessary. See our discussion the Section 4. Suppose, in general, that we would like to update a parameter τ and write $\log pr(\tau|\cdot) = f(\tau)$ to represent the log posterior density of this parameter with respect to the observations and all other parameters. For designing the proposal distribution, we employ a Gaussian approximation of this posterior density. A quadratic Taylor expansion of the log-posterior $f(\tau)$ around the value τ gives

$$f(\tau') \approx f(\tau) + f'(\tau)(\tau' - \tau) + \frac{1}{2}f''(\tau)(\tau' - \tau)^2$$
$$= a + b\tau' - \frac{1}{2}c(\tau')^2,$$

where $b = f'(\tau) - f''(\tau)\tau$ and $c = -f''(\tau)$. The posterior distribution $pr(\tau|\cdot)$ can therefore be approximated by

$$\widetilde{pr}(\tau|\cdot) \propto \exp\left(-\frac{1}{2}c(\tau')^2 + b\tau'\right),$$

the density of the Gaussian distribution $\mathcal{N}(b/c, c^{-1})$. Using this relationship, we choose $\mathcal{N}(b/c, c^{-1})$ as our proposal distribution, where τ is the current state in the MCMC chain. This formulation alleviates the user from specifying a large number of sampling tuning parameters and achieves high acceptance proportions.

The following subsections outline the specific forms of f, f', and f'' for all variates that are updated in this manner. Since the I_{it} depend on all r equations they are handled in a final, separate subsection.

Appendix A.3.1. Logistic Regression

If equation r is a logistic model then it has the form

$$pr(Y_{ir}|\cdot) = \left(\frac{\exp(\mu_{ir})}{1 + \exp(\mu_{ir})}\right)^{Y_{ir}} \left(\frac{1}{1 + \exp(\mu_{ir})}\right)^{1-Y_{ir}}$$

where

$$\mu_{ir} = \alpha_r + \sum_{t=1}^{T} \gamma_{rt} I_{it}$$

The formulas for α_r, γ_{rt} require derivation (as noted above we leave I_{it} to a final subsection). First, note

$$\log pr(Y_{ir}) = Y_{ir}\mu_{ir} - \log(1 + \exp(\mu_{ir}))$$

Then for the global parameter α_r with prior distribution $\alpha_r \sim \mathcal{N}(0,1)$ we have that

$$f(\alpha_r) = \sum_{i=1}^{n} \{Y_{ir}\mu_{ir} - \log(1 + \exp(\mu_{ir}))\} - \frac{\alpha_r^2}{2}$$
$$f'(\alpha_r) = \sum_{i=1}^{n} \left\{Y_{ir} - \frac{\exp(\mu_{ir})}{(1 + \exp(\mu_{ir}))}\right\} - \alpha_r$$
$$f''(\alpha_r) = -\sum_{i=1}^{n} \left\{\frac{\mu_{ir}}{(1 + \mu_{ir})^2}\right\} - 1$$

Similarly, for γ_{rt} not constrained to be 0 or 1 we assume $\gamma_{rt} \sim \mathcal{N}(0,1)$ and have

$$f(\gamma_{rt}) = \sum_{i=1}^{n} \{Y_{ir}\mu_{ir} - \log(1 + \exp(\mu_{ir}))\} - \frac{\gamma_{rt}^2}{2}$$

$$f'(\gamma_{rt}) = \sum_{i=1}^{n} \left\{ Y_{ir}I_{it} - I_{it}\frac{\exp(\mu_{ir})}{(1 + \exp(\mu_{ir}))} \right\} - I_{it}$$

$$f''(\gamma_{rt}) = -\sum_{i=1}^{n} I_{it}^2 \left\{ \frac{\exp(\mu_{ir})}{(1 + \exp(\mu_{ir}))^2} \right\} - 1$$

Finally, as it will be important in derivations for the updates of I_{it} we write

$$l_r(Y_{ir}, I_{it}) = Y_{ir}\mu_{ir} - \log(1 + \exp(\mu_{ir}))$$

$$\dot{l}_r(Y_{ir}, I_{it}) = \gamma_{ir}Y_{ir} - \gamma_{ir}\frac{\exp(\mu_{ir})}{1 + \exp(\mu_{ir})}$$

$$\ddot{l}_r(Y_{ir}, I_{it}) = -\gamma_{ir}^2 \frac{\exp(\mu_{ir})}{1 + \exp(\mu_{ir})^2}$$

