A Note on Sigma-Mu Efficiency Analysis as a Methodology for Evaluating Units through Composite Indicators

Mike G. Tsionas*

Abstract

Recent research introduces a methodology for constructing composite indicators, called "$\sigma - \mu$ efficiency analysis", illustrating its potential in a case study of world happiness. Building on the landmark research paper, we propose a novel model that allows statistical inference for both weights in the composite indicator as well as inefficiency, fully accounting for outliers in the data and unit-specific heterogeneity in weights. The new techniques are based on Bayesian analysis via Markov Chain Monte Carlo.

Key Words: Decision Processes; Composite Indicators; Sigma-Mu efficiency; Stochastic Multi-attribute Acceptability Analysis; Bayesian Analysis.

*Lancaster University Management School, LA1 4YX, United Kingdom, m.tsionas@lancaster.ac.uk

1 Introduction

Greco, Ishizaka, Tasiou, and Torrisi (2019a) introduce a methodology for constructing composite indicators that they call "$\sigma - \mu$ efficiency analysis", illustrating its potential in a case study of world happiness. The paper is also a fun read as it involves considerations from many different scientific fields. Many international organizations e.g. the OECD, UN, World Bank etc.) use various indicators which are based on equal weights for the underlying variables. A fundamental step in the construction of composite indicators regards the weighting of the elementary indicators. Quite often, the indicators are based on the arithmetic mean (e.g. see, among others, the Index of Economic Freedom (Miller et al., 2018) and the Inclusive Development Index (Samans et al., 2018), or the geometric mean (the 2010 Human Development Index; see UNDP, 2010). As a matter of fact, "by taking into account the whole set of admissible weight vectors, one can consider the whole spectrum of preferences of individuals, as well as multiple selves within each individual interested in the composite indicator" (Greco et al., 2019a. p. 943). In recent research the problem of equal weights which is associated with another major problem, viz. the problem of "representative agent", is mitigated by use of multi-criteria decision analysis and Stochastic Multi-Attribute Acceptability Analysis (SMAA). SMAA involves a decision-maker that is unable to provide the parameters required for the evaluation process (see e.g. Doumpos et al., 2016, 2017, as well as Greco et al., 2008, 2010, 2016, 2019b). As Greco et al (2019a) write: "More specifically, by considering a probability distribution on the set of feasible weight vectors, SMAA reveals the probability that a unit attains a given ranking position, as well as the probability that a given unit is better than another" (Greco et al., 2019, p. 943).

The methodology in Greco et al. (2019a) is, in fact, a novel SMAA approach, where instead of computing the probability that each considered alternative could obtain a given rank position and the probability of being preferred to another alternative; one could compute the mean μ and the standard deviation of the values assigned by the weighted sum, σ. In turn, μ and σ can be employed to obtain a single overall evaluation using the overall (global) efficiency measure that is proposed by Greco et al (2019a, pp. 942–943). In fact, this represents a new method in the SMAA family.

SMAA takes weight uncertainty into account by considering a probability density $f(w)$ over the space of all weight vectors, $w \in W$, where W is defined below.

Suppose we have a vector of variables $x_i \in \Re^J$ $(i = 1, ..., n)$. A composite indicator is defined as:

$$CI(x_i, w) = x_i' w = \sum_{j=1}^{J} x_{ij} w_j, i = 1, \ldots, n, \tag{1}$$

where $w \in W = \{(w_1, \ldots, w_J) : w_j \geq 0, j = 1, \ldots, J, \sum_{j=1}^{J} w_j = 1\}$, is a set of weights reflecting the importance of the different variables in the composite indicator. Given a distribution $f(w), w \in W$, we can define the mean and standard deviation of the composite indicator:

$$\begin{aligned} \mu_i &= {}_W\, x_i' w f(w) dw,\ i = 1, \ldots, n, \\ \sigma_i &= \sqrt{{}_W (x_i' w - \mu_i)^2 f(w) dw},\ i = 1, \ldots, n. \end{aligned} \tag{2}$$

The expressions in (2) are based, essentially, on specifying a uniform distribution of weights in set W, and, in turn, computing μ_i, σ_i by simulation when priority constraints like $w_1 \geq w_2 \geq \ldots \geq w_J$ are available. The question is why the density $f(w)$ should be uniform in W. Surely, this is acceptable if the density represents a "prior" but then a posterior of the form $p(w|\text{data})$ should have been derived which can be used to provide approximations to μ_i, σ_i in (2). An additional problem that Greco et al. (2019) dealt with is computation of inefficiency, viz. a measure of whether given σ_i, the mean μ_i is as high as possible. They deal with this problem in a second stage where they apply Data Envelopment Analysis (DEA), in particular m-order frontiers.

Finally, we feel obliged to quote from Greco et al. (2019) on the following serious issue: "We stand by the principle that a meaningful composite indicator should ideally reflect a multiplicity of viewpoints. Technically speaking, this can be achieved in the weighting stage, in which individuals that the indicator concerns can participate, by expressing their preferences on the importance of indicator dimensions. These individuals could constitute different clusters, e.g. experts, policymakers, or even citizens at whom policies are addressed. Therefore, the main driver of this concept refrains from the classic scheme of a single, allegedly representative weight vector in the construction of an indicator, by taking into account all these individuals' viewpoints. In the past, this has been feasible with the use of SMAA (see, e.g., Greco et al., 2018)" (Greco et al., 2019, p. 945).

In this paper, we propose a unifying model where weights and inefficiencies are derived simultaneously from the same problem, which is stated in the next section. In principle, uniform weights are not optimal and it is of interest to examine the implications of formal statistical inference for these weights. Of course, it would have been great if statistical inference provides support for uniform weights but this cannot be known in advance with obvious implications for rankings.

2 The model

Our new model is defined as follows:

$$CI_i = x_i' w + v_i - u_i, i = 1, \ldots, n, \tag{3}$$

$$x_i = f_i 1_J + \epsilon_i, i = 1, \ldots, n, \tag{4}$$

$$CI_i \text{ is "as close as possible" to } f_i. \tag{5}$$

where 1_J denotes a vector of ones in $\Re^J$, f_i represents a common factor, v_i is a two-sided error term representing noise, and u_i is a non-negative error term. According to equation (3), the composite indicator is a weighted average of elements of x_i, there is noise represented by v_i as well as inefficiency u_i relative to the frontier value of CI_i. The term v_i reflects usual measurement error, and u_i is a non-negative error term representing "inefficiency": For a given value of the regressors, x_i, and given measurement error v_i, u_i represents the amount by which the observed composite indicator CI_i falls short of its maximum possible value.

Equation (4) represents x_i as a factor model, where ϵ_i is a two-sided error term in $\Re^J$. Finally, (5) states that, for consistency purpose, the common factor and the composite indicator should be similar, to the extent possible. Without

(5), consistency (viz. $f_i = CI_i$) would require: $x_i' w = CI_i + \epsilon_i' w$ which implies the following restriction among the different error terms of the model:

$$\epsilon_i' w = -v_i + u_i \ \sim \ v_i + u_i, i = 1, \ldots, n, \tag{6}$$

where "$\sim$" denotes equality in distribution. In practice, it seems impossible to impose (6).

