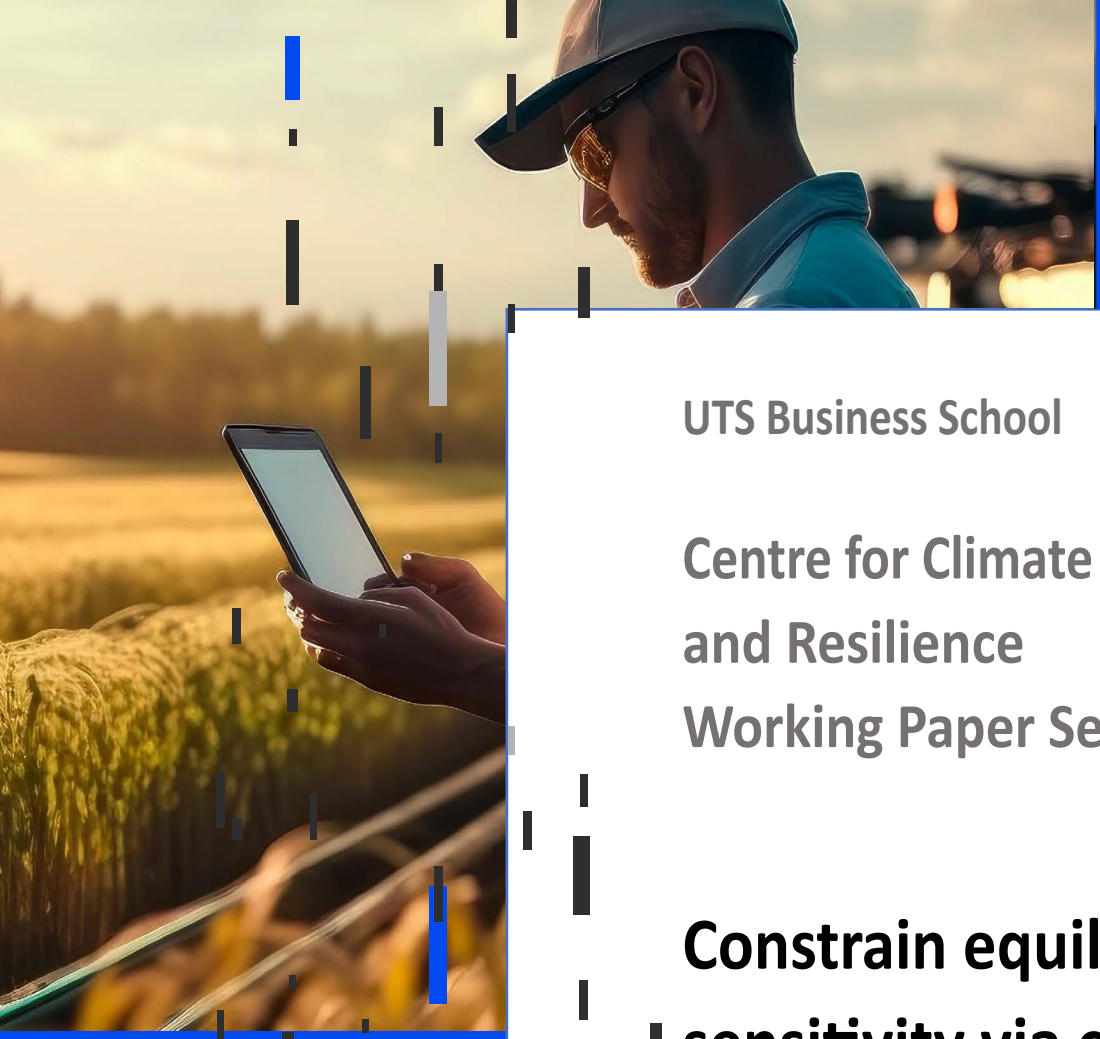

UTS Business School

Centre for Climate Risk and Resilience Working Paper Series

Constrain equilibrium climate sensitivity via composite likelihood

By Mengheng Lia, Donald P. Cummins

Working Paper No. 2024/1

Series Editor: Martina Linnenluecke, CCRR Director

Connect with CCRR
Web https://www.uts.edu.au/about/uts-business-school/research/centre-climate-risk-and-resilience
Email ccrr@uts.edu.au

Constrain equilibrium climate sensitivity via composite likelihood

Mengheng Li[a,b*] *Donald P. Cummins*[c]

[a] *Economics Discipline Group, University of Technology Sydney, Australia*

[b] *Centre for Climate Risk and Resilience, University of Technology Sydney, Australia*

[c] *Centre National de Recherches Météorologiques, Université de Toulouse, Météo-France, France*

The equilibrium climate sensitivity (ECS) stands as a pivotal parameter in climate science and other disciplines. This study estimates ECS as a common equilibrium parameter across energy balance models (EBMs). Fitting EBMs to the counterfactual data simulated from 31 climate models under a quadrupled CO_2 experiment, we introduce a Bayesian composite likelihood approach to simultaneously integrate and estimate all the constituent EBMs. In contrast to inferential methods based on storylines and emergent constraints commonly employed by climate scientists, our econometric alternative provides a data-driven ECS estimator with an intuitive probabilistic interpretation. We find an ECS estimate of 3.3K, characterized by a unimodal posterior distribution that facilitates uncertainty quantification. Our approach also yields a 95% credible interval of (3.18K, 3.44K), consistent with, yet notably tighter than, the likely range of ECS previously reported.

Keywords: *Composite likelihood, Equilibrium climate sensitivity, Energy balance model, Bayesian inference, model averaging*
JEL Classification: C11, C22, C51

*Corresponding author: M. Li, University of Technology Sydney, UTS Business School, Ultimo NSW 2007, Australia. mengheng.li@uts.edu.au

1 Introduction

Equilibrium climate sensitivity (ECS), the mean surface air temperature change with a sustained doubling of the atmospheric concentration of CO_2, is a key metric in anthropogenic climate change (Geoffroy et al., 2013b; Meehl et al., 2020). Determining its value and narrowing the range have been challenging tasks, with atmosphere-ocean general circulation models (AOGCMs) showing ECS variations from 1.8K to 6K (Stevens et al., 2016; Scafetta, 2022). The discrepancies are attributed to differing understandings of feedback mechanisms from water vapor, cloudiness, and aerosols (Knutti et al., 2017; Lewis and Curry, 2018). Efforts to constrain ECS values led to the Intergovernmental Panel on Climate Change (2021) recommending "likely" range of (2.5K, 4K) and "very likely" range of (2K, 5K) in its latest assessment report, based on AOGCMs' deviations from understood feedback mechanisms. Beyond climate science, accurate ECS estimation is crucial for Integrated Assessment Models (IAMs) in climate-economy studies, as different models yield drastically varying policy recommendations due to their sensitivity to ECS (Stern, 2007; Watkiss and Hope, 2011; Hahn and Ritz, 2015). IAMs' uncertainty in estimating economic costs hampers policymakers' ability to inform domestic and international climate policies (Clarke and Coauthors, 2014). Despite efforts to constrain the value of ECS, AOGCM-implied ECS uncertainty is sometimes deemed "irreducible" due to different physical processes under consideration and model resolutions (Lehner et al., 2020; Geoffroy et al., 2013a), emphasizing the need for an ECS estimate with sound inferential and risk properties consistent with physics.

This paper aims to integrate econometrics and climate science through a novel composite likelihood approach for estimating and constraining ECS (Varin et al., 2011; Canova and Matthes, 2021). Utilizing energy balance models (EBMs) as climate emulators for AOGCMs, which inevitably misspecify Earth's climate system, we employ EBMs to model the group temperature response to CO_2 radiative forcing. While EBMs have limitations compared to AOGCMs, their rising popularity among climate scientists and economists is attributed to their low computational cost and analytical tractability (Aldrin et al., 2012; Geoffroy et al., 2013b; Cummins

et al., 2020; Pretis, 2020; Bennedsen et al., 2023). Jackson et al. (2022) demonstrated satisfactory performance of EBMs for temperature projections under various forcing experiments. In essence, we treat EBMs as physically consistent time series models and fit them to AOGCM simulated data. Our key innovation lies in pooling information across AOGCMs via composite likelihood, recognizing that ECS is a well-defined physical quantity common across models with some uncertainty. Allowing other parameters in EBMs to differ, reflecting specificities in different AOGCMs, this approach offers a data-driven alternative for combining ECS estimates with statistical quantification of uncertainty. Additionally, it facilitates straightforward constraint adjustments using observational data or adding more models to the composite, letting data determine the value added by additional components.

Our contribution spans two key areas of literature. Firstly, within econometrics, where model selection and averaging methodologies, including Bayesian model averaging, model confidence set, and finite mixture models, typically involve considering different models fitted to the same dataset Raftery et al. (1997); Hansen et al. (2011); McLachlan et al. (2019). Our innovation lies in addressing an averaging problem with diverse datasets simulated by AOGCMs featuring distinct complexities and feedback mechanisms. Likelihood- and information criteria-based methods are inadequate due to the meaningless comparison of (marginal) likelihood with different datasets and the failure to distinguish the common estimate of interest from nuisance parameters. In our framework, allowing for differing nuisance parameters is essential, given the dissimilarity of AOGCMs. Additionally, we estimate model weights alongside other parameters, eliminating concerns about pre- or post-estimation weighting common in multimodel studies (Canova and Matthes, 2021). Our composite likelihood approach conceptually aligns with methods estimating a common quantity of interest from different sources (Durbin and Koopman, 2012, Chapter 3; Bennedsen et al., 2023). However, rooted in factor modeling, such methods do not discount different information sources. In comparison, our approach determines model weights in a data-driven way by concentrating explanatory power towards the common ECS. This ensures that even if an EBM fits an AOGCM simulation well, it receives a low weight if inconsistent with the composite ECS.

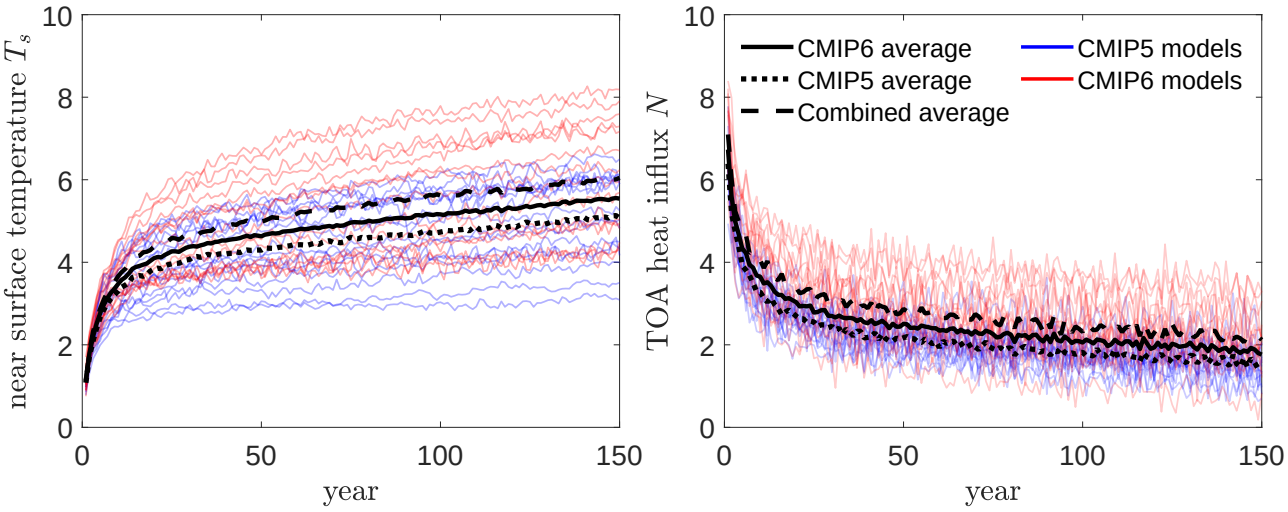

Figure 1: Simulated surface temperature T_s and TOA heat influx N by 31 AOGCMs under a quadrupled CO_2 experiment. CMIP5 and CMIP6 datasets are indicated by blue and red lines, respectively. Averaged data series are shown in black, with the solid, the dotted and the dashed line indicating CMIP5 average, CMIP6 average, and the average of the combination of CMIP5 and CMIP6 datasets, respectively.

Secondly, this study introduces a novel statistical toolkit to the literature on ECS and its admissible range. Stemming from the classical "Hasselmann model" (Hasselmann, 1976), the simple "Geoffroy regression" (Geoffroy et al., 2013b) relates global mean surface temperature T_s to radiative forcing $\mathcal{F}$ through $C_s \frac{dT_s}{dt} = \mathcal{F} - \lambda T_s = N$, where C_s is the surface heat capacity, N is the net radiative flux at top of the atmosphere (TOA), and λ is the feedback parameter. In equilibrium, for doubled ambient CO_2 forcing ($\mathcal{F} = F_{2\times CO2}$), ECS sets $dT_s = N = 0$, yielding ECS $= F_{2\times CO_2}/\lambda$. This establishes a regression that one can use to infer AOGCMs' ECSs via model runs, utilizing simulated data of $\mathcal{F}$, N, and T_s. Our study uses simulated T_s and N from 31 AOGCMs of Coupled Model Intercomparison Project (CMIP) phase 5 (as in (Geoffroy et al., 2013a)) and phase 6 ((Cummins et al., 2022)) under a quadrupled CO_2 experiment. Figure 1 illustrates our datasets, while Table 1 provides model names and expansions.