Appendix A.3.2. Bayesian Quantile Regression

Let

$$pr(Y_{ir}|\mu_{ir}, \kappa, q) \propto \exp\left\{\kappa - e^\kappa \rho_q(Y_{ir} - \mu_{ir})\right\}$$

$$\rho_q(Y_{ir} - \mu_{ir}) = (Y_{ir} - \mu_{ir})(q - \mathbf{1}\{Y_{ir} < \mu_{ir}\})$$

be a Bayesian Quantile Regression, that is, Y_{ir} is considered asymmetric Laplace distributed with log-precision parameter κ and

$$\mu_{ir} = \alpha_{ir} + \sum_{t=1}^{T} \gamma_{rt}I_{it}.$$

We therefore need to derive the relevant formulas for α_{ir}, γ_{rt} and likelihood derivatives for I_{it}. We note

$$\log pr(Y_{ir}|\mu_{ir}, \kappa, q) = \kappa - e^\kappa \rho_q(Y_{ir} - \mu_{ir})$$

and thus,

$$\frac{\partial \log pr(Y_i|\cdot)}{\partial \mu_i} = e^\kappa(q - \mathbf{1}\{Y_i < \mu_i\})$$

$$\frac{\partial^2 \log pr(Y_i|\cdot)}{(\partial \mu_i)^2} = 0.$$

Therefore, for α_r with $\mathcal{N}(0,1)$ prior we have

$$f(\alpha_r) = \kappa - e^\kappa \sum_{i=1}^n \rho_q(Y_{ir} - \mu_{ir}) - \frac{\alpha_r^2}{2}$$

$$f'(\alpha_r) = e^\kappa \sum_{i=1}^n (q - \mathbf{1}\{Y_i < \mu_i\}) - \alpha_r$$

$$f''(\alpha_r) = -1.$$

Similarly when γ_{rt} is not constrained to 0 or 1 we set $\gamma_{rt} \sim \mathcal{N}(0,1)$ and have

$$f(\gamma_{rt}) = \sum_{i=1}^n \left\{ \kappa - e^\kappa \rho_q(Y_{ir} - \mu_{ir}) \right\} - \frac{\gamma_r^2}{2}$$

$$f'(\gamma_{rt}) = \sum_{i=1}^n \left\{ I_{it} e^\kappa (q - \mathbf{1}\{Y_i < \mu_i\}) \right\} - \gamma_r$$

$$f''(\gamma_{rt}) = -1.$$

Likewise, we note that

$$\frac{\partial \log pr(Y_i|\cdot)}{\partial \kappa} = 1 - e^\kappa (q - \mathbf{1}\{Y_i < \mu_i\})$$

$$\frac{\partial^2 \log pr(Y_i|\cdot)}{(\partial \kappa)^2} = -e^\kappa (q - \mathbf{1}\{Y_i < \mu_i\}).$$

and thus if $\kappa \sim \mathcal{N}(0,1)$ in the prior, then

$$f(\kappa|\cdot) = n\kappa - e^\kappa \sum_{i=1}^n \rho_q(Y_i - \mu_i) - \frac{1}{2}\kappa^2$$

$$f'(\kappa|\cdot) = n - e^\kappa \sum_{i=1}^n \rho_q(Y_i - \mu_i) - \kappa$$

$$f''(\kappa|\cdot) = -e^\kappa \sum_{i=1}^n \rho_q(Y_i - \mu_i) - 1.$$

Finally, for I_{it} we have

$$l(Y_{ir}, I_{it}) = \kappa - e^\kappa \rho_q(Y_{ir} - \mu_{ir})$$
$$\dot{l}(Y_{ir}, I_{it}) = \gamma_{rt} e^\kappa (q - \mathbf{1}\{Y_i < \mu_i\})$$
$$\ddot{l}(Y_{ir}, I_{it}) = 0$$

Appendix A.3.3. GEV Regression

When Y_{ir} has the form of a GEV Regression with global log-precision κ and shape ξ we have

$$pr(Y_{ir}|\mu_{ir}, \kappa, \xi) = e^\kappa h(Y_{ir})^{-(\xi+1)/\xi} \exp\left(- h(Y_{ir})^{-\xi^{-1}} \right)$$

$$h(Y_{ir}) = 1 + \xi e^\kappa (Y_{ir} - \mu_{ir})$$

$$\mu_{ir} = \alpha_r + \sum_{t=1}^T \gamma_{rt} I_{it}.$$

with the additional restriction that $h(\cdot) > 0$. Calculations for this density have a tendency to become somewhat involved. We first note

$$a(Y_{ir}) \equiv \log pr(Y_{ir}|\mu_{ir}, \kappa, \xi) = \kappa - \frac{\xi+1}{\xi} \log h(Y_{ir}) - h(Y_{ir})^{-\xi^{-1}}$$