Equations (3), (4) and (5) are equivalent to the following problem. Conditional on $(f_i, i = 1, \ldots, n)$:

$$\begin{gathered} \min_{\{f_i, CI_i,\, i=1,\ldots,n\},w} \sum_{i=1}^n \left(CI_i - f_i\right)^2, \\ CI_i = x_i' w + v_i - u_i, i = 1, \ldots, n, \\ x_i = f_i 1_J + \epsilon_i, i = 1, \ldots, n, \\ w \in \Re_+^J,\ w' 1_J = 1, \end{gathered} \tag{7}$$

where 1_J is a $J \times 1$ vector of ones. This requires the interpretation of (5) in the sense that the L_2-norm of the difference between the composite indicator and the common factor should be as small as possible. Since the factors are not known, if we are willing to assume:[1]

$$f_i \sim iid\, \mathcal{N}(0, \sigma_f^2),\ i = 1, \ldots, n, \tag{8}$$

then the problem in (7) is equivalent to a weighted least squares problem of the following form:

$$\lambda_1 \sum_{i=1}^n (CI_i - f_i)^2 + \lambda_2 \sum_{i=1}^n (CI_i - x_i' w + u_i)^2 + \lambda_3 \sum_{i=1}^n u_i^2 + \lambda_4 \sum_{i=1}^n (x_i - f_i 1_J)^2 + \lambda_5 \sum_{i=1}^n f_i^2, \tag{9}$$

where $\lambda_1, \ldots, \lambda_5 > 0$ are weights (whose sum is not necessarily one). Since the weights are unknown, they have to be estimated as well. In fact (9) is in a form which is implied by the log likelihood / log posterior in (11) below. To derive the posterior we make the following distributional assumptions:

$$v_i \sim iid\, \mathcal{N}(0, \sigma_v^2),\ u_i \sim iid\, \mathcal{N}_+(0, \sigma_u^2), \epsilon_i \sim iid\, \mathcal{N}_J(\mathbf{0}, \Omega), i = 1, \ldots, n, \tag{10}$$

where $iid\, \mathcal{N}_+(0, \sigma_u^2)$ denotes the half-normal distribution. Moreover, $\mathbf{0}$ denotes a $J \times 1$ zero vector. All random variables in (10) are independent of each other.

Based on (10) and (8) we define the following augmented posterior distribution:[2]

$$\begin{gathered} p(CI, f, w, u, h, \sigma_v, \sigma_u, \sigma_f, \Omega | X) \propto h^{-n/2} \sigma_v^{-n/2} \sigma_u^{-n/2} \sigma_f^{-n/2} |\Omega|^{-n/2} \cdot \\ \exp\left\{ -\tfrac{1}{2h^2} \sum_{i=1}^n (CI_i - f_i)^2 - \tfrac{1}{2\sigma_v^2} \sum_{i=1}^n (CI_i - x_i' w + u_i)^2 - \tfrac{1}{2\sigma_u^2} \sum_{i=1}^n u_i^2 \right\} \cdot \\ \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^n (x_i - f_i 1_J)' \Omega^{-1} (x_i - f_i 1_J) - \tfrac{1}{2\sigma_f^2} \sum_{i=1}^n f_i^2 \right\} \cdot p(w) \cdot p(\sigma_v, \sigma_u, \Omega | w), \end{gathered} \tag{11}$$

where $CI = [CI_1, ..., CI_n]'$, $f = [f_1, ..., f_n]'$, $w = [w_1, ..., w_n]'$, $u = [u_1, ..., u_n]'$, $X = [x_1', ..., x_n']'$, and $p(w)$ is a prior for w

[1] An alternative assumption is to perform factor analysis using $(x_i, i = 1, ..., n)$ prior to solving this problem. Moreover, in equation (8) below, we assume that the variance of the factor is one, for identification purposes. In Bayesian analysis, this is not strictly necessary if a proper prior is used for the variance of the factor.

[2] "Augmented" in the sense that latent variables CI, f, u are present in this posterior instead of integrating them out.

defined over W. Specifically, following Greco et al. (2019b), we assume:

$$p(w) \propto \mathbb{I}_W(w),\ w_1 \geq w_2 \geq \ldots \geq w_J, \tag{12}$$

where $\mathbb{I}_W(w) = 1$, if $w \in W$, and zero otherwise. The inequality restrictions arise because we assume that different priorities can be given to different variables. Moreover, $p(\sigma_v, \sigma_u, \Omega)$ is a prior on the parameters $\sigma_v, \sigma_u, \Omega$. We use the reference prior:

$$p(\sigma_v, \sigma_u, \sigma_f, \Omega|w) \propto \sigma_v^{-1} e^{-(\overline{q}/2\sigma_v^2)} \sigma_u^{-1} \sigma_f^{-1} |\Omega|^{-(J+1)/2}, \tag{13}$$

where $\overline{q} = 10^{-6}$. This prior is flat or "non-informative" for the scale parameters σ_u, σ_f and the different elements of Ω.[3]

In particular, we are interested in the posterior:

$$p(w|X) \propto p(w|X)d\vartheta, \tag{14}$$

where $\vartheta = [CI, f, u, \sigma_v, \sigma_u, \Omega]$. Posterior means of weights can be obtained as:

$$\overline{w} = S^{-1} \sum_{s=1}^{S} w^{(s)}, \tag{15}$$

and posterior mean inefficiency as:

$$\overline{u}_i = S^{-1} \sum_{s=1}^{S} u_i^{(s)},\ i = 1, \ldots, n. \tag{16}$$

The posterior mean composite indicator can be estimated as

$$\overline{CI_i} = S^{-1} \sum_{s=1}^{S} CI_i^{(s)},\ i = 1, \ldots, n. \tag{17}$$

Posterior standard deviations, and in fact, posterior densities can be computed easily based on MCMC draws for these parameters of interest.

3 Outliers and other problems

Greco et al. (2019a) argue convincingly that outliers can certainly be a problem and they propose standardizing the data, as in Greco et al. (2018a, see online Appendix A.2). We use the same normalizations. This takes care of outliers, at least in a partial way. As they write they adopt m-order frontiers to account for outliers in a second stage: "In this respect, the robust m-order frontiers aid significantly in adjusting the estimators to account for these outliers, and given these noticeable differences (especially around the frontiers containing the outliers), we strongly encourage their use alongside our proposed approach". In this work we follow a different approach which is easy to implement in practice. Specifically,

[3] Although this prior is rather flat, we need $\overline{q} > 0$ to make sure that the posterior is finitely integrable.

instead of the normal distribution of ϵ_i in (10) we assume a **multivariate Student-t distribution**:

$$\epsilon_i \sim t_{J,\nu}(\mathbf{0}, \Omega),\ i = 1, \dots, n, \tag{18}$$

where $t_{J,\nu}(0, \Omega)$ denotes the J-variate Student-t distribution, whose location is a zero vector and the scale matrix is Ω (we use the same symbol as in (10) to economize on notation but the interpretation is, apparently, different). A well-known way to construct the distribution is as follows:

$$\varepsilon_i | U \sim \mathcal{N}_J(\mathbf{0},\ (U/\nu)^{-1/2}\Omega),\ U \sim \chi^2_\nu. \tag{19}$$

Therefore, a multivariate Student-t distribution can be constructed as a multivariate normal with zero location vector and scale matrix $(U/\nu)^{-1/2}\Omega$ where U follows the chi-square distribution with ν degrees of freedom.

In turn, this mixing process gives rise to the multivariate Student-t distribution with density:

$$p(\epsilon_i; \nu, \Omega) = \frac{\Gamma((\nu + J)/2)}{\Gamma(\nu/2)(\nu J)^{J/2}|\Sigma|^{1/2}} \left(1 + \tfrac{1}{\nu}\epsilon_i'\Omega^{-1}\epsilon_i\right),\ i = 1, \dots, n, \tag{20}$$

whose covariance matrix is $\frac{\nu}{\nu-2}\Omega$, provided $\nu > 2$, and $\Gamma(\cdot)$ denotes the *gamma* function. Relative to (11) only minor modifications are needed in MCMC to accommodate drawing ν and U_is, as working directly with (20) is quite cumbersome (for details see Geweke, 1993, and Fernandez and Steel, 1999). Our prior for the degrees of freedom follows Geweke (1993) and we assume:

$$p(\nu) = \lambda e^{-\lambda\nu},\ \nu > 0, \tag{21}$$

where $\lambda = \frac{1}{5}$, implying that the prior mean and variance of ν are 5, i.e. we have considerable deviations from normality but the Student-t distribution possesses moments up to fourth order. Deviations from normality can be justified, *a priori*, based on the evidence and concerns of Greco et al. (2019).