Several researchers, including Lewis and Curry (2018); Geoffroy et al. (2013a); Cummins et al. (2020); Jackson et al. (2022), have highlighted the limitations of the Geoffroy regression in capturing deep ocean heat uptake and Earth's thermal inertia, leading to inaccuracies in reflecting temperature changes over extended timescales. Addressing these concerns, we adopt

the approach proposed by Cummins et al. (2020) to formulate optimal EBMs. In contrast to earlier works that simply list estimated model-specific ECSs, such as Cummins et al. (2020) and Scafetta (2022), our study uses the composite ECS to effectively combine and weight information from different AOGCMs, providing a statistical reconciliation amid the ongoing debate on "low" vs. "medium" vs. "high" ECSs (Stevens et al., 2016; Meehl et al., 2020; Lehner et al., 2020; Scafetta, 2022). While not explicitly demonstrated here, our proposed composite method readily accommodates observational data from diverse sources (Aldrin et al., 2012; Pretis, 2020; Bennedsen et al., 2023) by treating them as additional model components or emergent constraints (Cox et al., 2018).

The remainder of this paper is structured as follows. Section 2 introduces the composite likelihood approach and Bayesian inference within this framework. In Section 3, we delve into the specifics of the EBM that is fitted to AOGCM simulated data, highlighting the parallels between the EBM (AOGCM) and a reduced-form (structural) model in economics. Section 4 presents the results of our estimations. We conclude with final remarks in Section 5.

2 Composite likelihood and quasi-Bayesian inference

This section presents the principles of composite likelihood (CL) based on the work of Varin et al. (2011) and, notably, Canova and Matthes (2021). Emphasizing the latter, who initially explored common parameter estimation and identification in macroeconomic models through quasi-Bayesian inference, we elucidate our model setup and interpretation of estimation results in a statistically robust manner that aligns with the perspective of climate scientists.

2.1 Composite likelihood for hypothetical climate models

The composite likelihood (CL) approach, initially designed to untangle model likelihoods in situations with latent variables or near-singular covariance matrices, offers an alternative objective function—an easily manageable weighted average of marginal or conditional distributions of submodels (Varin et al., 2011). Extending this framework, Canova and Matthes (2021) en-

Table 1: Model expansions (Geoffroy et al., 2013a; Cummins et al., 2022)

Model	Expansion
CMIP5 models	
BCC-CSM1.1	Beijing Climate Center, Climate System Model, version 1.1
BNU-ESM	Beijing Normal University-Earth System Model
CanESM2	Canadian Earth System Model, version 2
CCSM4	Community Climate System Model, version 4
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5
CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization Mark, version 3.6.0
FGOALS-s2	Flexible Global Ocean-Atmosphere- Land System Model gridpoint, second spectral version
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth Science Model 2M
GISS-E2-R	Goddard Institute for Space Studies Model E, coupled with Russell ocean model
HadGEM2-ES	Hadley Centre Global Environmental Model 2, Earth System
INM-CM4	Institute of Numerical Mathematics Coupled Model, version 4.0
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO, low resolution
MIROC5	Model for Interdisciplinary Research on Climate, version 5
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution
MRI-CGCM3	Meteorological Research Institute Coupled General Circulation Model, version 3
NorESM1-M	Norwegian Earth System Model, intermediate resolution
CMIP6 models	
ACCESS-ESM1.5	Australian Community Climate and Earth System Simulator, Earth System Model, version 1.5
BCC-CSM2-MR	Beijing Climate Center, Climate System Model, version 2, medium resolution
CanESM5	Canadian Earth System Model, version 5
CESM2	Community Earth System Model, version 2
CMCC-ESM2	Euro-Mediterranean Center on Climate Change, Earth System Model, version 2
CNRM-ESM2.1	Centre National de Recherches Météorologiques, Earth System Model, version 2.1
EC-Earth3-CC	European Community Earth System Model, version 3, carbon cycle
GFDL-ESM4	Geophysical Fluid Dynamics Laboratory, Earth System Model, version 4
GISS-E2.1-G	Goddard Institute for Space Studies, ModelE/GISS Ocean 2×2.5×L40
IPSL-CM6A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO, low resolution
MIROC-ES2L	Model for Interdisciplinary Research on Climate, Earth System version 2 for Long-term simulations
MPI-ESM1.2-LR	Max Planck Institute, Earth System Model, version 1.2, low resolution
MRI-ESM2.0	Meteorological Research Institute, Earth System Model, version 2.0
NorESM2-LM	Norwegian Earth System Model, version 2, low-resolution low-resolution atmosphere-land and medium-resolution ocean-sea ice
UKESM1.0-LL	United Kingdom Earth System Model, version 1.0, low resolution

hances the CL approach by allowing the estimation of parameters across models that are not or weakly identifiable individually. It further facilitates pooling information from diverse sources and enables data-driven determination of model weights. This setup aligns with our goal of constraining ECS across multiple EBMs using simulated datasets from various AOGCMs.

We extend the insights of Canova and Matthes (2021)'s work on dynamic models in economics to climate science. Akin to the perspective presented by Hasselmann (1976), Earth's climate system, subject to various forcings, behaves as a stochastic dynamic system. In the absence of forcings, the system maintains equilibrium with internal variability. External forcings induce deviations from equilibrium, stochastically amplifying or distorting climatic variations. This implies that with a complete understanding of Earth's climate system, there is a known data generating process (DGP) that induces a probability measure Ω and a density function $F(\{\boldsymbol{y}_t\}_{t=1}^{T}, \boldsymbol{\varphi})$ for an $n \times 1$ vector of climate variables $\boldsymbol{y}_t$ observed over T periods. The $k \times 1$ parameter vector $\boldsymbol{\varphi}$ splits into $\boldsymbol{\varphi} = (\alpha, \boldsymbol{\beta}')'$, where α signifies the ECS, not explicitly defined by Earth's climate system, and $\boldsymbol{\beta}$ includes physical quantities governing system dynamics and reflecting our understanding and measurement of the system.

Hypothetically, consider $\boldsymbol{y}_t$ is observed heterogeneously at various locations using different technologies, denoted as events. Let $\{A_i, i = \{1, ..., K\}\}$ be the set of marginal or conditional events of $\boldsymbol{y}_t$. Each event defines a submodel with subdensities $f(\{\boldsymbol{y}_{it}\}_{t=1}^{T_i} \in A_i, \alpha, \boldsymbol{\beta}_i)$ that induces a probability measure σ_i with $\sigma_i \subset \Omega$. $\boldsymbol{y}_{it}$ is the vector of climate variables observed for T_i periods at a location using a technology specified in A_i and characterized by $\boldsymbol{\beta}_i$. As such, α is the climate sensitivity that applies to all events in equilibrium, while $\boldsymbol{\beta}_i$'s are submodel-specific nuisance parameters. Given a vector of weights $0 < \omega_i < 1$, $\sum_i^K \omega_i = 1$, the CL is given by

$$F_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^{T_i}\}_{i=1}^{K}, \alpha, \{\boldsymbol{\beta}_i\}_{i=1}^{K}) \propto \prod_{i=1}^{K} f(\{\boldsymbol{y}_{it}\}_{t=1}^{T_i} \in A_i, \alpha, \boldsymbol{\beta}_i)^{\omega_i}. \tag{1}$$

Same as in Varin et al. (2011), Chan et al. (2020), and Canova and Matthes (2021), the above CL objective function ignores potential dependence across A_i's. However, as we argue in section 2.2, since we consider simulated counter-factual observations from AOGCMs, $\boldsymbol{y}_{it}$'s

are independent across i other than the fact that they all respect physical laws. Ideally, if each AOGCM captures all feedback mechanisms correctly and has infinitely high spatial resolution, all submodels with $T_1 = ... = T_K = T$ are correctly specified model for Earth's climate system, rendering $\{\boldsymbol{y}_{it}\}_{i=1}^K$ independent realizations from $F(\{\boldsymbol{y}_t\}_{t=1}^T, \boldsymbol{\varphi})$. Then it is trivial to show that under fixed weights, the maximum CL estimator $\hat{\alpha}$ satisfying

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}_1', ..., \hat{\boldsymbol{\beta}}_K') = \arg \max_{\alpha, \boldsymbol{\beta}_i, i=1,...,K} \sum_{i=1}^K \omega_i \log f(\{\boldsymbol{y}_{it}\}_{t=1}^T \in A_i, \alpha, \boldsymbol{\beta}_i) \tag{2}$$

is consistent, or $\hat{\alpha} \xrightarrow{p.} \alpha$, under large T or K asymptotics. Consistency is established as the first-order condition given by each correctly specified likelihood component in (2) is an unbiased estimating function, and so is any linear combination of them. In econometrics, this amounts to a linear combination of valid finite-sample moment conditions. Asymptotic normality immediately follows from the fact that under standard continuity and differentiability conditions, the likelihood functions of all submodels admit a Laplace approximation around the same mode, i.e, the maximum CL estimator $\hat{\alpha}$. Specifically,

$$\sqrt{KT}(\hat{\alpha} - \alpha) \overset{a.}{\sim} N(0, \mathcal{J}(\alpha)\mathcal{H}(\alpha)^{-2}), \tag{3}$$

where $\mathcal{H}(\alpha)$ and $\mathcal{J}(\alpha)$ are the asymptotic variability and sensitivity term, respectively (Varin et al., 2011). In finite samples, frequentist inference is straightforward with a consistent variance estimator $\widehat{\text{Var}}(\hat{\alpha}) = J(\hat{\alpha})H(\hat{\alpha})^{-2}$ satisfying $\frac{1}{KT}H(\hat{\alpha}) \xrightarrow{p.} \mathcal{H}(\alpha)$ and $\frac{1}{KT}J(\hat{\alpha}) \xrightarrow{p.} \mathcal{J}(\alpha)$; for example $H(\hat{\alpha})$ is the hessian, and $J(\hat{\alpha})$ is the outer product of the gradient.

Benefiting from independent AOGCM simulations, our setting is simpler than cases where events overlap, as considered by Engle and Kelly (2012), Varin et al. (2011), and Chan et al. (2020) with $A_i \cap A_j \neq \emptyset$. In these applications, some finite-sample correction is needed. Additionally, as (1) is not a likelihood function, its maximizer cannot achieve the Cram'er-Rao lower bound, as evident from the inefficient variance in (3). Model weights can be chosen based on maximizing efficiency. Canova and Matthes (2021) also explored data-driven weights with

$\omega_i \propto \exp(\mu_i)$, where μ_i is a function of statistics derived from past data in A_i.

2.2 Composite likelihood for EBMs

In practice, our approach deviates from the ideal setting in two aspects. Firstly, our understanding of $F(\{\boldsymbol{y}_t\}_{t=1}^T, \boldsymbol{\varphi})$ for Earth's climate system is limited due to current scientific constraints. Computational limitations prevent arbitrary refinement of AOGCMs, rendering them inherently under- and misspecified. Chan et al. (2020) and Canova and Matthes (2021) discuss a quasi-Bayesian CL approach to address misspecification in economic models. Secondly, we cannot econometrically estimate AOGCMs since they are not statistical models. Instead, we estimate EBMs as stochastic climate emulators approximating the global temperature response under a counter-factual quadrupled CO_2 experiment. As AOGCM-simulated data must adhere to energy balance principles, we follow Cummins et al. (2020) and Jackson et al. (2022) and fit EBMs to such data. Therefore, in this study we define the event A_i in CL as "fitting an EBM to climate data simulated by the i-th misspecified AOGCM", different but not incompatible with the broader interpretation where A_i represents different locations or technologies. An advantage of our approach is its flexibility to expand (1) with additional submodels.

Due to the approximating nature of EBMs and misspecifications in AOGCMs, the subdensities $f(\{\boldsymbol{y}_{it} \in A_i, \alpha, \boldsymbol{\beta}_i\})$ are also misspecified in the sense that under all $(\alpha', \boldsymbol{\beta}_i')'$ the Kullback-Leibler (KL) divergence of $f(\{\boldsymbol{y}_{it} \in A_i, \alpha, \boldsymbol{\beta}_i\})$ from $F(\{\boldsymbol{y}_t\}_{t=1}^T, \boldsymbol{\varphi})$ is strictly positive. Consequently, the CL (1) as a weighted geometric mean of subdensities has a strictly positive KL divergence. As the maximum likelihood first-order condition of each submodel is a biased estimating function, the associated maximum CL estimator $\hat{\alpha}$ is inconsistent. However, following Varin et al. (2011), it is possible to show that $\hat{\alpha}$ still converges in probability to a pseudo ECS α_0 that minimizes the KL divergence.