Since $\partial h(Y_{ir})/\partial\mu_{ir} = -e^{\kappa}\xi$ we have that

$$\dot{a}(Y_{ir}) \equiv \frac{\partial}{\partial\mu_{ir}} \log pr(Y_{ir}|\cdot) = (\xi+1)e^{\kappa}h(Y_{ir})^{-1} - e^{\kappa}h(Y_{ir})^{-\xi^{-1}-1}$$

$$\ddot{a}(Y_{ir}) \equiv \frac{\partial^2}{(\partial\mu_{ir})^2} \log pr(Y_{ir}|\cdot) = \xi(\xi+1)e^{2\kappa}h(Y_{ir})^{-2} - (\xi+1)e^{2\kappa}h(Y_{ir})^{-\xi^{-1}-2}.$$

Therefore, to update $\alpha_r \sim \mathcal{N}(0,1)$ we have

$$f(\alpha_r) = \sum_{i=1}^{n} a(Y_{ir}) - \frac{\alpha_r^2}{2}$$

$$f'(\alpha_r) = \sum_{i=1}^{n} \dot{a}(Y_{ir}) - \alpha$$

$$f''(\alpha_r) = \sum_{i=1}^{n} \ddot{a}(Y_{ir}) - 1.$$

Likewise, to update any γ_{rt} not constrained to 0 or 1 we have

$$f(\gamma_{rt}) = \sum_{i=1}^{n} a(Y_{ir}) - \frac{\gamma_{rt}^2}{2}$$

$$f'(\gamma_{rt}) = \sum_{i=1}^{n} \dot{a}(Y_{ir})I_{rt} - \gamma_{rt}$$

$$f''(\gamma_{rt}) = \sum_{i=1}^{n} \ddot{a}(Y_{ir})I_{rt}^2 - 1.$$

For the term I_{rt} we note

$$l(Y_{ir}, I_{it}) = a(Y_{ir})$$
$$\dot{l}(Y_{ir}, I_{it}) = \dot{a}(Y_{ir})\gamma_{rt}$$
$$\ddot{l}(Y_{ir}, I_{it}) = \ddot{a}(Y_{ir})\gamma_{rt}^2.$$

Now focus on the global log precision term $\kappa \sim \mathcal{N}(0,1)$ we have

$$f(\kappa) = \sum_{i=1}^{n} a(Y_{ir}) - \frac{\kappa^2}{2}$$

$$f'(\kappa) = \sum_{i=1}^{n} \{1 - e^{\kappa}(\xi+1)(Y_{ir} - \mu_{ir}) + b_1(Y_{ir})\} - \kappa$$

$$f''(\kappa) = \sum_{i=1}^{n} \{-e^{\kappa}(\xi+1)(Y_{ir} - \mu_{ir}) + b_1(Y_{ir}) - b_2(Y_{ir})\} - 1$$

where

$$b_1 = e^{\kappa}(Y_{ir} - \mu_{ir})h(Y_{ir})^{-\xi^{-1}-1}$$
$$b_2 = (\xi + 1)e^{2\kappa}(Y_{ir} - \mu_{ir})^2 h(Y_{ir})^{-\xi^{-1}-2}.$$

The calculations for the shape parameter ξ are somewhat more involved. Let

$$g_1(Y_{ir}) = \frac{\xi + 1}{\xi}\log h(Y_{ir})$$
$$g_2(Y_{ir}) = \exp\left\{-(\xi^{-1} + 1)\log h(Y_{ir})\right\}$$

We then obtain

$$\dot{g}_1(Y_{ir}) = \frac{\partial g_1(Y_{ir})}{\partial \xi} = -\frac{\log h(y_{ts})}{\xi^2} + \frac{\xi + 1}{\xi}h(Y_{ir})^{-1}e^{\kappa}(Y_{ir} - \mu_{ir})$$
$$\dot{g}_2(Y_{ir}) = \frac{\partial g_2(Y_{ir})}{\partial \xi} = g_2\left[\frac{\log h(Y_{ir})}{\xi^2} - (\xi^{-1} + 1)h(Y_{ir})^{-1}e^{\kappa}(Y_{ir} - \mu_{ir})\right],$$

from which it follows that

$$\frac{\partial}{\partial \xi}\log pr(Y_{ir}|\cdot) = -\dot{g}_1 - \dot{g}_2.$$

For the second derivative, similar calculations return

$$\frac{\partial^2}{(\partial \xi)^2}\log pr(Y_{ir}|\cdot) = \frac{\partial}{\partial \xi}\left(-\dot{g}_1 - \dot{g}_2\right) = d_1 + d_2 - d_3 + d_4,$$