It is notable that in Greco et al. (2019a) technical efficiency is measured using

$$u_i^* = \min_{\alpha,\beta\geq 0,\ \alpha+\beta=1}\ \min_{(\sigma',\mu')\in\hat{\Psi}} F_{\alpha,\beta}(\sigma_i, \mu_i) - F_{\alpha,\beta}(\sigma', \mu'), \tag{22}$$

where $F_{\alpha,\beta}(\sigma, \mu) = \alpha\mu + \beta\sigma$, $\hat{\Psi} = \{(\sigma, \mu\}\,|\nexists(\sigma', \mu') \in \Psi\ \&\ (\sigma', \mu') \neq (\sigma, \mu),\ \sigma' \leq \sigma,\ \mu' \geq \mu\}$, and

$$\Psi = \left\{ \left(\sum_{i=1}^n \omega_i\mu_i, \sum_{i=1}^n \omega_i\sigma_i \right) \mid \omega_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \omega_i = 1 \right\}.$$

The well-known Debreu-Farrell efficiency measure may be defined as $\theta_\mu(\sigma_i, \mu_i) = \min\left\{\theta|(\sigma_i, \theta\mu_i) \in \hat{\Psi}\right\}$ in the case of μ-orientation, and $\theta_\sigma(\sigma_i, \mu_i) = \max\left\{\theta|\ (\theta\sigma_i, \mu_i) \in \hat{\Psi}\right\}$, in the case of σ-orientation. Pareto-Koopmans efficiency requires $\theta_\mu(\sigma_i, \mu_i) = 1$ or $\theta_\sigma(\sigma_i, \mu_i) = 1$ depending on the orientation. In our model, inefficiency is measured directly using (10) and (11). In (3) there is no choice of orientation as inefficiency is directly attached to the composite indicator.

Similar to (17) we can define the posterior standard deviation of the composite indicator:

$$SDCI_i = \sqrt{S^{-1}\sum_{s=1}^{S}\left(CI_i^{(s)} - \overline{CI_i}\right)^2}, i = 1, \ldots, n. \tag{23}$$

In turn, we can consider a model of the form:

$$\overline{CI}_i = b_0 + b_1 SDCI_i + b_2 SDCI_i^2 + \tilde{v}_i - \tilde{u}_i,\ i = 1, \ldots, n, \tag{24}$$

where $v_i \sim iid\,\mathcal{N}(0, \sigma_{\tilde{v}}^2)$, and, independently, $u_i \sim iid\,\mathcal{N}_+(0, \sigma_{\tilde{u}}^2)$. The model should be a reasonable approximation based on Figure 3, 4 and 7 in Greco et al. (2019) and it is consistent with μ-orientation.[4] A potential problem is that we do not allow for different groups. The problem can corrected easily, using the following model:

$$\overline{CI}_i = b_{0,G_i} + b_{1,G_i} SDCI_i + \tfrac{1}{2} b_{2,G_i} SDCI_i^2 + \tilde{v}_{i,G_i} - \tilde{u}_{i,G_i},\ i = 1, \ldots, n, \tag{25}$$

where $\tilde{v}_i \sim iid\,\mathcal{N}(0, \sigma_{\tilde{v},G_i}^2)$, and, independently, $\tilde{u}_i \sim iid\,\mathcal{N}_+(0, \sigma_{\tilde{u},G_i}^2)$, and $G_i \in \mathcal{G} = \{1, \ldots, \overline{G}\} \subset \mathbb{Z}$, indicates the group membership ($\mathbb{Z}$ denotes the set of integers in $\Re$). There are, potentially, $\overline{G}$ groups at most, and the question is to estimate G_i and the parameters conditional on the data. It is well-known that this is a mixture model that can be analyzed using standard Bayesian MCMC techniques. We impose monotonicity and concavity restrictions using:[5]

$$b_{1,G_i} + b_{2,G_i} SDCI_i \geq 0,\ b_{2,G_i} \leq 0\ \forall G_i \in \mathcal{G},\ i = 1, \ldots, n. \tag{26}$$

An alternative is to modify the model as follows:

$$CI_i = x_i' w + v_i, i = 1, \ldots, n, \tag{27}$$

$$x_i = f_i 1_J + \epsilon_i, i = 1, \ldots, n, \tag{28}$$

$$CI_i \text{ is "as close as possible" to } f_i,\ i = 1, \ldots, n, \tag{29}$$

from which inefficiency (u_i) is absent. For models that include inefficiency, for each MCMC draw, consider

$$CI_i^{(s)} = x_i' w^{(s)},\ SD_i^{(s)} = \sqrt{x_i' V_w x_i},\ i = 1, \ldots, n,\ s = 1, \ldots, S, \tag{30}$$

where $V_w = S^{-1}\sum_{s=1}^{S}(w^{(s)} - \overline{w})(w^{(s)} - \overline{w})' = S^{-1}\sum_{s=1}^{S} w^{(s)} w(s)' - \overline{w} \cdot \overline{w}'$, and $\overline{w}$ is defined in (15). Given $\overline{w}$ these measures can be computed easily. Then, for each MCMC draw ($s = 1, ..., S$), we estimate the following model by MCMC

[4] Clearly, $\sigma-$ orientation can be easily accommodated.

[5] These conditions are imposed easily via rejection sampling in MCMC, i.e. draws which do not satisfy the constraints are rejected.

for mixtures (Geweke and Amisano, 2007):

$$CI_i^{(s)} = b_{0,G_i}^{(s)} + b_{1,G_i}^{(s)} SD_i^{(s)} + \tfrac{1}{2} b_{2,G_i}^{(s)} SD_i^{(s)\,2} + \tilde{v}_{i,G_i}^{(s)} - \tilde{u}_{i,G_i}^{(s)},\ i = 1, \ldots, n, \tag{31}$$

subject to (26) for each $G_i \in \mathcal{G}$. The model provides directly $\mu - \sigma$ inefficiency measures as:

$$u_i = S^{-1} \sum_{s=1}^{S} \tilde{u}_{i,G_i}^{(s)},\ i = 1, \ldots, n, \tag{32}$$

with the understanding that $\tilde{u}_{i,G_i}^{(s)}$ is really $\tilde{u}_{i,G_i^{(s)}}^{(s)}$, since group membership may change across MCMC iterations. Equation (32) provides **local inefficiency**, that is inefficiency of a unit relative to its own frontier. **Global inefficiency** is defined relative to a meta-frontier as follows. Given SD_i, suppose G_i^* maximizes the value of: $b_{0,G_i}^{(s)} + b_{1,G_i}^{(s)} SD_i^{(s)} + \frac{1}{2} b_{2,G_i}^{(s)} SD_i^{(s)\,2}$, with respect to G_i for all $s = 1, ..., S$. In turn, global inefficiency is:

$$u_i^{global} = u_i + \left[\left(b_{0,G_i^*}^{(s)} - b_{0,G_i}^{(s)} \right) + \left(b_{1,G_i^*}^{(s)} - b_{1,G_i}^{(s)} \right) SD_i^{(s)} + \tfrac{1}{2} \left(b_{2,G_i^*}^{(s)} - b_{2,G_i}^{(s)} \right) SD_i^{(s)\,2} \right]. \tag{33}$$

The models in (25) and (31) are called **Models A and B**, respectively. **Model A** using posterior means of the composite indicator and its standard deviation, whereas **Model B** accounts for parameter uncertainty in the composite indicator and its standard deviation. On prior grounds, this is an important feature and the results delivered by Models A and B could be very different. Our **Model C** allows for random country-specific weights:

$$w_i \sim iid\, \mathcal{N}_J(\underline{w},\ \underline{\sigma}_w^2 \mathbf{I}_J),\ i = 1, \ldots, n, \tag{34}$$

where $\underline{w} = 0.5$, $\underline{\sigma}_w = 0.5$, and $\mathbf{I}_J$ denotes the $J \times J$ identity matrix. For practical purposes this prior is rather diffuse. The importance of Model C is in the fact that we account for heterogeneity. In effect, *if this model receives support in the light of the data, it means that a single composite indicator with common weights, may be problematic.* In effect, this would put into doubt the notion that there is a single composite indicator that can summarize the data.