To see this, suppress the dependence on $\boldsymbol{\beta}_i$ and A_i, but denote each subdensity as $f_i(.)$ for exposition. Let $\boldsymbol{Y}$ be a random vector of climate variables of interest that is Ω-measurable. The KL divergence between $F_{CL}(\boldsymbol{Y}, \alpha)$ and $F(\boldsymbol{Y}, \alpha)$ as a function of the composite or group

ECS α is given by

$$KL(\alpha) = E\left[\log F_{CL}\left(\boldsymbol{Y}, \alpha\right) - \log F\left(\boldsymbol{Y}, \alpha\right)\right] = \int \log \frac{\prod_{i=1}^{K} f_i(\boldsymbol{Y}, \alpha)^{\omega_i}}{F(\boldsymbol{Y}, \alpha)} F(\boldsymbol{Y}, \alpha) d\Omega(\boldsymbol{Y}),$$

where the expectation is taken under the true density $F(\boldsymbol{Y}, \alpha)$. It follows

$$\alpha_0 = \arg\min_{\alpha} KL(\alpha) = \arg\max_{\alpha} E\left[\log F_{CL}\left(\boldsymbol{Y}, \alpha\right)\right]. \tag{4}$$

Let $s(\alpha) = \frac{\partial}{\partial \alpha} \log F_{CL}(\boldsymbol{Y}, \alpha\}$ denote the score function. Under standard regularity conditions in White (1982), (4) gives us the asymptotic identifiability condition

$$E[s(\alpha)] \neq 0, \quad \forall \alpha \neq \alpha_0. \tag{5}$$

In finite samples, we have

$$\begin{aligned} \log F_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}, \alpha) &= \sum_{i=1}^{K} \omega_i \log f_i(\{\boldsymbol{y}_{it}\}_{t=1}^{T}, \alpha) \\ &= \sum_{t=1}^{T} \sum_{i=1}^{K} \omega_i \log f_i(\boldsymbol{y}_{it} | \sigma_i(\{\boldsymbol{y}_{is}\}_{s=1}^{t-1}), \alpha), \end{aligned}$$

where $\sigma_i(\{\boldsymbol{y}_{is}\}_{s=1}^{t-1})$, $i = 1, ..., K$, is the natural filtration generated by past data and $\sigma_i(\boldsymbol{y}_{i0}) = \emptyset$. Given the maximum CL estimator $\hat{\alpha} = \sup_{\alpha} \log F_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}, \alpha)$, under finite-sample identifiability we have

$$\frac{1}{KT} \sum_{i=1}^{K} \sum_{t=1}^{T} s_{it}(\alpha) \neq 0, \quad \forall \alpha \neq \hat{\alpha}. \tag{6}$$

where $s_{it}(\alpha) = K\omega_i \frac{\partial}{\partial \alpha} \log f_i(\boldsymbol{y}_{it} | \sigma_i(\{\boldsymbol{y}_{is}\}_{s=1}^{t-1}), \alpha)$, $t = 1, ..., T$, is the score of log likelihood contribution from the i-th subdensity. Given fixed weights and based on a uniform law of large numbers, $\frac{1}{KT} \sum_{i=1}^{K} \sum_{t=1}^{T} s_{it}(\alpha) \xrightarrow{p.} E[s(\alpha)]$ whenever α is an interior point of a compact set around $\hat{\alpha}$. Combining (5) and (6), we have $\hat{\alpha}_i \xrightarrow{p.} \alpha_0$.

Taking a first-order Taylor expansion of $\sqrt{KT} \sum_{i=1}^{K} \sum_{t=1}^{T} s_{it}(\hat{\alpha}) = 0$ around the pseudo

ECSα_0. It gives

$$\sqrt{KT}(\hat{\alpha} - \alpha_0) = -\left(\frac{1}{KT}H(\hat{\alpha})\right)^{-1} \frac{1}{\sqrt{KT}}s(\alpha_0) + o_p(1).$$

It follows that $\frac{1}{KT}H(\hat{\alpha}) \xrightarrow{p.} \mathcal{H}(\alpha_0)$ due to consistency with respect to the pseudo ECS, and $\frac{1}{\sqrt{KT}}s(\alpha_0) \overset{a.}{\sim} N(0, \mathcal{J}(\alpha_0))$ due to the Lindeberg-Feller theorem (White, 1982). We thus have

$$\sqrt{KT}(\hat{\alpha} - \alpha_0) \overset{a.}{\sim} N(0, \mathcal{J}(\alpha_0)\mathcal{H}(\alpha_0)^{-2}),$$

identical to (3) but with the asymptotic variability and sensitivity terms evaluated at the KL-divergence minimizer α_0 instead of the true α. This also implies that the finite-sample variance estimator in the previous section is still consistent but with respect to α_0.

2.3 Quasi-Bayesian inference with composite likelihood

Results from Sections 2.1 and 2.2 rely on fixed model weights, critical not only for inference on the composite ECS, but also for the numerical maximization of CL given in (1). As ω_i goes to zero, the i-th submodel becomes unidentified in the population, making the CL flat along $\boldsymbol{\beta}_i$. Besides convergence problems, zero model weights mean that we cannot compute some climatic quantities that are usually functions of α and $\boldsymbol{\beta}_i$, such as the transient climate response and characteristic time scales (Geoffroy et al., 2013b,a; Cummins et al., 2020; Jackson et al., 2022).

2.3.1 Composite posterior distribution

We consider quasi-Bayesian inference by estimating weights together with all other model parameters. Specifically, we use a Markov chain Monte Carlo (MCMC) sampler to approximate the CL posterior distribution

$$p(\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega} | \{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K) \propto p_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K | \alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega}) \pi(\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega}), \quad (7)$$

where $\boldsymbol{\omega}$ is the vector of weights and $\pi(\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega})$ is the prior distribution. In the above, $p_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K|\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega})$ is numerically the same as the CL in (1) but viewed as the density function of data that is evaluated at given vectors of parameters and weights. Its details are given in section 3.2. Because the CL is not a likelihood function, the CL posterior bears the standard quasi-Bayesian interpretation: under a flat prior, the constructed CL posterior has the same asymptotic coverage as a maximum CL estimator in total variation of moments norm (Chernozhukov and Hong, 2003; Tian et al., 2007).

2.3.2 Prior distribution

We specify the prior as

$$\pi(\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega}) = \pi(\boldsymbol{\omega})\pi(\alpha)\pi\left(\{\boldsymbol{\beta}_i\}_{i=1}^K|\alpha\right).$$

A flat prior for ECS, or $\pi(\alpha) \propto 1$, is employed to encapsulate our ignorance about the climate response. Section 4.1 considers alternative priors. The prior for other parameters is

$$\pi\left(\{\boldsymbol{\beta}_i\}_{i=1}^K|\alpha\right) = \prod_{i=1}^K \pi(\{\boldsymbol{\beta}_i\}|\alpha), \quad \pi(\{\boldsymbol{\beta}_i\}|\alpha) \propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_i)\right).$$

So conditional on α, $\boldsymbol{\beta}_i$ is $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$-distributed, $i = 1, ..., K$. Specifically,

$$\boldsymbol{\mu}_i = \arg\max_{\boldsymbol{\beta}_i} \log p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i), \quad \boldsymbol{\Sigma}_i = -\left[\frac{\partial^2}{\partial \boldsymbol{\beta}_i^2} \log p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i)\Big|_{\boldsymbol{\beta}_i=\boldsymbol{\mu}_i}\right]^{-1},$$

where $\log p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i)$ is the log likelihood of the i-th EBM, equivalent to $\log f(\{\boldsymbol{y}_{it}\}_{t=1}^T \in A_i, \alpha, \boldsymbol{\beta}_i)$ in (2). The conditional prior $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is a Laplace approximation of the model-specific log likelihood, given α. Seemingly *ad hoc*, this setup follows an empirical Bayes approach to imbue the prior with data information (Canova and Matthes, 2021). A notable benefit is evident when ω_i approaches zero, as the nuisance parameters are drawn from the asymptotic distribution of the conditional maximum likelihood estimator. Therefore, the empirical prior permits the identification of nuisance parameters from their EBM in case of zero model weights.

Also, the prior is not tight, because it is conditional on α which has a flat prior.

Lastly, the prior for weights $\pi(\boldsymbol{\omega})$ is a K-dimensional Dirichlet distribution with a common concentration parameter a, or $\pi(\boldsymbol{\omega}) \propto \prod_{i=1}^{K} \omega_i^{a-1}$. We choose $a = 0.5$, so *a priori*, ω_i equals $1/K$ in expectation with a standard deviation about $3/K$, which is non-informative.

2.3.3 MCMC algorithm

The MCMC algorithm iterates over three blocks: $p(\{\boldsymbol{\beta}_i\}_{i=1}^K|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \alpha, \boldsymbol{\omega})$, $p(\alpha|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \{\boldsymbol{\beta}_i\}_{i=1}^K$, and $p(\boldsymbol{\omega}|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K)$. The full CL posterior sample consists of every 5-th draw from 15,000 MCMC runs with the first 5,000 burn-in draws discard.

Sampling $\{\boldsymbol{\beta}_i\}_{i=1}^K|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \alpha, \boldsymbol{\omega}$

The model-specific parameters $\boldsymbol{\beta}_i$, $i = 1, ..., K$, are sampled in a Metropolis-Hastings step in parallel on a multi-core worker. The conditional CL posterior satisfies

$$\begin{aligned} p(\{\boldsymbol{\beta}_i\}_{i=1}^K|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \alpha, \boldsymbol{\omega}) &\propto p_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K|\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega})\pi(\{\boldsymbol{\beta}_i\}_{i=1}^K|\alpha) \\ &\propto \prod_{i=1}^{K} p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i)^{\omega_i}\pi(\boldsymbol{\beta}_i|\alpha). \end{aligned} \tag{8}$$

We sample a candidate $\boldsymbol{\beta}_i$ from a multivariate Student's t distribution $mSt(\boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*, v)$ with mean vector $\boldsymbol{\mu}_i^*$, covariance matrix $\boldsymbol{\Sigma}_i^*$, and a chosen degrees of freedom v. In our application, $\boldsymbol{\mu}_i^*$ is the vector of the model-specific maximum likelihood estimates for $\boldsymbol{\beta}_i$; $\boldsymbol{\Sigma}_i^*$ is the $\boldsymbol{\beta}_i$-subblock of the negative inverse hessian of the log likelihood evaluated at $\boldsymbol{\mu}_i^*$. We choose $v = 10$ to have large swings around the center of the proposal density. Notice, the proposal density is different from $\pi(\boldsymbol{\beta}_i|\boldsymbol{\alpha})$, as the former does not depend on α and is fixed throughout the MCMC runs. Let $\boldsymbol{\beta}_i^\bullet$ be the previous draw in the Markov chain. The candidate draw is accepted with probability

$$\min\left[1, \frac{p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i)^{\omega_i}\pi(\boldsymbol{\beta}_i|\alpha)mSt(\boldsymbol{\beta}_i^\bullet; \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*, v)}{p(\{\boldsymbol{y}_{it}\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i^\bullet)^{\omega_i}\pi(\boldsymbol{\beta}_i^\bullet|\alpha)mSt(\boldsymbol{\beta}_i; \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}_i^*, v)}\right].$$

Sampling $\boldsymbol{\alpha}|\{\{\boldsymbol{y}_{it}\}_{t=1}^T\}_{i=1}^K, \{\boldsymbol{\beta}_i\}_{i=1}^K, \boldsymbol{\omega}$

The composite ECS α is sampled in a Metropolis-Hastings step. A candidate α is drawn

from $mSt(\hat{\alpha}, \widehat{\text{Var}}(\hat{\alpha}), v)$, where $\hat{\alpha}$ and $\widehat{\text{Var}}(\hat{\alpha})$ are the conditional maximum CL estimator of the CL log likelihood and the associated variance, respectively; see Section (2.1). Alternatively, a random walk proposal can be used (Canova and Matthes, 2021). The candidate draw is accepted with probability

$$\min\left[1, \frac{p_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}|\alpha, \{\boldsymbol{\beta}_i\}_{i=1}^{K}, \boldsymbol{\omega}) mSt(\alpha^{\bullet}; \hat{\alpha}, \widehat{\text{Var}}(\hat{\alpha}), v)}{p_{CL}(\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}|\alpha^{\bullet}, \{\boldsymbol{\beta}_i\}_{i=1}^{K}, \boldsymbol{\omega}) mSt(\boldsymbol{\alpha}; \hat{\alpha}, \widehat{\text{Var}}(\hat{\alpha}), v)}\right].$$

Sampling $\boldsymbol{\omega}|\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}, \alpha, \{\boldsymbol{\beta}_i\}_{i=1}^{K}$

The weights are sampled via an importance sampling procedure. We draw $M = 100,000$ vectors of candidate $\boldsymbol{\omega}^{(m)}$, $m = 1, ..., M$, from the Dirichlet prior $\pi(\boldsymbol{\omega})$ in parallel. One draw from the conditional CL posterior $p(\boldsymbol{\omega}|\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}, \alpha, \{\boldsymbol{\beta}_i\}_{i=1}^{K})$ can be obtained by resampling once from the M candidates, with probability proportional to

$$\exp\left(\sum_{i=1}^{K} \omega_i^{(m)} \log p(\{\boldsymbol{y}_{it}\}_{t=1}^{T}|\alpha, \boldsymbol{\beta}_i) - (a-1)\sum_{i=1}^{K} \log \omega_i^{(m)}\right).$$

With a large M, the importance sampling procedure directly draws from the conditional posterior. Due to parallelization, it is also computationally efficient.

2.4 Relation to physical storylines

Stevens et al. (2016) among others have highlighted that the significance of quantifying ECS and its bounds in climate science is the alignment with diverse lines of scientific evidence. The current approach involves developing and scrutinizing physical storylines or hypotheses to refine these bounds. For example, various ranges of ECS reported by IPCC result from a thorough assessment of global scientific collaboration, evaluating physical and observational evidence (Intergovernmental Panel on Climate Change, 2021).