where

$$d_1 = -2\xi^{-3}\log h(Y_{ir}) + \xi^{-2}h(Y_{ir})^{-1}e^{\kappa}(Y_{ir} - \mu_{ir})$$
$$d_2 = \frac{h(Y_{ir})^{-1}(Y_{ir} - \mu_{ir})e^{\kappa}}{\xi^2} + \frac{\xi + 1}{\xi}h^{-2}(Y_{ir})(Y_{ir} - \mu_{ir})^2 e^{2\kappa}$$
$$d_3 = \dot{g}_2(Y_{ir})\left[\frac{\log h(Y_{ir})}{\xi^2}\right] + g_2(Y_i)\left[-\frac{2\log h(Y_{ir})}{\xi^3} + \frac{h(Y_{ir})^{-1}e^{\kappa}(Y_{ir} - \mu_{ir})}{\xi^2}\right]$$
$$d_4 = \dot{g}_2(Y_{ir})\left[\frac{h(Y_{ir})^{-1}e^{\kappa}(Y_{ir} - \mu_{ir})}{\xi}\right] - g_2(Y_i)(Y_{ir} - \mu_{ir})e^{\kappa}\left[\frac{h(Y_{ir})^{-1}}{\xi^2} + \frac{h(Y_{ir})^{-2}(Y_{ir} - \mu_{ir})e^{\kappa}}{\xi}\right].$$

Hence, for updating $\xi \sim \mathcal{N}(0,1)$ we have

$$f(\xi|\cdot) = \sum_{i=1}^{n}\left\{\kappa - g_1(Y_{ir}) - g_2(Y_{ir})\right\} - \frac{\xi^2}{2}$$
$$f'(\xi|\cdot) = \sum_{i=1}^{n}\left\{-\dot{g}_1(Y_{ir}) - \dot{g}_2(Y_{ir})\right\} - \xi$$
$$f''(\xi|\cdot) = \sum_{i=1}^{n}\left\{d_1 + d_2 - d_3 + d_4\right\} - 1.$$

Appendix A.3.4. Updating Theory Indices

We now consider updating of theory indices I_{it}. Noting that

$$I_{it} = \mathbf{X}'_{rt}\boldsymbol{\beta}_t + \epsilon_{it}, \quad \epsilon \sim \mathcal{N}(0, v_t^{-1})$$

We have the formulas

$$f(I_{it}|\cdot) = \sum_{r=1}^{R} l_r(I_{it}|\cdot) - \frac{v_t}{2}(I_{it} - \mathbf{X}'_{rt}\boldsymbol{\beta}_t)^2$$

$$f'(I_{it}|\cdot) = \sum_{r=1}^{R} \dot{l}_r(I_{it}|\cdot) - v_t(I_{it} - \mathbf{X}'_{rt}\boldsymbol{\beta}_t)$$

$$f''(I_{it}|\cdot) = \sum_{r=1}^{R} \ddot{l}_r(I_{it}|\cdot) - v_t$$

Were the $l_r, \dot{l}_r$ and $\ddot{l}_r$ terms are those discussed in the sections above for each respective outcome equation r in the system.

Appendix A.4. Updating Theory Inclusion Parameters via Reversible Jump

Suppose now that $\gamma_{rt} = 0$ in the current state of the chain. In the relatively straightforward case in which there is an $r' < r$ for which $\gamma_{rt} = 1$ – and thus the inclusion of the γ_{rt} will not affect identification matters–we may attempt to make γ_{rt} non zero by proposing $\gamma'_{rt} \sim \mathcal{N}(0,1)$. We thus transition from (γ_r, γ_{rt}), where $\gamma_{rt} = 0$ to γ'_t with $(\gamma'_t)_r = \gamma_{rt}$ and $(\gamma'_t)_s = (\gamma_t)_s$ for all other $s \neq r$, a transformation with Jacobian 1. Letting

$$\mu_{ir} = \alpha_r + \sum_{t=1}^{T} \gamma_{rt} I_{it}$$

and

$$\mu'_{ir} = \alpha_r + \sum_{t=1}^{T} \gamma'_{rt} I_{it}$$

Since our prior sets all $\gamma_{rs} \sim \mathcal{N}(0,1)$, the auxiliary density cancels with the larger prior we thus have that

$$\log pr(\gamma_r, \gamma'_{rt}|\cdot) \propto \sum_{i=1}^{n} l_r(Y_{it}|\mu_{it})$$

$$\log pr(\gamma'_r|\cdot) \propto \sum_{i=1}^{n} l_r(Y_{it}|\mu'_{it})$$

where l_r is the associated log-likelihood for equation r. This gives the necessary log densities for comparing $\gamma_{rt} \in \mathbb{R}$ and $\gamma_{rt} = 0$. See our discussion in the Conclusions section regarding more focused proposals of γ_{rt} which could aid in mixing and would also make the expressions above slightly more involved.