Models D, E, F are the same as A, B, C, except that we adopt a multivariate Student-t distribution for ϵ_is as in (20).

4 Empirical application

As in Greco et al. (2019a) we apply $\sigma - \mu$ efficiency analysis to the data set in the 2017 Report of 'World Happiness'. The 'World Happiness' report (WHR, Helliwell et al., 2017) presents and analyses data of a survey conducted by the Gallup World Poll.[6] There are six key variables (GDP per capita, healthy life expectancy at birth, social support, freedom to

[6]Specifically, 3,000 respondents in each of the -roughly- 150 countries considered, evaluate their lives on a 0-10 scale known as 'Cantril Ladder' (see Helliwell et al., 2017, p.123). As in Greco et al. (2019) "[w]e use a three-year rolling-window for the six variables, in order to be consistent with the procedure used by theWorld Happiness Report for the subjective evaluation. This means that the values we consider in each dimension in year 2016 are in fact non-weighted arithmetic averages of the period 2014-2016. We restrict the sample to only these countries that possess data for all 6 dimensions for the 2016 and at least one of the years 2014 and 2015. After this data cleaning procedure we are left

Table 1: Bayes factors

model	Bayes factor
Model A without inefficiency	1.000 (6 groups)
Model A	17.42 (4 groups)
Model B	85.12 (5 groups)
Model C	21.30 (3 groups)
Model D	44.82 (6 groups)
Model E	113.32 (5 groups)
Model F	17.12 (6 groups)

Notes: Bayes factors have been computed for all combinations of models and number of groups $G \in \mathcal{G} = \{1, ..., 10\}$. The optimal number of groups in indicated in parentheses. Models A, B, C use the multivariate normality assumption as in (10). Models D, E, F use the multivariate Student-t assumption as in (20).

make life choices, generosity and perceptions of corruption) for 119 countries. Weights of the composite indicator are ranked in this order.

Following Greco et al. (2019a) we discuss and report only the efficiency of the top-10 ranked countries of the 2017 'World Happiness' report. Greco et al. (2019a) find that the top ten countries are countries found in the top ten rankings are the following: Norway, Denmark, Iceland, Switzerland, Finland, the Netherlands, Canada, New Zealand, Australia and Sweden, in this order. However, we find five or six groups when using Models D, E or F, instead of a total of 31 as in Greco et al. (2019a).

Greco et al. (2019a) find that "the countries which are self-claimed to be ranked in the top-10 positions (i.e. having the top-10 highest subjective evaluation) are positioned in our top-10 list as well, with the exception of Iceland and Finland, which we position in the 11th and 13th places accordingly". Our findings are similar. Before discussing the empirical results, it is important to perform model selection as we have six competing specifications (viz. Models A–F). The Bayes factors in favor of each model relative to the benchmark are presented in Table 1 along with the optimal number of groups in each case. We have considered 60 models (models A–F along with unknown number of groups $G \in \{1, \ldots, 10\}$). As a benchmark we consider Model A *without* inefficiency and we compute all Bayes factors relative to this model. [7] Evidently, Model E is strongly preferred over the alternatives so, we proceed using Model E. The importance of this finding is that a single set of weights can be used and heterogeneous weights are not needed in terms of fit versus parsimony. From another, perspective, *comparing the benchmark model with Model E, provides a principled way to test that a single common indicator is acceptable in the light of the data, instead of taking it for granted.* In other words, a single common composite indicator is valid but in addition a Student-t distribution is also strongly favored by the data.The Bayes factors favor by far Model E.

Posterior results for the WHR data are provided in Table 2 along with the WHR and Greco et al. (2019) rankings. Besides the apparent difference in rankings which is due to the non-uniform weights and statistical inference, it is of some interest to notice that, for example, Singapore is ranked 2 by using uniform weights but falls all the way to 35 in model D

with a final sample of 119 countries".

[7] Suppose $p(\theta|X) \propto L(\theta; X)p(\theta)$ is the posterior of a certain model, and $\mathscr{M}(X)$ is the normalizing constant of the posterior, viz. $\mathscr{M}(X) = {}_{\Theta} L(\theta; X)p(\theta)\mathrm{d}\theta$ where $\Theta \subset \Re^d$ is the parameter space. This is known as marginal (or integrated) likelihood and summarizes all evidence (data-based and prior-based) in favor of a given specification. Clearly, $p(\theta|X)d\theta = \frac{L(\theta;X)p(\theta)}{{}_{\Theta} L(\theta;X)p(\theta)\mathrm{d}\theta} = \frac{L(\theta;X)p(\theta)}{\mathscr{M}(X)}$. The marginal likelihood is obtained using the procedure in Chib (1995) which makes use of the conditional posterior distributions. Given marginal likelihoods, say $\mathscr{M}_m(X), m = 1, ..., M'$, and the marginal likelihood of a benchmark model, say $\mathscr{M}_o(X)$, the Bayes factor in favor of model m and against model o is: $BF_{m,o}(X) = \frac{\mathscr{M}_m(X)}{\mathscr{M}_o(X)}$, $m = 1, ..., M'$. For any model $m' \neq m$ the Bayes factor $BF_{m,m'}(X) = \frac{\mathscr{M}_m(X)}{\mathscr{M}_{m'}(X)} = \frac{\mathscr{M}_m(X)}{\mathscr{M}_o(X)} \frac{\mathscr{M}_o(X)}{\mathscr{M}_{m'}(X)} = \frac{BF_{m,o}(X)}{BF_{m',o}(X)}$, so the choice of the benchmark model does not affect the results.

Table 2: Empirical results for WHR data

Country	WHR rank$^{(a)}$	Greco et al. (2019) rank and inefficiency$^{(a)}$	Model D	Model E
Norway	1	6, 0.034	2, 0.042	2, 0.034
Denmark	2	3, 0.033	1, 0.017	1, 0.033
Iceland	3	11, 0.016	13, 0.025	13, 0.016
Switzerland	4	7, 0.020	9, 0.014	9, 0.020
Finland	5	13, 0.030	12, 0.024	12, 0.030
Netherlands	6	10, 0.028	11, 0.017	11, 0.028
Canada	7	9, 0.036	3, 0.056	3, 0.036
New Zealand	8	1, 0.052	2, 0.037	2, 0.052
Australia	9	4, 0.038	5, 0.042	4, 0.038
Sweden	10	5, 0.028	4, 0.039	5, 0.028
Austria	13	17, 0.032	16, 0.043	17, 0.032
United States	14	19, 0.011	18, 0.025	19, 0.011
Ireland	15	8, 0.038	9, 0.031	9, 0.038
Germany	16	15, 0.031	17, 0.075	15, 0.031
Belgium	17	18, 0.007	19, 0.022	19, 0.007
Luxembourg	18	12, 0.010	14, 0.028	14, 0.010
United Kingdom	19	14, 0.018	13, 0.041	16, 0.018
Singapore	26	2, 0.034	35, 0.044	7, 0.034
Nicaragua	41	33, 0.005	52, 0.012	50, 0.005
Ecuador	44	38, 0.002	41, 0.007	40, 0.002
Kazakhstan	60	30, 0.006	35, 0.002	37, 0.006
Hong Kong	71	16, 0.012	22, 0.012	24, 0.012
Honduras	91	40, 0.016	91, 0.019	40, 0.016
F.Y.R. of Macedonia	92	41, 0.004	44, 0.001	44, 0.004
Egypt	111	55, 0.000	57, 0.000	55, 0.000
Iraq	117	54, 0.000	59, 0.000	56, 0.000