From climate simulations with observational constraints, Aldrin et al. (2012) and Otto et al. (2013) found a modest cooling effect of anthropogenic aerosols and the potential for negative cloud feedbacks, supporting a low ECS (<1.5K). Conversely, Boucher et al. (2013) and Stevens

(2015) argued against a low ECS, favoring a medium effect. This study, using a Bayesian approach, allows researchers to incorporate diverse beliefs into the prior distribution of model weights. By assigning different concentration parameters, say $\alpha_i > \alpha_j$, one can express $\pi(\boldsymbol{\omega}) \propto \prod_{i=1}^{K} \omega_i^{a_i - 1}$, reflecting the greater consistency of the i-th AOGCM with a specific storyline compared to the j-th climate model.

Stevens et al. (2016) proposed Bayesian averaging with storyline-driven priors, yet this method requires a common piece of "evidence", such as observational records, to assess marginal likelihood. This limitation makes it unsuitable for our scenario where the "evidence" consists of AOGCM-simulated datasets. Canova and Matthes (2021) demonstrated that when a common dataset is fitted by models with varying complexities, The CL posterior estimates of model weights closely align with those in Bayesian model averaging. Additionally, if one of the models accurately represents the DGP, the CL approach yields model weights converging to the true model.

The CL approach not only consolidates models from different datasets but also enables inference on model weights, a feature absent in Bayesian model averaging or the storyline approach. In the latter, weights are fixed *ex ante*, offering no *ex post* adjustments even when the "evidence" strongly supports or opposes a particular weighting scheme. In contrast, the CL approach estimates weights concurrently with other model parameters, allowing for dynamic updates based on the evidence strength reflected in the CL posterior density. The readily attainable credible bands of model weights can measure the uncertainty in pooling storylines, which is largely overlooked by the literature.

2.5 Relation to emergent constraints

In climate science, emergent constraints offer another common approach to constrain the value of ECS. These constraints arise from shared physical laws in various AOGCMs, implying consistent relationships in internal variability across models. Researchers can apply emergent constraints to simple models as in Caldwell et al. (2018) who impose a common relationship between ECS and temperature variability.

Although not explicitly imposing emergent constraints, our CL approach effectively constrains model-specific ECSs among candidate models. If AOGCMs share emergent constraints, the composite ECS must align with them. Additionally, the CL approach is adaptable in cases where constraining models is less straightforward. This includes scenarios where some models share emergent constraints, while others do not, or when certain AOGCMs prioritize specific feedbacks or resolutions.

Through the introduced sampling iteration in Section 2.3.3, we observe that ω_i tends to increase when $\boldsymbol{\beta}_i$ better fits the simulated dataset of the i-th AOGCM, indicating a more realistic feedback mechanism in terms of aggregate energy balance. Conversely, ω_i decreases when the model only partially aligns with the composite ECS, revealing instances where the AOGCM lacks consensus in climate response compared to other models. Thus, the weights in the CL approach represent models' proximity to the composite ECS, considering how well each EBM fits its corresponding AOGCM-simulated dataset.

3 Energy balance models as climate emulators

While AOGCMs explicitly describe the fluid dynamics of atmosphere and oceans, EBMs provide a simplified view on how changes in global mean surface temperature can be explained by radiative forcings that alter Earth's energy budget. This simplification makes an EBM a climate emulator that offers analytical tractability (Geoffroy et al., 2013b; Cummins et al., 2020; Jackson et al., 2022). This section introduces the setup of a 3-box EBM that we view as a physically consistent time series model.

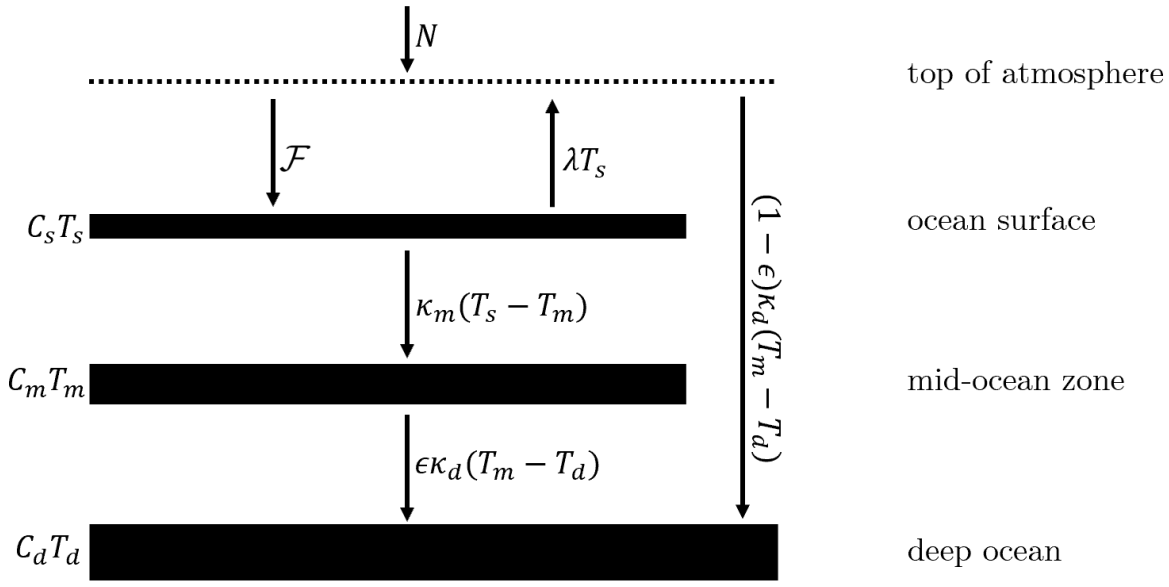

Figure 2: Vertical heat diffusion diagram in the 3-box energy balance model. The thickness of a box indicates the heat capacity of an ocean layer with associated thermal budget indicated on the left. The top of the atmosphere is represented by a dashed line as it has no heat capacity.

3.1 Model specification in continuous time

Following Cummins et al. (2020), the 3-box EBM considered in this study is characterized by the system of 3 linear differential equations:

$$C_s dT_s(t) = [\mathcal{F}(t) - \lambda T_s(t)]dt - \kappa_m[T_s(t) - T_m(t)]dt + \sigma_s dW_s(t), \tag{9}$$

$$C_m dT_m(t) = \kappa_m[T_s(t) - T_m(t)]dt - \kappa_d[T_m(t) - T_d(t)]dt, \tag{10}$$

$$C_d dT_d(t) = \kappa_d[T_m(t) - T_d(t)]dt, . \tag{11}$$

We divide the ocean into three vertical boxes—surface, middle, and deep, denoted by subscript s, m, and d, respectively. For $i \in \{s, m, d\}$, $T_i(t)$ denotes the instantaneous temperature of box i, with heat capacity C_i. κ_i's are heat transfer coefficients, capturing the exchange of energy between the atmosphere and the ocean surface and further down to the deep ocean. $W_s(t)$ in (9) is a Brownian motion scaled by σ_s, partly capturing the climatic internal variability modeled by an AOGCM. Also, $\mathcal{F}(t)$ is the CO_2 radiative forcing that heats up the surface box.

We present the exchange of energy in the EBM in Figure 2. The natural heat influx is

the TOA radiative forcing denoted by $N(t)$. It is straightforward to reduce or increase the number of ocean layers or boxes. The simplest 1-box model has been studied by Sellers (1969) and Hasselmann (1976). Geoffroy et al. (2013b) has derived analytical solutions for the 2-box EBM. When $k \to \infty$, a k-box EBM can approximate continuous vertical heat diffusion (see *e.g.*, Aldrin et al., 2012 and Bruns et al., 2020). We follow Cummins et al. (2020) who carried out extensive model comparisons via information criterion and concluded that with the CMIP5 model run that contains 150 years of climatic response after a quadrupled CO_2 forcing, the 3-box EBM is the best in terms of model fit and parsimony.

It is worth noting that Figure 2 suggests that the deep ocean receives $1-\epsilon$ and ϵ proportion of heat influx from the atmosphere and middle layer, respectively. This breaks up (11) into two parts, in line with Geoffroy et al. (2013b,a) and Cummins et al. (2020) who showed that an EBM can better capture the deep ocean heat uptake by incorporating the efficacy factor ϵ. Cummins et al. (2020) and Jackson et al. (2022) also proposed to use a colored noise process for the CO_2 forcing so as to reproduce the persistence of residual forcing generated by AOGCMs. We thus allow $\mathcal{F}_t$ to follow an Ornstein-Uhlenbeck process:

$$d\mathcal{F}(t) = \theta(F_{4\times CO_2} - \mathcal{F}(t))dt + \sigma_F dW_F(t), \tag{12}$$

where the innovation driving $\mathcal{F}(t)$ is a Brownian motion $W_F(t)$ (scaled by σ_F) that is independent of $W_s(t)$. We introduce two shocks, following Cummins et al. (2020) who showed that one shock alone cannot adequately capture internal variability generated by AOGCMs. The drift term $F_{4\times CO_2}$ elicits the fourfold step-jump of ambient CO_2. Lastly, the energy budget of the atmosphere in the EBM as presented in Figure 2 is subject to the TOA heat influx $N(t)$, CO_2 radiative forcing $-\mathcal{F}(t)$, deep ocean sink $-(1-\epsilon)\kappa_d[T_m(t) - T_d(t)]$, and the surface radiation $\lambda T_s(t)$. Putting together, we get

$$N(t) = \mathcal{F}(t) - \lambda T_s(t) + (1-\epsilon)\kappa_d[T_m(t) - T_d(t)]. \tag{13}$$

Equation (13) plays an important role in determining ECS. In equilibrium $T_s(t) = T_m(t) =$

$T_s(t) = T(t)$, this equation reduces to the "Geoffroy regression" $N(t) = \mathcal{F}(t) - \lambda T(t)$. It is clear that data on $N(t)$ is instrumental in identifying the feedback parameter λ and thus ECS. Cummins et al. (2020) also pointed out that only relying on temperature data likely leads to meaningless estimates, considering the relaxation time of Earth's climate system is hundreds or even thousands of years. While satellite observational records of TOA heat influx have become increasingly accessible, they still suffer from the lack of quality control and short time span (Pretis, 2020; Jackson et al., 2022; Bennedsen et al., 2023). Therefore, we use AOGCM-simulated $N(t)$ series as counter-factual observations, following the climate science literature.

As ECS is common across AOGCM-fitted EBMs, the specificities of AOGCMs are captured by other EBM-specific parameters. For instance, heterogeneity in $F_{4\times CO_2}$ reflects how AOGCMs maps the quadrupled CO_2 concentration to forcings. As in section 2, let α denote the ECS. From the energy budget of (9) or (13), we see that

$$\alpha = \frac{F_{4\times CO_2}/2}{\lambda}.$$

$F_{4\times CO_2}/2$ in the numerator is from the definition of ECS as the equilibrium response to the doubling, not the quadrupling of atmospheric CO_2 concentration. This replaces the feedback parameter λ in (9) and (13) with $F_{4\times CO_2}/(2\alpha)$. Thus, parameters in the i-th EBM include the common ECS α and model-specific $\boldsymbol{\beta}_i$ that consists of the CO_2 forcing $F_{4\times CO_2}$, heat transfer coefficients κ_m and κ_d, heat capacities C_s, C_m, and C_d, the efficacy coefficient ϵ, the continuous-time autoregressive coefficient θ, and two scaling parameters σ_s and σ_F. For the physical definitions and units of the model parameters, readers can refer to Cummins et al. (2020).

Equation (9)-(12) allows us to write the 3-box EBM as a continuous-time state space model. Its state transition forms a system of stochastic differential equations:

$$d\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t)dt + \theta F_{4\times \mathrm{CO}_2}\boldsymbol{e}_1 dt + \boldsymbol{R}d\boldsymbol{W}(t), \tag{14}$$

where $\boldsymbol{x}(t) = (\mathcal{F}(t), T_s(t), T_m(t), T_d(t))'$, $d\boldsymbol{W}(t) = (dW_s(t), dW_F(t), 0, 0)'$, and $\boldsymbol{e}_1 = (1, 0, 0, 0)'$.

System matrices are given by

$$\boldsymbol{A} = \begin{bmatrix} -\theta & 0 & 0 & 0 \\ \frac{1}{C_s} & -\frac{F_{4\times CO_2}/(2\alpha)+\kappa_m}{C_s} & \frac{\kappa_m}{C_s} & 0 \\ 0 & \frac{\kappa_m}{C_m} & -\frac{\kappa_m+\epsilon\kappa_d}{C_m} & \frac{\epsilon\kappa_d}{C_m} \\ 0 & 0 & \frac{\kappa_d}{C_d} & -\frac{\kappa_d}{C_d} \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} \sigma_s & 0 \\ 0 & \frac{\sigma_F}{C_s} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{15}$$

Let $\boldsymbol{y}(t) = (T_s(t), N(t))'$ denote the AOGCM-simulated counter-factual observations. Using (13), we have the measurement equation

$$\boldsymbol{y}(t) = \boldsymbol{Z}\boldsymbol{x}(t), \quad \boldsymbol{Z} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & -F_{4\times CO_2}/(2\alpha) & (1-\epsilon)\kappa_d & -(1-\epsilon)\kappa_d \end{bmatrix}. \tag{16}$$

3.2 State space representation in discrete time

Due to the uniformly sampled TOA heat influx and surface temperature from the CMIP5 model runs for 150 years, the stochastic differential equation (14) can be discretized exactly and lead to a linear and Gaussian state space model that enables parameter estimation (Durbin and Koopman, 2012, Chapter 4).