When $\gamma_{rt} = 0$ and $\gamma_{st} = 1$ for $s > r$, some bookkeeping is necessary to adjust the system. In particular, we sample $u \sim \mathcal{N}(0,1)$. We then create a new vector γ_t where

$$\gamma_{st} = \begin{cases} 1, & \text{if } s = r \\ u\gamma'_{st}, & \text{otherwise} \end{cases}$$

And similarly we move from I_{it} to I'_{it} by setting $I'_{it} = I_{it}/u$, $\boldsymbol{\beta}'_t = \boldsymbol{\beta}/u$, $g'_t = g_t/u$ and $v'_t = uv_t$, $b'_t = b_t/u$. We therefore note that while we have changed all γ_{st} values and the associated theory indices I_{it}, only the likelihood for the dependent variable r is affected and comparisons can then be performed as discussed above.

References

Brock, William A., Steven N. Durlauf, and Kenneth D. West. 2003. *Policy Evaluation in Uncertain Economic Environments*. Technical Report. Cambridge: National Bureau of Economic Research.

Chen, Ray-Bing, Yi-Chi Chen, Chi-Hsiang Chu, and Kuo-Jung Lee. 2017. On the determinants of the 2008 financial crisis: A Bayesian approach to the selection of groups and variables. *Studies in Nonlinear Dynamics & Econometrics* 21: 1–17.

Dobra, Adrian, and Alex Lenkoski. 2011. Copula Gaussian graphical models and their application to modeling functional disability data. *The Annals of Applied Statistics* 5: 969–93. [CrossRef]

Dyrrdal, Anita Verpe, Alex Lenkoski, Thordis L. Thorarinsdottir, and Frode Stordal. 2015. Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics* 26: 89–106.

Fasiolo, Matteo, Yannig Goude, Raphael Nedellec, and Simon N. Wood. 2017. Fast calibrated additive quantile regression. *arXiv* arXiv:1707.03307.

Gamerman, Dani, and Hedibert F. Lopes. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall/CRC.

Green, Peter J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–32. [CrossRef]

Hastie, Trevor, and Robert Tibshirani. 1990. *Generalized Additive Models*. London: Chapman and Hall/CRC.

Hoff, Peter D. 2007. Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics* 1: 265–83. [CrossRef]

Karl, Anna, and Alex Lenkoski. 2012. Instrumental variable Bayesian model averaging via conditional Bayes factors. *arXiv* arXiv:1202.5846.

Kourtellos, Andros, Alex Lenkoski, and Kyriakos Petrou. 2019. Measuring the strength of the theories of government size. In *Empirical Economics*. Berlin: Springer, pp. 1–38.

Lenkoski, Alex. 2013. A direct sampler for G-Wishart variates. *Stat* 2: 119–28. [CrossRef]

Ley, Eduardo, and Mark F.J. Steel. 2009. On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics* 24: 651–74. [CrossRef]

Möller, Annette, Alex Lenkoski, and Thordis L. Thorarinsdottir. 2013. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society* 139: 982–91. [CrossRef]

Reich, Gary. 2002. Categorizing political regimes: New data for old problems. *Democratization* 9: 1–24. [CrossRef]

Robert, Christian, and George Casella. 2013. *Monte Carlo Statistical Methods*. Berlin: Springer Science & Business Media.

Roubini, Nouriel, and Paolo Manasse. 2005. *"Rules of Thumb" for Sovereign Debt Crises*. Working paper no. 05/42. Washington: IMF.

Rue, Håvard. 2001. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63: 325–38. [CrossRef]

Rue, Havard, and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall/CRC.

Savona, Roberto, and Marika Vezzoli. 2015. Fitting and forecasting sovereign defaults using multiple risk signals. *Oxford Bulletin of Economics and Statistics* 77: 66–92. [CrossRef]

Steel, Mark F.J. 2019. Model averaging and its use in economics. *arXiv* arXiv:1709.0822.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R.* London: Chapman and Hall/CRC.

Zellner, Arnold. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American statistical Association* 57: 348–68. [CrossRef]