Notes: $^{(a)}$ Taken from Table 4 of Greco et al. (2019a). Inefficiency corresponds to δ_{i7} in Table 4 of Greco et al. (2019a) which is based on seven frontiers. The ranks correspond to $\sigma - \mu$ frontier (sixth column in Table 4 of Greco et al. (2019a).

but is 7 in model E. This fact is due to the deviation of optimal weights from uniformity. In addition, the change in rank is justified by handling of outliers through the Student-t distribution, inefficiency in the happiness index and incorporation of statistical uncertainty about the weights. The fact that there is considerable statistical uncertainty about weights is also illustrated in Figure 1, panel (b). In panel (d) of the same Figure, it is evident that there is considerable inefficiency (ranging from zero to roughly 40% averaging 15-20%. Local inefficiency is less pronounced (panels (d) and (f) of Figure 1) and averages roughly 2% across the different models. Besides parameter (weight) uncertainty there is also, and perhaps more importantly, model uncertainty as illustrated in Figure 2 which reports results for different models. Clearly, results from the model with the highest Bayes factor (which is the posterior odds ratio when the prior odds are 1:1) should be preferred.

It is, perhaps, of interest to notice that (from results not reported here in the interest of space) we obtain the following posterior median rankings for selected countries: Saudi Arabia 21, United Arab Emirates 28, Qatar 29, Bahrain 39 and Kuwait 40. From the 2019 WHR report (covering the period 2016–18) the reported rankings were, respectively, 28, 21, 29, 37, and 51. From the 2017 WHR report (covering the period 2014–16) the reported rankings were, respectively, 37, 21, 35, 41, and 39. So, in these cases, our results roughly match those in the WHR report.

Table 3: Posterior statistics for weights (Model E)

	posterior mean	posterior s.d.
GDP per capita	0.2433	0.0299
healthy life expectancy at birth	0.2016	0.0239
social support	0.1701	0.0145
freedom to make life choices	0.1472	0.016
generosity	0.1275	0.0168
perceptions of corruption	0.1104	0.0151

Notes: Weights are ordered from high to low corresponding to the variables in the order they appear in the Table.

Reported in Figure 1 are marginal posterior densities of the composite indicator weights in panel (a), using Model E, the posterior distribution of percentage difference between the common factor and the composite indicator, in panel (a), and marginal posterior densities of global inefficiencies for models A–F in panels (c)-(f).

In panel (b) of Figure 1, reported are percentage differences between the composite indicator and the common factor, which range from -0.2% to 0.3% with considerable probability mass in the neighborhood of zero. In effect, the hypothesis that these differences are zero cannot be rejected in the light of the data. In panel (a) of Figure 1, reported are posterior distributions of weights (for the preferred model E). Marginal posterior densities of global and local inefficiency, for all models are reported in panels (c)–(f) of the same Figure. Evidently, global inefficiencies are much larger as they are defined relative to a metafrontier as in (33). In Figure 2, reported are marginal posterior cumulative distribution functions of the composite indicator weights. It is evident that the distributions satisfy a no-crossing property so, the weights can be ordered without ambiguity.

In Figure 3, reported are posterior densities of the percentage difference between the composite indicator and the common factor for all models A–F. Apart for Model E, posterior densities of the percentage difference have considerably more spread. For Models D and F, due to bimodality, it is not even clear that this condition holds.

Posterior densities of degrees of freedom (ν) for the multivariate Student-t, are reported in Figure 4. For the preferred model E, they range from 1 to 8 with modes around 3 and 6 so, it is doubtful that fourth moments exist. For models D and F the posteriors are unimodal with posterior mean around 10 but they range from 2 to 16 so, there is considerable posterior uncertainty. However, based on the probable values of ν it is clear that outliers are accounted for, as ν can be quite low and, in any case, nowhere near 30, which corresponds, approximately, to a multivariate normal distribution.

Finally, we discuss posterior sensitivity to prior values of λ in (21) as it relates to (19) or (20). We draw randomly 1,000 values of λ in the interval $\left[\frac{1}{30}, 2\right]$ which implies that prior expectations about the degrees of freedom parameter (ν) of the multivariate Student-t distribution range from $\frac{1}{2}$ to 30. For $\nu \simeq 30$ the Student-t and the normal behave roughly the same.[8] To conserve space, we report the absolute percentage differences of new marginal posterior moments of weights (across all 1,000 priors and relative to the benchmark prior) in panel (g) of Figure 4. Evidently, the posterior moments are approximately the same and they do not deviate significantly from the benchmark prior ($\lambda = \frac{1}{5}$).

[8] The new posterior moments are computed using the sampling-importance-resampling method of Rubin (1987) and Smith and Gelfand (1992). We use a 20% sub-sample size from the original MCMC sample.