As the discretization scheme is standard, we omit it for brevity. For $t = 1, ..., T$ and $T = 150$, the discrete-time state space model is given by

$$\boldsymbol{y}_t = \boldsymbol{Z}\boldsymbol{x}_t, \tag{17}$$

$$\boldsymbol{x}_t = \boldsymbol{d} + \boldsymbol{T}\boldsymbol{x}_{t-1} + \boldsymbol{W}_t, \quad \boldsymbol{W}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}), \tag{18}$$

$$\boldsymbol{x}_0 = (F_{4\times CO_2}, 0, 0, 0)'. \tag{19}$$

The subscript t in the above denotes the state and the observation processes sampled annually.

System matrices in (18) are given by

$$\begin{aligned}\boldsymbol{T} &= \exp(\boldsymbol{A}),\\ \boldsymbol{d} &= \theta F_{4\times \mathrm{CO}_2}\boldsymbol{A}^{-1}(\exp(\boldsymbol{A}) - \boldsymbol{I}_4)\boldsymbol{e}_1,\\ \boldsymbol{Q} &= \int_{s=0}^{1} \exp(\boldsymbol{A})\boldsymbol{R}\boldsymbol{R}'\exp(\boldsymbol{A}')ds,\end{aligned}$$

where $\boldsymbol{I}_n$ denotes an n-dimensional identity matrix. The integral in the last equation above can be analytically computed using the matrix exponential formula in Van Loan (1978). Because the AOGCM simulations are recorded right after the quadrupling of CO_2 concentration when the climate system was in equilibrium, we use deterministic initialization (19).

The Kalman filter can be used to evaluate the likelihood of the state space model given by (17)-(19) (Durbin and Koopman, 2012; Cummins et al., 2020; Jackson et al., 2022). However, we can directly integrate out $\boldsymbol{x}_t$ and carry out more efficient computation, which is key to the parallel computation of $p(\{\boldsymbol{y}_{it}\}_{t=1}^{T}|\alpha, \boldsymbol{\beta}_i)$ in (8). Suppress the dependence on i. The state transition (18) can be stacked and written as

$$\boldsymbol{\Phi}\boldsymbol{X} = \boldsymbol{D} + \boldsymbol{\mathcal{W}}, \quad \boldsymbol{\mathcal{W}} \sim N(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{Q}),$$

where $\boldsymbol{X} = (\boldsymbol{x}_1', ..., \boldsymbol{x}_T')'$, $\boldsymbol{\mathcal{W}} = (\boldsymbol{W}_1', ..., \boldsymbol{W}_T')'$, and

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{I}_4 & 0 & 0 & \dots & 0 \\ -\boldsymbol{T} & \boldsymbol{I}_4 & 0 & \dots & 0 \\ 0 & -\boldsymbol{T} & \boldsymbol{I}_4 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\boldsymbol{T} & \boldsymbol{I}_4 \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} \boldsymbol{d} + \boldsymbol{T}\boldsymbol{x}_0 \\ \boldsymbol{d} \\ \vdots \\ \boldsymbol{d} \end{bmatrix}.$$

Importantly, $\boldsymbol{\Phi}$ is a sparse and band matrix whose inverse can be efficiently computed using the band matrix routine of linear complexity developed in Chan and Jeliazkov (2009). Similarly, we can also stack observations $\boldsymbol{Y} = (\boldsymbol{y}_1', ..., \boldsymbol{y}_T')'$ in (17) such that $\boldsymbol{Y} = (\boldsymbol{I}_T \otimes \boldsymbol{Z})\boldsymbol{X}$. Using

$\boldsymbol{X} = \boldsymbol{\phi}^{-1}\boldsymbol{D} + \boldsymbol{\phi}^{-1}\boldsymbol{W}$, we have $\boldsymbol{Y}|\alpha, \boldsymbol{\beta} \sim N(\boldsymbol{U}, \boldsymbol{\Sigma})$ with $\boldsymbol{U} = (\boldsymbol{I}_T \otimes \boldsymbol{Z})\boldsymbol{\phi}^{-1}\boldsymbol{D}$ and $\boldsymbol{\Sigma}$ satisfying

$$\text{vec}(\boldsymbol{\Sigma}) = (\boldsymbol{I}_T \otimes \boldsymbol{Z}) \otimes (\boldsymbol{I}_T \otimes \boldsymbol{Z})(\Phi \otimes \Phi)^{-1}\text{vec}(\boldsymbol{I}_T \otimes \boldsymbol{Q}).$$

$\Phi \otimes \Phi$ is also a sparse and band matrix such that $(\Phi \otimes \Phi)^{-1}\text{vec}(\boldsymbol{I}_T \otimes \boldsymbol{Q})$ can be computed with linear cost. $\boldsymbol{\Sigma}$ is a Toeplitz matrix, for which there exist various efficient inversion and Cholesky decomposition algorithms of quadratic complexity; see *e.g.* Martinsson et al. (2005). We can compute the conditional likelihood of the EBM via

$$p(\{\boldsymbol{y}_t\}_{t=1}^T|\alpha, \boldsymbol{\beta}_i) = \frac{1}{\sqrt{(2\pi)^{2T}|\boldsymbol{\Sigma}|}} \exp\left((\boldsymbol{Y} - \boldsymbol{U})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{U})\right), \tag{20}$$

which speeds up the parallelization step in our MCMC algorithm introduced in Section (2.3.3).

3.3 Physical quantities versus mathematical abstractions

The quasi-Bayesian inference demonstrated by Canova and Matthes (2021), combining multiple models through the CL, aligns with an adaptive learning perspective, relevant to our study. While it's reasonable for all EBMs to share a common ECS when constraining its range, other parameters in EBMs, serving as mathematical abstractions for consistency with AOGCMs, should not assume common values. Distinguishing parameters such as λ (feedback parameter), $F_{4\times\text{CO}_2}$ (forcing), or C_s (surface heat capacity) allows the CL approach to leverage the unique characteristics of each AOGCM, a key strength setting it apart from Bayesian averaging and finite mixture models.

While acknowledging some nuances, we propose viewing an AOGCM as a structural model and an EBM as a reduced-form model, concepts often used in economics. In this analogy, the ECS serves as a structural parameter, while other parameters represent reduced-form abstractions. Consider the standard consumption-saving model in economics as an example, where economists formulate a utility maximization problem for a consumer with quadratic preference. The resulting supermartingale consumption path mirrors an AOGCM, with differences across

models akin to subtleties in how economists specify income processes driving consumption. Empirically, economists fit reduced-form models to consumption data, with our CL approach advocating for a common autoregressive coefficient across models, acknowledging the super-martingale nature while allowing other parameters to be model-specific to capture diverse income processes.

While the structural vs. reduced-form model distinction aids economists in linking AOGCM and EBM models, differences persist. Structural economic models rely on simplifying assumptions for analytical ease, while AOGCMs are rooted in physics. In contrast, reduced-form models reparameterize without necessarily misspecifying the underlying structure, whereas EBMs misrepresent AOGCMs. For example, introducing shocks in (9) and (12) aims to match climatic internal variability, which are deterministic signals, not random noise, in an AOGCM. Forging a common language and collaboration in modeling between economists and climate scientists is essential. This study aims to be an early step in this joint endeavor.

4 Estimation results

4.1 The composite equilibrium climate sensitivity

The use of CL to pool information across multiple EBMs is expected to constrain the ECS. Figure 3 illustrates the hump-shaped posterior distributions of the composite ECS from 16 CMIP5 AOGCM-simulated datasets, 15 CMIP6 datasets, and 31 combined datasets. These distributions, tighter than the IPCC's "likely" range, provide a clear risk interpretation. The flat prior for ECS ensures the posterior distributions, indicative of the composite ECS, reflect the rich information contained in the data. Furthermore, the presence of a composite ECS also supports potential emergent constraints shared by AOGCMs.

Jonko et al. (2018) estimated a bivariate vector autoregressive model of $(T_{s,t}, N_t)'$ using Bayesian method. They considered a hierarchical prior for ECS, shrinking model-specific ECS posteriors towards a common one. As Cummins et al. (2020) pointed out, such a statistical approach lacks a sound physical foundation and involve many hard-to-interpret reduced-form

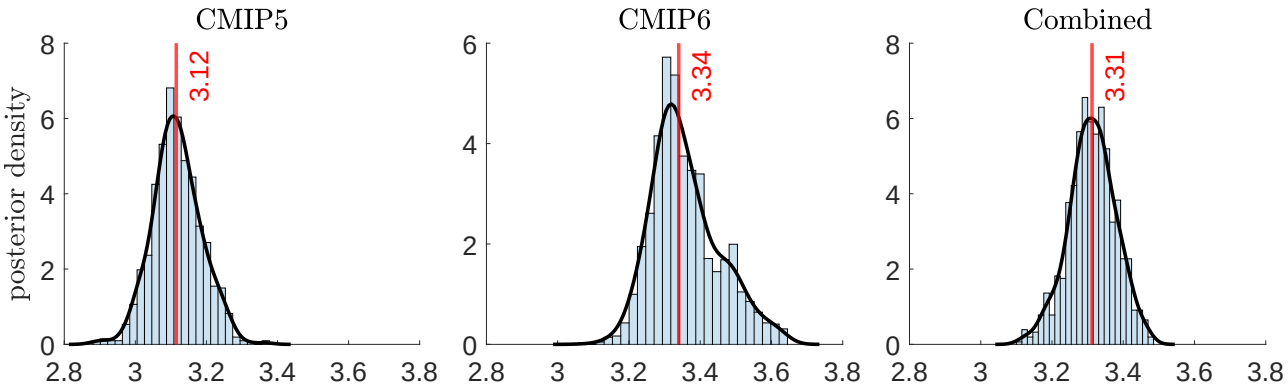

Figure 3: The composite posterior distributions of ECS. The posterior distribution of the ECS obtained from CMIP5 models is shown in the left panel, followed by that obtained from CMIP6 models (middle) and all models from CMIP5 and CMIP6 (right). Red vertical bars indicate the posterior medians. The ECS has a flat prior for all datasets.

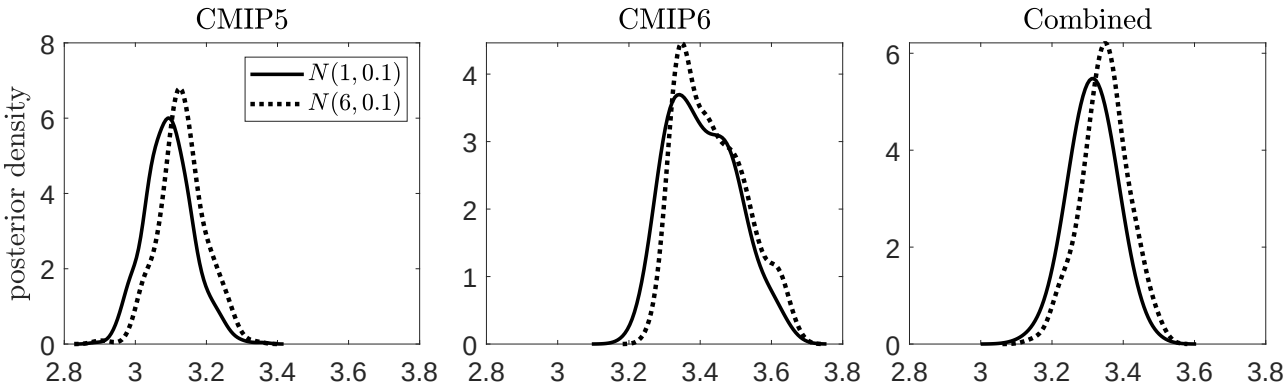

Figure 4: The composite posterior distributions of ECS under alternative priors. The posterior distribution of the ECS obtained from CMIP5 models is shown in the left panel, followed by that obtained from CMIP6 models (middle) and all models from CMIP5 and CMIP6 (right). The solid and the dashed line indicate the posterior obtained under informative priors $N(6, 0.1)$ and $N(1, 0.1)$, respectively.

parameters. We also notice that their hierarchical prior may impose strong restrictions on ECS: all datasets contribute equally to the posterior. Instead, CL downweights an AOGCM-simulated dataset if it is incompatible with the composite ECS; *e.g.*, due to disagreements with emergent constraints shared by other AOGCMs.

To further gauge prior sensitivity, we consider very informative priors $\pi(\alpha) = N(6, 0.1)$ and $\pi(\alpha) = N(1, 0.1)$. From the results shown in Figure 4, we see that the posterior distributions, in terms of the range and the shape of the density function, change little. Reassuringly, this implies that the AOGCM datasets contain dominating information over the strong priors. Although

Table 2: Composite posterior statistics of the ECS

Posterior statistics	mean	median	sta.dev.	95% CI	KS test
	Flat prior: $\pi(\alpha) \propto 1$				
CMIP5	3.117	3.115	0.069	$(2.988, 3.237)$	–
CMIP6	3.360	3.340	0.095	$(3.207, 3.579)$	0.021
Combined	3.312	3.312	0.066	$(3.176, 3.442)$	0.083
	Informative prior: $\pi(\alpha) = N(1, 0.1)$				
CMIP5	3.094	3.095	0.067	$(2.969, 3.241)$	–
CMIP6	3.406	3.392	0.093	$(3.267, 3.604)$	0.006
Combined	3.314	3.313	0.061	$(3.188, 3.425)$	0.032
	Informative prior: $\pi(\alpha) = N(6, 0.1)$				
CMIP5	3.132	3.130	0.067	$(3.011, 3.264)$	–
CMIP6	3.433	3.419	0.094	$(3.300, 3.627)$	0.027
Combined	3.348	3.349	0.066	$(3.212, 3.480)$	0.160

Reported are the posterior mean, median, standard deviation (sta.dev.), 95% credible interval (CI), and p-value of the Kolmogorov-Smirnov (KS) test for equivalent distributions. The null hypothesis of the KS test is that the posterior under CMIP6 or the combined datasets is the same as the one under CMIP5.

we only use 150 years of data, the presence of the TOA heat influx greatly removes the burden of the limited effective sample size of the surface temperature, considering the equilibration time of the oceans can be thousands of years (Cummins et al., 2020; Jackson et al., 2022).