Figure 1: Aspects of the posterior distribution

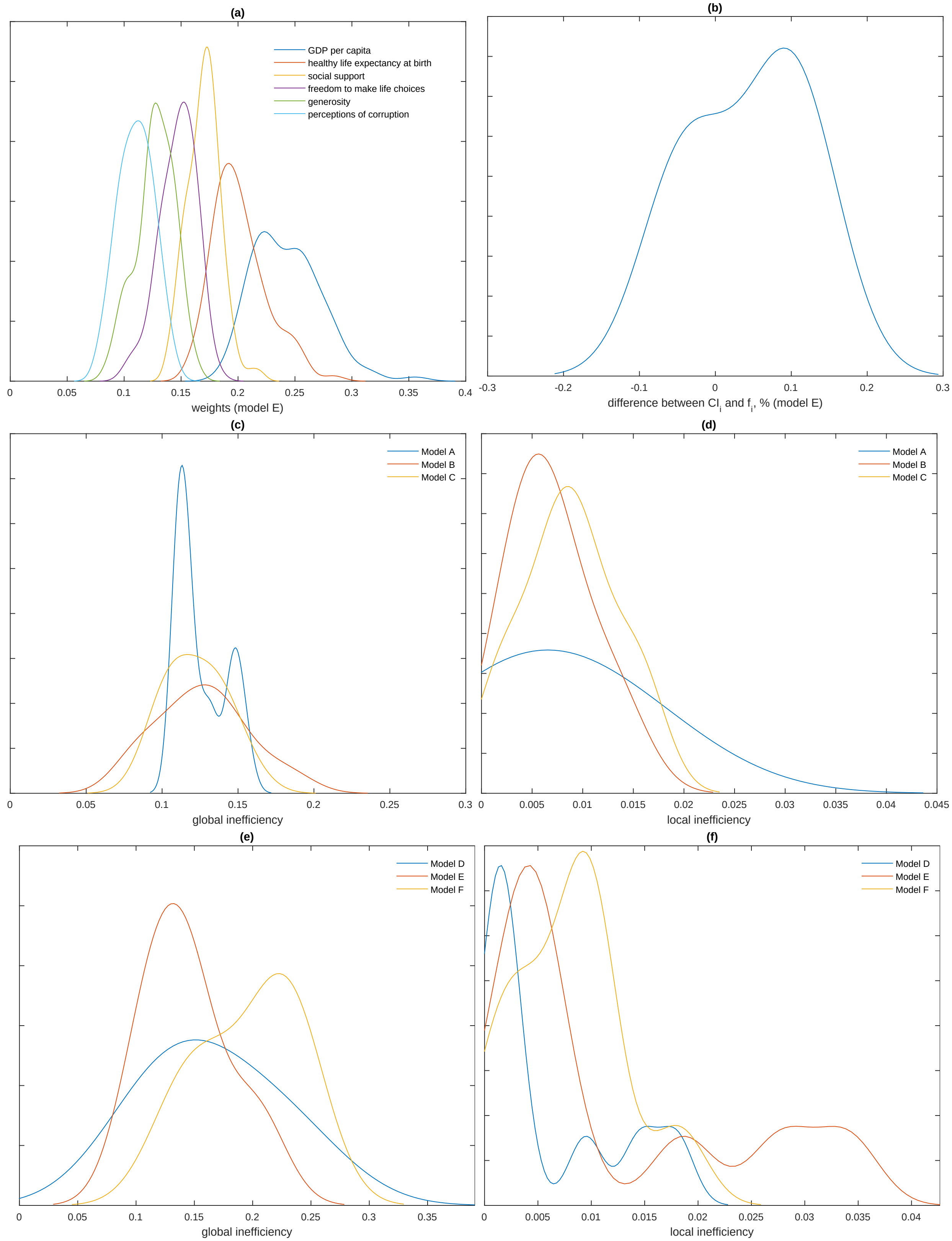


Figure 2: Marginal posterior distributions of weights

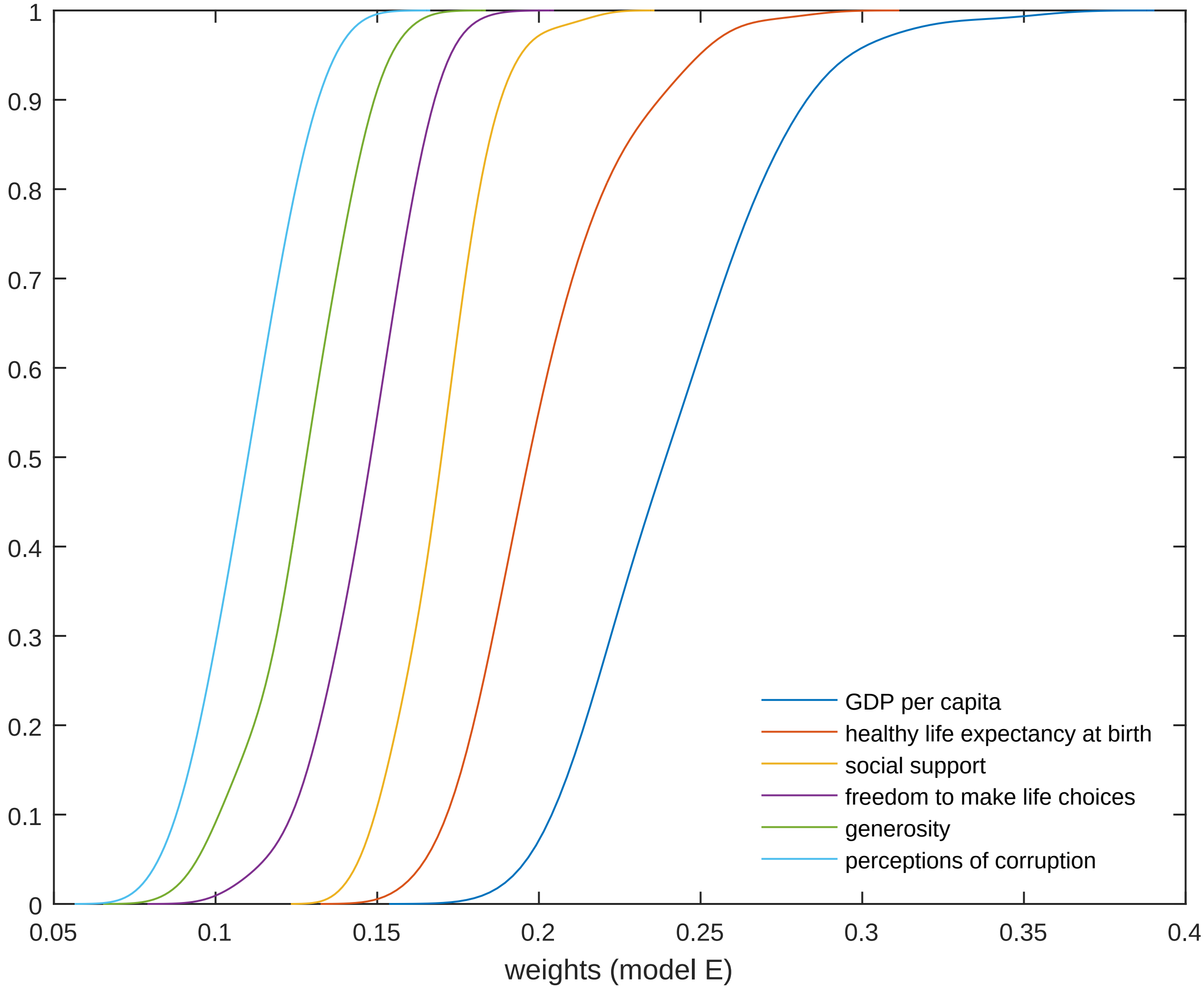


Figure 3: Posterior densities of difference between CI_i and f_i for different models

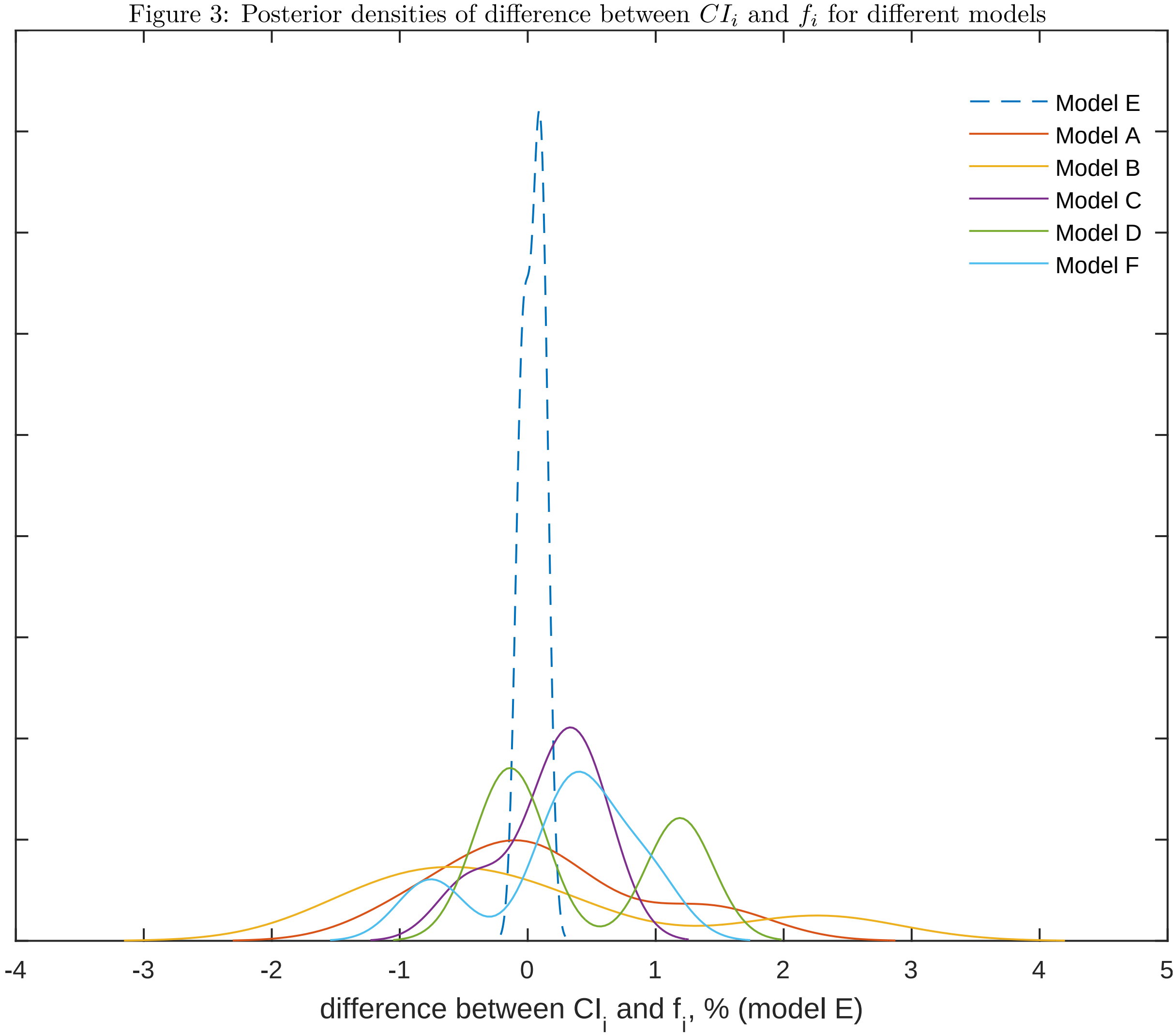
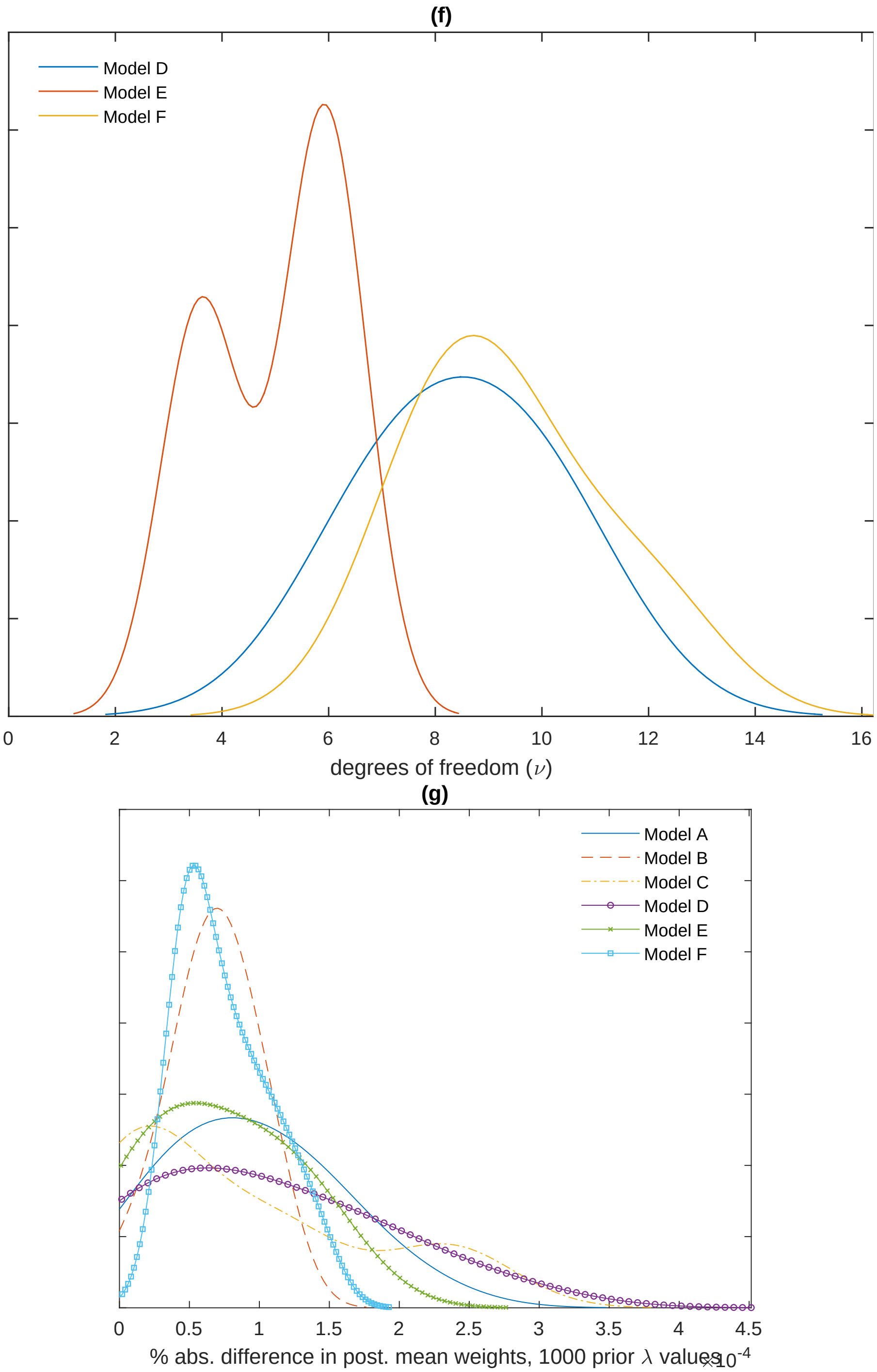


Figure 4: Posterior densities of degrees of freedom (ν) for multivariate Student-t distribution of ϵ_is and prior sensitivity analysis



Concluding Remarks

We provided a novel model of constructing composite indicators that takes into account statistical uncertainty about weights of different predetermined variables on the indicator, and can handle outliers in a formal way via a Student-t distribution. In an empirical application to WHR data we document important differences in rankings of different countries in terms of the happiness indicator. We also document that there is significant model uncertainty so model comparison and model selection (via Bayes factors) should be part of any analysis of composite indicators as they seem to be sensitive to model specification and / or outlying observations.

Approximate Bayesian Computation (ABC) approaches are strongly encouraged for future work, along with Info-Metrics techniques (to avoid many distributional assumptions required here, which could be a drawback). ABC methods, usually, rely on moments but, on the other hand, moments are sensitive to outliers. Ways to resolve this problem in ABC are necessarily left for future research.