Table 2 reports some posterior statistics of ECS. The posterior medians of ECS obtained from CMIP6 datasets are typically larger than those from CMIP5 by 0.5K. While the combined dataset leads to a 0.3K increase from CMIP5. This finding suggests that CMIP6 models consider a warmer future climate. Interestingly, it is also in line with the update from the "very likely" range of 1.5K-4.5K in older IPCC assessments to 2.0K-5.0K in the most recent report (Intergovernmental Panel on Climate Change, 2021). In the last column, we test if the posterior distribution of ECS under CMIP6 or the combined datasets is the same as that under CMIP5, or $H_0 : p(\alpha|\text{CMPI6}) = p(\alpha|\text{CMPI5})$ and $H_0 : p(\alpha|\text{Combined}) = p(\alpha|\text{CMPI5})$, via the Kolmogorov-Smirnov (KS) test. The KS test shows that under all prior settings, the data clearly distinguishes the CMPI6 ECS from the CMIP5 one, implying that there is a general improvement of AOGCMs, in terms of their agreement on a higher ECS.

Beyond the level shift, posterior standard deviations under CMIP6 exceed those under

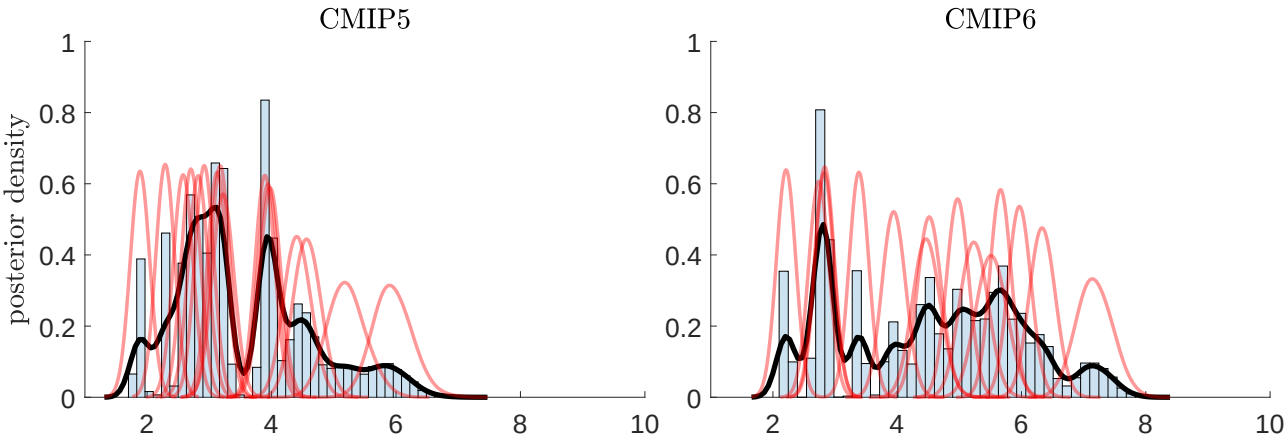

Figure 5: Individual and joint posterior distributions of ECS using CMIP5 and CMIP6 AOGCM-simulated datasets. The posterior distributions of the ECS obtained from individual models are shown by the red line. Their densities are divided by 4 to allow for easy comparison with the joint distribution indicated by the black line. The joint distribution is a mixture distribution of 16 (15) components under CMIP5 (CMIP6).

CMIP5, with both 95% credible intervals (CIs) notably tighter than the IPCC's "likely" range. The broader 95% CI of ECS under CMIP6 aligns with the wider range of model-specific ECSs reported in Figure 5. As discussed by Meehl et al. (2020) and Dietz et al. (2021), an improved understanding of physical processes suggest a higher likelihood of ECS at the upper end of the "likely" range. This is supported by our results which also show an increased uncertainty. Combining CMIP5 and CMIP6 models, ECS centers around 3.31K, bridging the medians of CMIP5 and CMIP6 values but leaning closer to the latter. This highlights that the CL approach does not fully disregard CMIP5 information, a valuable feature given the consistency of AOGCM models with emergent constraints that help narrow down the range of ECS. As the IPCC heavily relies on the latest CMIP models, our CL approach offers a scalable method to enhance inferential strength by incorporating older AOGCMs.

4.2 The individual equilibrium climate sensitivity

Estimated individually, the EBM fitted to 16 CMIP5 datasets leads to a bimodal joint posterior distribution of ECS, as shown in the left panel of Figure 5. Modalities are clearly due to the clustering of model-specific ECS posteriors—one centered around 3K and another centered

around 4K. Such clustering is less visible in the CMIP6 case, where we have a larger disagreement among models, expanding the uncertainty range of ECS especially towards the upper end of the multi-modal joint posterior. In either case, quantifying the plausible range of ECS seems to be a challenging task without deep understanding of physical storylines or evidences that can back up or rule out a certain AOGCM-implied ECS (Stevens et al., 2016; Knutti et al., 2017; Scafetta, 2022). In contrast, the CL approach targets the common ECS of interest while leaving other parameters model-specific to capture the specificities of the underlying AOGCMs. By striking a balance between model fit and the deviation to the composite ECS, our approach delivers hump-shaped posteriors and tight credible intervals. From a risk assessment viewpoint, a unimodal distribution of ECS can be instrumental to quantifying the risks associated with climate change and policies (Stern, 2007; Dietz et al., 2021).

Figure 6 displays boxplots illustrating posterior distributions of model weights for CMIP5 models, CMIP6 models, and their combination. The leftmost panel indicates that the dataset simulated from HadGEM2-ES receives zero weight, guided by the Dirichlet prior that, while not mechanically shrinking towards zero, allows for exclusion *a posteriori*. This prior design facilitates the exclusion of models when inferring a common parameter value. It is noteworthy that fitting an EBM to the HadGEM2-ES dataset doesn't imply poor model fit; rather, it is the individual estimate of ECS, shown in Table 3, deviating significantly from the composite ECS that incurs the weight penalty. Similar patterns are observed with the low weight of CESM2 which exhibits a climatic response running too hot. In general, the relative importance of models from CMIP5 and CMIP6 is largely maintained in the combined exercise. Models from the Beijing Climate Center, particularly BCC-CSM1.1 and its new version BCC-CSM2-MR which carry significant information in identifying the composite ECS in both CMIPs, emerge as winning models in the combined exercise. Interestingly, a clear bifurcation is evident between models with small and large weights in both CMIP5 and CMIP6 results. Understanding the different climatic feedback mechanisms modeled by these AOGCM groups and their impact on clustering (Boucher et al., 2013; Stevens, 2015) could enhance our understanding of how physical processes influence ECS. We reserve this topic for future research.

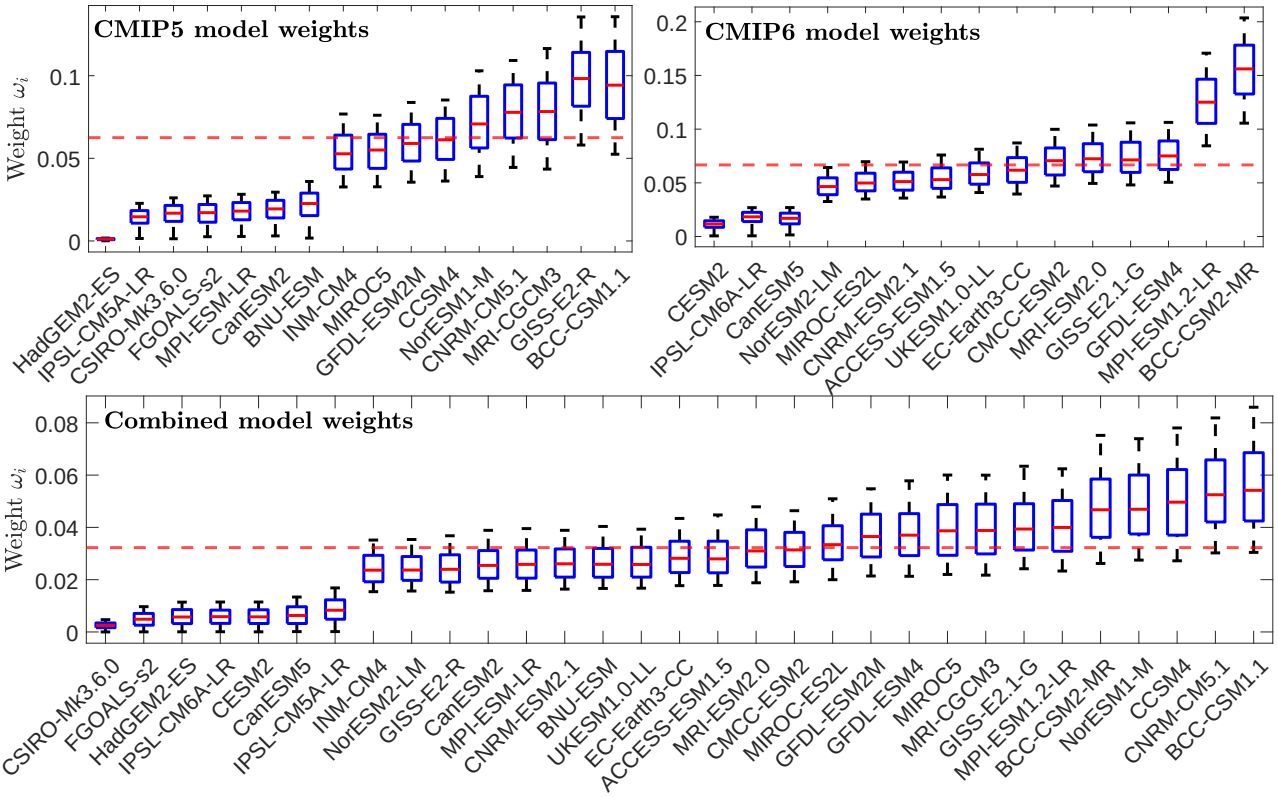

Figure 6: Posterior distributions of model weights using only CMIP5 models, only CMIP6 models, and all models. Each posterior distribution is indicated by the median, the interquartile range (bar), and the 95% credible interval (whisker). The red dashed lines indicate the prior mean of equal weights. Posterior distributions are sorted based on the 3-rd posterior quartile.

4.3 Comparison with naïve weightings

It is worth noticing that the unimodality of the posterior distribution of ECS is not an imposed result. The prior of ECS is a flat uniform distribution, leaving no restriction on the posterior shape. In fact, the posterior distribution under CMIP6, as shown in Figure 3 and 4, is less regular than the composite posterior under the CMIP5 and the combined datasets.

This raises the question that how the posterior distribution will look like if we consider different naïve weighting or pooling schemes. First, we consider the CL approach with fixed weights. We either consider equal weights or weights proportional to $1/\hat{\sigma}^2(\alpha_i)$—the inverse of the posterior variance of the ECS from the i-th model. The latter is motivated by the literature on forecast combination where the mean-squared-error-optimal combination of a set of unbiased forecast is the precision- or variance inverse-weighted forecast (Varin et al., 2011; Canova and Matthes, 2021). Second, we conduct Bayesian model averaging (BMA) where we weight the model-specific posteriors based on the posterior model probability, or

$$p(\alpha|\{\{\boldsymbol{y}_{it}\}_{t=1}^{T}\}_{i=1}^{K}) = \sum_{i=1}^{K} \omega_i p(\alpha|\{\boldsymbol{y}_{it}\}_{t=1}^{T}) \propto \sum_{i=1}^{K} \omega_i p(\{\boldsymbol{y}_{it}\}_{t=1}^{T}) \tag{21}$$

Two cases are considered: (1) we equally treat each model *a priori*, which leads to a posterior probability that is equal to the data likelihood (21); (2) we impose deterministic prior $\omega_i \propto 1/\hat{\sigma}^2(\alpha_i)$. When computing the data likelihood in (21), we evaluate the conditional likelihood (20) using Monte Carlo integration to average out the prior effect. Note that the joint posteriors shown in Figure 5 are BMA posteriors with a deterministic prior of equal weight. Lastly, we follow Cummins et al. (2020) and Jackson et al. (2022) who consider estimating ECS from the average of AOGCM-simulated datasets. This data averaging approach gives us a single dataset to which we fit an EBM.

Figure 7 presents posteriors under different weighting schemes. Fixed weights in both CL approaches tend to yield larger ECS than the composite ECS with estimated weights, resulting in an overestimation due to insufficient discounting of models that run too hot. Additionally, they produce less regular distributions with fatter tails compared to composite posteriors.