TECHNICAL APPENDIX

To provide access to the posterior in (11), we use a Gibbs sampler which is based upon drawing successively from the posterior conditional distributions associated with (11). Given starting values $f_i^{(0)}, w^{(0)}, u_i^{(0)}, \sigma_v^{(0)}, \sigma_u^{(0)}, \sigma_f^{(0)}, \Omega^{(0)}$, we draw from the following posterior conditional distributions, for $s = 1, \ldots, S$.

$$CI_i^{(s)} \sim CI_i | f_i^{(s-1)}, w^{(s-1)}, u_i^{(s-1)}, \sigma_v^{(s-1)}, \sigma_u^{(s-1)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X),\ i = 1, ..., n, \tag{A.1}$$

$$f_i^{(s)} \sim f_i | CI_i^{(s)}, w^{(s-1)}, u_i^{(s-1)}, \sigma_v^{(s-1)}, \sigma_u^{(s-1)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X), \tag{A.2}$$

$$w^{(s)} \sim w | CI_i^{(s)}, f_i^{(s)}, u_i^{(s-1)}, \sigma_v^{(s-1)}, \sigma_u^{(s-1)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X), \tag{A.3}$$

$$u_i^{(s)} \sim u_i | CI_i^{(s)}, f_i^{(s)}, w^{(s)}, \sigma_v^{(s-1)}, \sigma_u^{(s-1)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X),\ i = 1, \ldots, n, \tag{A.4}$$

$$\sigma_v^{(s)} \sim \sigma_v | CI_i^{(s)}, f_i^{(s)}, w^{(s)}, u_i^{(s)}, \sigma_u^{(s-1)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X), \tag{A.5}$$

$$\sigma_u^{(s)} \sim \sigma_u | CI_i^{(s)}, f_i^{(s)}, w^{(s)}, u_i^{(s)}, \sigma_v^{(s)}, \sigma_f^{(s-1)}, \Omega^{(s-1)}, X), \tag{A.6}$$

$$\sigma_f^{(s)} \sim \sigma_u | CI_i^{(s)}, f_i^{(s)}, w^{(s)}, u_i^{(s)}, \sigma_v^{(s)}, \sigma_u^{(s-1)}, \Omega^{(s-1)}, X), \tag{A.7}$$

$$\Omega^{(s)} \sim \Omega | CI_i^{(s)}, f_i^{(s)}, w^{(s)}, u_i^{(s)}, \sigma_v^{(s)}, \sigma_u^{(s)}, \sigma_f^{(s-1)}, X). \tag{A.8}$$

In turn, $\left[f_i^{(s)}, w^{(s)}, u_i^{(s)}, \sigma_v^{(s)}, \sigma_u^{(s)}, \sigma_f^{(s)}, \Omega^{(s)}, s = 1, ..., S\right]$ is a sample from the distribution whose non-normalized density is given by (11). The Gibbs sampler operates by drawing from the following conditional posterior distributions:

$$CI_i | \cdot \sim N\left(\frac{h^2(x_i'w - u_i) + \sigma_v^2 f_i}{\sigma_v^2 + h^2},\ \frac{\sigma_v^2 h^2}{\sigma_v^2 + h^2}\right), i = 1, \ldots, n, \tag{A.9}$$

$$f_i|\cdot \sim N\left(\frac{\sigma_f^2 \cdot 1'_J\Omega^{-1}x_i}{\sigma_f^2 1'_J\Omega^{-1}1_J+1},\ \frac{\sigma_f^2}{\sigma_f^2 1'_J\Omega^{-1}1_j+1}\right), i=1,\ldots,n, \tag{A.10}$$

$$w|\cdot \sim N_J\left((\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\left(CI+u\right),\ \sigma_v^2(\mathbb{X}'\mathbb{X})^{-1}\right), \tag{A.11}$$

$$u_i|\cdot \sim N_+\left(\frac{(x_i'w-CI_i)\sigma_u^2}{\sigma_v^2+\sigma_u^2},\ \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2+\sigma_u^2}\right), i=1,\ldots,n, \tag{A.12}$$

$$\frac{\sum_{i=1}^n(CI_i-x_i'w+u_i)^2+\bar{q}}{\sigma_v^2}|\cdot \sim \chi_n^2, \tag{A.13}$$

$$\frac{\sum_{i=1}^n u_i^2}{\sigma_u^2}|\cdot \sim \chi_n^2, \tag{A.14}$$

$$\frac{\sum_{i=1}^n(CI_i-f_i)^2}{\sigma_f^2}|\cdot \sim \chi_n^2, \tag{A.15}$$

$$p(\Omega|\cdot) \propto |\Omega|^{-(n+J+1)/2}\exp\left\{-\tfrac{1}{2}tr\Omega^{-1}\mathbb{Q}\right\}, \tag{A.16}$$

where $\mathbb{Q}=\sum_{i=1}^n(x_i-f_i1_J)(x_i-f_i1_J)'$, and $\mathbb{X}$ is the $n\times J$ matrix containing all data on x_is. All these distributions are well known and random number generation is quite easy. In particular, the last one is an inverted Wishart distribution (Zellner, 1971, pp. 395–396).

References

[1] Chib, S. (1995). Marginal Likelihood from the Gibbs Output. Journal of the American Statistical Association, 90, 1313–1321.

[2] Doumpos, M., Gaganis, C., and Pasiouras, F. (2016). Bank Diversification and Overall Financial Strength: International Evidence. Financial Markets, Institutions & Instruments, 25 (3):169–213.

[3] Doumpos, M., Hasan, I., and Pasiouras, F. (2017). Bank overall financial strength: Islamic versus conventional banks. Economic Modelling, 64:513–523.

[4] Fernandez, C., and Steel, M. F. J. (1999). Multivariate Student-t regression models: Pitfalls and inference. Biometrika 86, 153–67.

[5] Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. Journal of Applied. Econometrics 8, 519–40.

[6] Geweke, J., and Amisano, G. (2007). Interpretation and inference in mixture models: Simple MCMC works. Computational Statistics & Data Analysis 51(7): 3529–3550.

[7] Greco, S., Ehrgott, M., and Figueira, J. (2016). Multiple Criteria Decision Analysis: State of the Art Surveys. International Series in Operations Research & Management Science. 2nd edition, New York: Springer.

[8] Greco, S., Ishizaka, A., Matarazzo, B., and Torrisi, G. (2018a). Stochastic multi-attribute acceptability analysis (SMAA): an application to the ranking of Italian regions. Regional Studies, 52(4): 585–600.

[9] Greco, S., Mousseau, V., and Słowinski, R. (2008). Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. European Journal of Operational Research, 191(2): 416–436.

[10] Greco, S., Słowinski, R., Figueira, J., and Mousseau, V. (2010). Robust ordinal regression, in Trends in multiple criteria decision analysis, Editors: Greco, Salvatore, Ehrgott, Matthias, Figueira, José Rui (Eds.), Springer, pp. 241–283.

[11] Greco, S. , Ishizaka, A. , Matarazzo, B. , & Torrisi, G. (2018). Stochastic multi-attribute acceptability analysis (SMAA): An application to the ranking of Italian regions. Regional Studies, 52 (4), 585–600 .

[12] Greco, S., A. Ishizaka, M. Tasiou, G. Torrisi (2019a). Sigma-Mu efficiency analysis: A methodology for evaluating units through composite indicators. European Journal of Operational Research, 278 (3), 942–960.

[13] Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2019b). On the methodological framework of composite indices: A review of the issues of weighting, aggregation and robustness, Social Indicators Research, 141 (1): 61–94.

[14] Helliwell, J., Layard, R., and Sachs, J. (2017). World Happiness Report 2017. New York: Sustainable Development Solutions Network.

[15] Miller, T., Kim, A. B., and Roberts, J. M. (2018). 2018 Index of Economic Freedom. Technical report, The Heritage Foundation.

[16] Rubin, D. B. (1987). A noniterative sampling-importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. Journal of the American Statistical Association 82, 543–546.

[17] Samans, R., Blanke, J., Corrigan, G., and Drzeniek, M. (2018). The Inclusive Growth and Development Report 2018. In Geneva: World Economic Forum.

[18] Smith, A. F. M., and Gelfand, A. (1992). Bayesian statistics without tears: A sampling–resampling perspective. The American Statistician 46, 84–88.

[19] UNDP (2010). Human Development Report (HDR) 2010: The Real Wealth of Nations: Pathways to Human Development. Technical report, United Nations Development Programme (UNDP). Retrieved from http://hdr.undp.org/en/content/human-development-report-2010

[20] Zellner, A, (1971). An Introduction to Bayesian Inference in Econometrics, Wiley, New York.