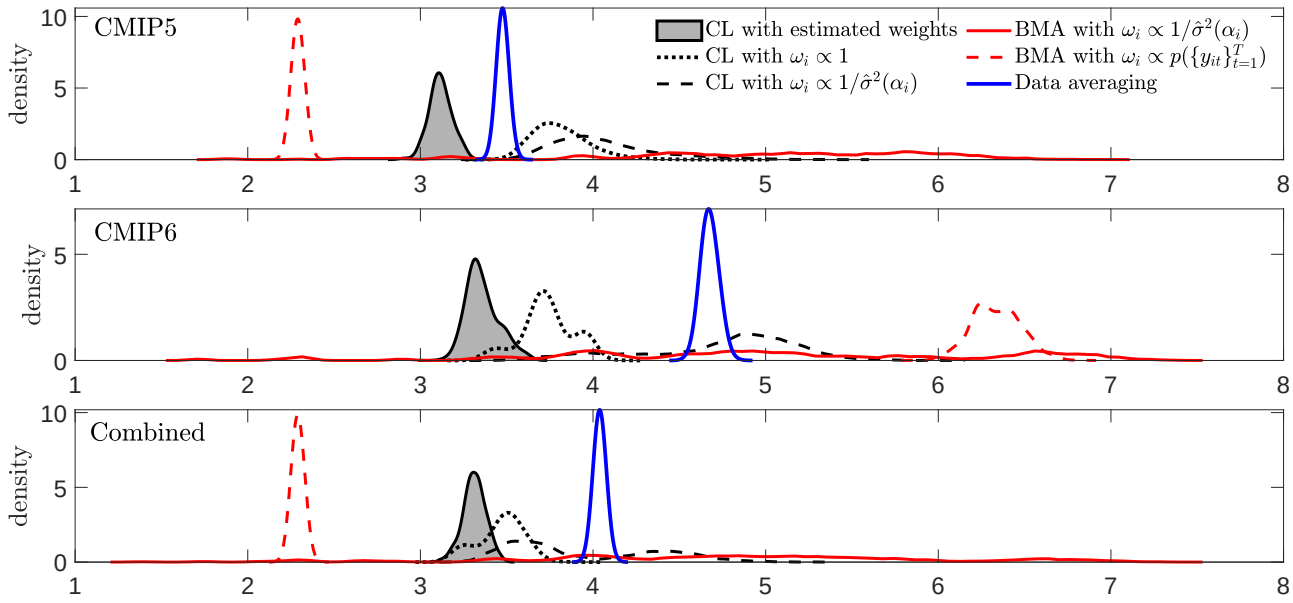

Figure 7: Posterior distribution of ECS under some naïve weighting schemes. Weighting and pooling schemes are with respect to CMIP5 models, CMIP6 models, and the combined models.

Weighting CL model components by precision $1/\hat{\sigma}^2(\alpha_i)$ (see Table 3) offers no new insight, displaying a significantly wider range than other CL approaches. In terms of BMA, Stevens et al. (2016) proposed to use BMA where the prior of model probability reflects physical storylines. However, comparing data likelihood of different datasets carries no probabilistic interpretation Canova and Matthes (2021), rendering BMA meaningless. We see that under $\omega_i \propto p(\{\boldsymbol{y}_{it}\}_{t=1}^T)$ (see Table 3), the CMIP5 and the combined models are dominated by GISS-E2-R, whereas the CMIP6 models are dominated by CMCC-ESM2 due to their near-unity weights after normalization. Additionally, BMA with weights inversely related to the posterior variance of ECS provides the least informative ECS range, lacking identifiable modes and exhibiting high posterior uncertainty, making ECS constraining less evident in all three weighted posteriors.

The data averaging approach yields a unimodal and tight posterior distribution under each model set. However, it's crucial to approach this result with caution, as the tight range is a mechanical consequence of averaging out climatic internal variability in AOGCM-simulated data series, as illustrated in Figure 1. Critics such as Lehner et al. (2020) and Scafetta (2022) argue against treating internal variability as random noise, emphasizing that climatic signals contain dynamic information and should not be smoothed out. Each of these simple weighting

or pooling schemes exhibits major drawbacks, making the proposed CL approach an appealing addition to the literature for narrowing down the ECS range. Further discussion on model averaging in the context of ECS estimation is provided in the next section.

4.4 Other posterior analysis

Combining model outputs in climate modeling is a contentious issue, distinct from assumption-based economic models. AOGCMs adhere to physical laws, making the average of their outputs inconsistent across all models due to the involvement of non-linear differential equations in their physical processes (Caldwell et al., 2018; Meehl et al., 2020). Consequently, the reported values of ECS, such as those by Intergovernmental Panel on Climate Change (2021), represent a range of "opportunities" or model outputs. Meaningful model combination methods must respect AOGCM specificities and should avoid blind pooling (Jonko et al., 2018). Therefore, BMA methods, demonstrated in the previous section, should not be used to constrain ECS, as weighting by $p(\{\boldsymbol{y}_{it}\}_{t=1}^{T})$ or $1/\hat{\sigma}^2(\alpha_i)$ lacks both physical and probabilistic interpretation. The magnitude of these quantities, as seen in Table 3, exhibits little relevance to the physics of AOGCMs. At best, they inform whether or not an EBM can be fitted to AOGCM simulations.

Although not reported, differences exist between the CL estimates of model parameters and model-specific estimates, indicating possible inconsistency between the composite model and individual AOGCMs. It's crucial to recognize that an EBM misspecifies an AOGCM. The CL approach constrains ECS by incorporating information from various AOGCM datasets while accommodating AOGCM specificities. Given the composite ECS, an estimated EBM remains optimal in terms of Kullback-Leibler minimization under maximum CL estimation. Essentially, an EBM with different parameterizations, nested in the CL, is not necessarily "more misspecified" than one fitted to an individual AOGCM dataset. Table 3 provides some statistics from model-specific estimates, including the transient climate response (TCR) and characteristic time scale $\tau(j)$, where $j = 1, 2, 3$.

TCR measures the annual surface temperature response to atmospheric CO_2 concentration increasing at a rate of 1% per year (doubling requires 70 years). Geoffroy et al. (2013a) and

Table 3: Posterior results of key parameters and data likelihood

Model	ECS–α_i	$10^3 \times \hat{\sigma}^2(\alpha_i)$	TCR_i	$\tau_i(1)$	$\tau_i(2)$	$\tau_i(3)$	$\log p(\{\boldsymbol{y}_{it}\}_{t=1}^{T})$
CMIP5 models							
BCC-CSM1.1	2.9	1.0	1.86	1.7	8.1	165	236
BNU-ESM	3.9	5.2	2.44	1.4	8.9	275	71
CanESM2	3.9	3.1	2.31	1.4	7.5	219	149
CNRM-CM5.1	3.2	1.0	2.14	0.9	8.7	262	230
CSIRO-Mk3.6.0	5.2	74.3	1.88	1.0	6.8	318	54
GFDL-ESM2M	2.6	3.0	1.54	0.9	5.6	265	119
GISS-E2-R	2.3	0.7	1.38	1.3	3.7	236	295
FGOALS-s2	4.6	26.8	2.30	1.1	5.6	396	35
INM-CM4	1.9	2.1	1.36	0.8	5.9	559	272
IPSL-CM5A-LR	4.4	25.5	2.18	0.8	13.2	401	133
MIROC5	2.8	3.1	1.78	1.3	7.7	330	54
HadGEM2-ES	5.9	80.2	2.40	1.0	8.2	531	196
MPI-ESM-LR	4.0	6.2	2.29	1.3	7.4	232	66
MRI-CGCM3	2.7	1.7	1.72	1.2	9.5	193	184
CCSM4	3.1	2.3	1.85	1.1	6.2	202	166
NorESM1-M	3.2	7.9	1.60	1.1	5.9	303	184
CMIP6 models							
ACCESS-ESM1.5	4.5	28.3	1.97	0.5	6.0	316	100
BCC-CSM2-MR	2.8	1.2	1.78	1.2	6.7	133	188
CanESM5	6.0	12.3	2.94	1.1	11.3	318	140
CESM2	7.2	64.3	2.47	0.5	5.1	391	135
CMCC-ESM2	6.3	20.8	2.41	1.0	7.2	477	214
CNRM-ESM2.1	5.2	28.5	2.42	0.7	8.7	537	101
EC-Earth3-CC	5.0	9.6	2.62	1.3	11.1	253	119
GFDL-ESM4	2.7	4.4	1.78	1.1	7.1	334	85
GISS-E2.1-G	2.8	2.5	1.75	0.9	5.4	378	110
IPSL-CM6A-LR	5.5	37.3	2.74	1.2	12.9	419	52
MIROC-ES2L	2.2	1.8	1.80	0.9	12.1	1939	14
MPI-ESM1.2-LR	3.4	2.4	2.00	1.2	5.3	298	140
MRI-ESM2.0	3.9	13.8	1.78	0.8	5.2	345	70
NorESM2-LM	4.5	16.8	2.24	1.0	5.0	293	79
UKESM1.0-LL	5.7	6.7	2.86	0.7	10.4	298	169

In the two left columns, the table reports the posterior median and variance (scaled by 10^3) of model-specific ECS. The posterior medians of model-specific TCR and three characteristic time scales are also reported. The last column give the marginal data likelihood of each model in logarithmic terms.

Knutti et al. (2017) among others argued that TCR can be a more policy-relevant measure of climate risk than ECS itself, as it summarizes the changes we are about to see in this century before the deep ocean fully equilibrates. The model-specific TCR is given by

$$TCR_i = \boldsymbol{e}_2' \frac{\log 1.01}{\log 2} ECS_i \left(70 - \boldsymbol{A}_i^{-1}(\exp(70\boldsymbol{A}_i) - \boldsymbol{I}_4)\mathbf{1}_4\right), \quad i = 1, ..., K,$$

where $\boldsymbol{e}_2 = (0, 1, 0, 0)'$ selects the second element (T_s) from the state vector $\boldsymbol{x}_t$ and $\mathbf{1}_4$ is a 4×1 vectors of ones; see *e.g.*, Geoffroy et al. (2013a) and Cummins et al. (2020) for details. With the CL approach, ECS_i in the above is replaced with the composite ECS. $-1/\tau(j)$ is the j-th eigenvalue of $\boldsymbol{A}$ in (15), so that $\tau(j)$ is the j-th characteristic time scale of the linear system that describes the time it takes for the response to a disturbance of each ocean layer to decay to $1/e$ (about 37%) of its initial value, a concept of relaxation time similar to half life.

In Figure 8 and 9, we compare the model-specific posterior estimates of TCR and $\tau(3)$ with those obtained under the CL approach applied to all 31 AOGCMs (UKESM1.0-LL is omitted in the figures for clarity). It is interesting to see that the posterior distributions of those physical quantities as functions of EBM parameters are very similar. While parameter estimates differ, the dynamics of state transition implied by the TCR and the characteristic time scales under the CL approach closely mimics each individual EBM, especially in the case of $\tau(3)$, or the relaxation time of the slowest mode that differs greatly among AOGCMs. This result shows that with enough room for AOGCM specificities, CL estimates are consistent with the vertical heat diffusion in the ocean as modeled by each AOGCM. It is this specificity-preserving property of the CL approach that makes it very suitable for constraining ECS.

Finally, Canova and Matthes (2021) demonstrated the possibility of weighting posteriors from the CL approach similarly to BMA, with CL weights replacing the posterior model probability. For instance, if the interest lies in a composite TCR, one can weigh the CL posteriors of TCR_i, where $i = 1, ..., K$, using estimated ω_i. However, this approach is somewhat *ad hoc* and lacks clear interpretability. While it is more natural to constrain an equilibrium quantity such as the ECS, if the goal is to constrain TCR, it is recommended to conduct the CL exercise treating

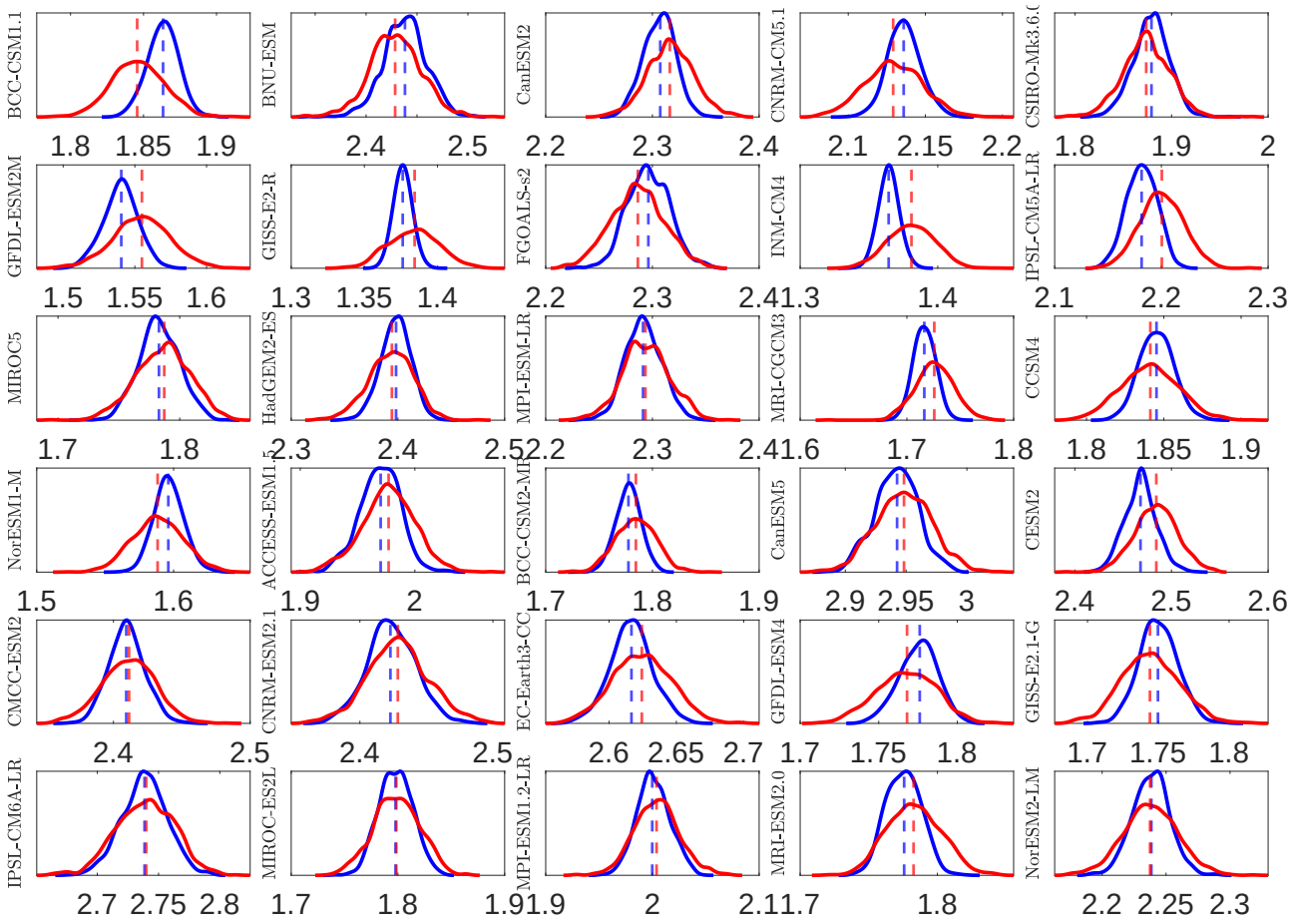

Figure 8: Posterior distributions of model-specific TCRs in comparison to posteriors obtained from the CL approach. The blue (red) line indicates the model-specific (CL) posterior. Dashed lines are posterior medians

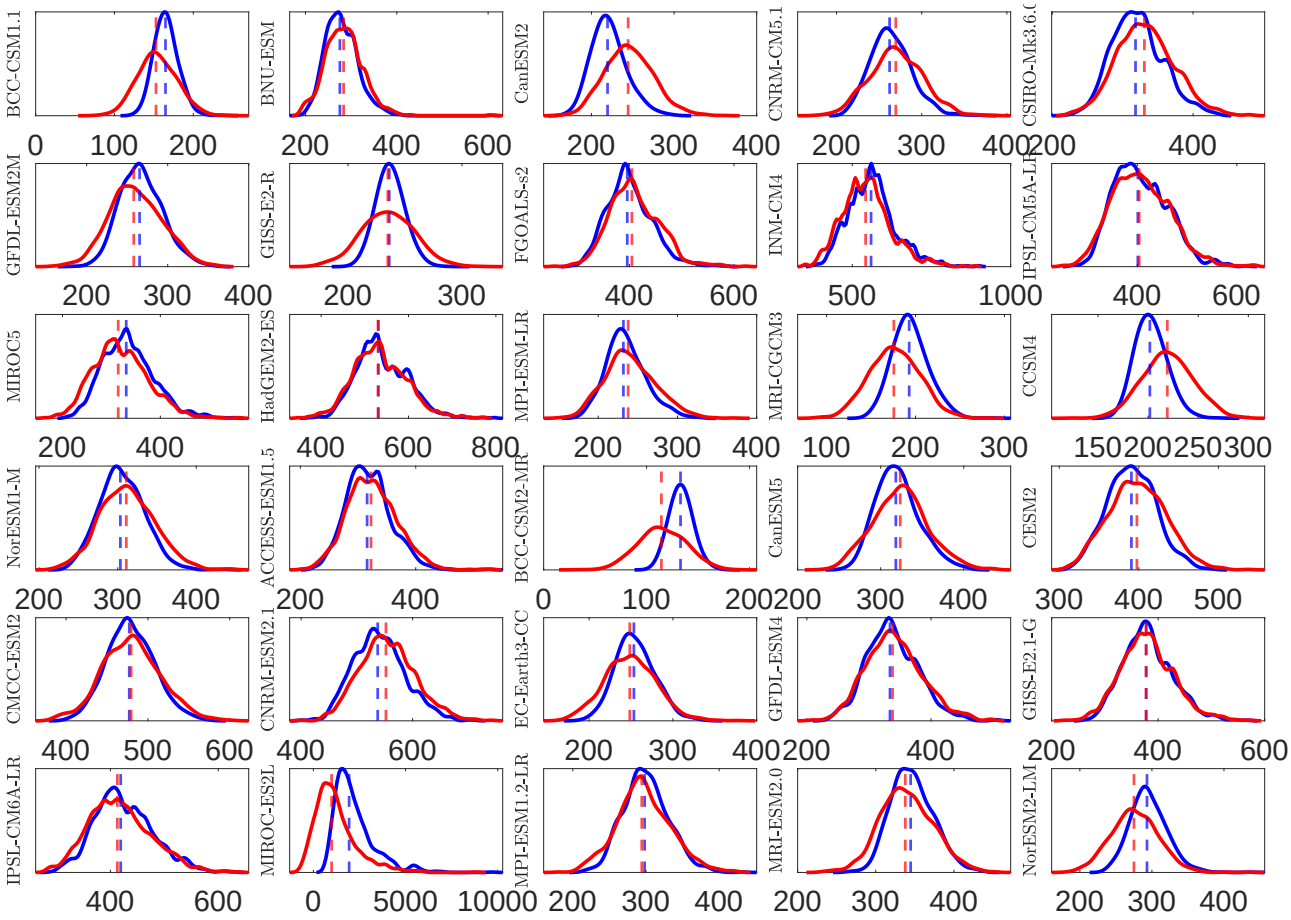

Figure 9: Posterior distributions of model-specific $\tau(3)$'s in comparison to posteriors obtained from the CL approach. The blue (red) line indicates the model-specific (CL) posterior. Dashed lines are posterior medians

TCR as a common parameter across EBMs while allowing ECS to vary across models. Imposing many common parameters contradicts the specificity-preserving property, so it is advisable to impose commonality only for the parameter of interest in practical applications.

5 Conclusion

Climate scientists and economists recognize the crucial need to narrow down the equilibrium climate sensitivity (ECS) range for effective policy decision-making, as demonstrated by the latest assessment report from Intergovernmental Panel on Climate Change (2021). Despite their efforts, the current range of (2K, 5K) remains too broad for informed policymaking, leading to substantial social costs under policy uncertainty (Stern, 2007; Dietz et al., 2021). In this study, we propose a composite likelihood estimator for ECS using Bayesian inference applied to a 3-box energy balance model. Our approach acknowledges the misspecification of the model when fitted to data of a quadrupled CO_2 experiment from 31 climate models, maintaining specificity for each model. By concentrating the distribution of ECS using information from all models, we provide a composite estimate with intuitive probabilistic interpretation. Through Bayesian methods, we estimate model weights balancing fit and distance to the constrained climate sensitivity, ensuring compatibility with potential emergent constraints among climate models, linking our econometric method to the climate science literature. Our findings suggest an ECS of 3.3K with a 95% credible interval of (3.18K, 3.44K), much tighter than the reported "likely range" by IPCC. The integration of econometric techniques and climate science positions the composite likelihood framework as a promising avenue for future research in data-driven methodologies for constraining parameters and projections in climate analytics.

References

Aldrin, M., Holden, M., Guttorp, P., Skeie, R. B., Myhre, G., and Berntsen, T. K. (2012). Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, 23(3):253–271.

Bennedsen, M., Hillebrand, E., and Zhou Lykke, J. (2023). Global temperature projections from a statistical energy balance model using multiple sources of historical data. *Journal of Climate*, pages 1–48.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., et al. (2013). Clouds and aerosols. In *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*, pages 571–657. Cambridge University Press.

Bruns, S. B., Csereklyei, Z., and Stern, D. I. (2020). A multicointegration model of global climate change. *Journal of Econometrics*, 214(1):175–197.

Caldwell, P. M., Zelinka, M. D., and Klein, S. A. (2018). Evaluating emergent constraints on equilibrium climate sensitivity. *Journal of Climate*, 31(10):3921–3942.

Canova, F. and Matthes, C. (2021). A composite likelihood approach for dynamic structural models. *The Economic Journal*, 131(638):2447–2477.

Chan, J. C., Eisenstat, E., Hou, C., and Koop, G. (2020). Composite likelihood methods for large bayesian VARs with stochastic volatility. *Journal of Applied Econometrics*, 35(6):692–711.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of econometrics*, 115(2):293–346.

Clarke, L. and Coauthors (2014). Assessing transformation pathways. *Climate Change 2014: Mitigation of Climate Change*, pages Edited by O. Edenhofer et al., 413–510.

Cox, P. M., Huntingford, C., and Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, 553(7688):319–322.

Cummins, D. P., Stephenson, D. B., and Stott, P. A. (2020). Optimal estimation of stochastic energy balance model parameters. *Journal of Climate*, 33(18):7909–7926.

Cummins, D. P., Stephenson, D. B., and Stott, P. A. (2022). Could detection and attribution of climate change trends be spurious regression? *Climate Dynamics*, 59(9-10):2785–2799.

Dietz, S., van der Ploeg, F., Rezai, A., and Venmans, F. (2021). Are economists getting climate dynamics right and does it matter? *Journal of the Association of Environmental and Resource Economists*, 8(5):895–921.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38.

OUP Oxford.

Engle, R. and Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, 30(2):212–228.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D., and Tytéca, S. (2013a). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, 26(6):1859–1876.

Geoffroy, O., Saint-Martin, D., Olivié, D. J., Voldoire, A., Bellon, G., and Tytéca, S. (2013b). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *Journal of Climate*, 26(6):1841–1857.

Hahn, R. W. and Ritz, R. A. (2015). Does the social cost of carbon matter? Evidence from US policy. *The Journal of Legal Studies*, 44(1):229–248.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Hasselmann, K. (1976). Stochastic climate models part I. Theory. *tellus*, 28(6):473–485.

Intergovernmental Panel on Climate Change (2021). *Working group I contribution to the Sixth Assessment Report (AR6), climate change 2021: The physical science basis*. Retrieved from: https://www.ipcc.ch/assessment-report/ar6/.

Jackson, L. S., Maycock, A. C., Andrews, T., Fredriksen, H.-B., Smith, C. J., and Forster, P. (2022). Errors in simple climate model emulations of past and future global temperature change. *Geophysical Research Letters*, 49(15):e2022GL098808.

Jonko, A., Urban, N. M., and Nadiga, B. (2018). Towards Bayesian hierarchical inference of equilibrium climate sensitivity from a combination of CMIP5 climate models and observational data. *Climatic Change*, 149:247–260.

Knutti, R., Rugenstein, M. A., and Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, 10(10):727–736.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R., and Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, 11(2):491–508.

Lewis, N. and Curry, J. (2018). The impact of recent forcing and ocean heat uptake data on estimates of climate sensitivity. *Journal of Climate*, 31(15):6051–6071.

Martinsson, P.-G., Rokhlin, V., and Tygert, M. (2005). A fast algorithm for the inversion of general Toeplitz matrices. *Computers & Mathematics with Applications*, 50(5-6):741–752.

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual review of statistics and its application*, 6:355–378.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, 6(26):eaba1981.

Otto, A., Otto, F. E., Boucher, O., Church, J., Hegerl, G., Forster, P. M., Gillett, N. P., Gregory, J., Johnson, G. C., Knutti, R., et al. (2013). Energy budget constraints on climate response. *Nature Geoscience*, 6(6):415–416.

Pretis, F. (2020). Econometric modelling of climate systems: The equivalence of energy balance models and cointegrated vector autoregressions. *Journal of Econometrics*, 214(1):256–273.

Raftery, A. E., Madigan, D., and Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191.

Scafetta, N. (2022). Advanced testing of low, medium, and high ECS CMIP6 GCM simulations versus ERA5-T2m. *Geophysical Research Letters*, 49(6):e2022GL097716.

Sellers, W. D. (1969). A global climatic model based on the energy balance of the earth-atmosphere system. *Journal of Applied Meteorology and Climatology*, 8(3):392–400.

Stern, N. H. (2007). *The economics of climate change: the Stern review*. cambridge University press.

Stevens, B. (2015). Rethinking the lower bound on aerosol radiative forcing. *Journal of Climate*, 28(12):4794–4819.

Stevens, B., Sherwood, S. C., Bony, S., and Webb, M. J. (2016). Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, 4(11):512–522.

Tian, L., Liu, J. S., and Wei, L. (2007). Implementation of estimating function-based inference procedures with Markov chain Monte Carlo samplers. *Journal of the American Statistical Association*, 102(479):881–888.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE transactions on automatic control*, 23(3):395–404.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

Watkiss, P. and Hope, C. (2011). Using the social cost of carbon in regulatory deliberations. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):886–901.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25.