# Predicting tail risks and the evolution of temperatures

Anthoulla Phella [a], Vasco J. Gabriel [b,*], Luis F. Martins [c,d]

[a] *University of Glasgow, United Kingdom*
[b] *University of Victoria and NIPE, Canada*
[c] *Instituto Universitário de Lisboa (ISCTE-IUL and BRU-IUL), Portugal*
[d] *CIMS, United Kingdom*

## ARTICLE INFO

## ABSTRACT

This paper explores a range of simple models to study the relationship between global temperature anomalies and climate forcings. In particular, we consider quantile regression models with potentially time-varying parameters (TVP), implemented by Bayesian methods. In its most general specification, this approach is flexible in that it models distinct regions of distribution of global temperature anomalies, while also allowing us to investigate changes in the relationship between (natural and anthropogenic) climate forcings and temperatures. Our results indicate that there is indeed considerable variation over time in the relationship between temperatures and its drivers, and that these effects may be heterogeneous across different quantiles. We then perform a long-range forecasting exercise for temperatures, which suggests that incorporating TVP or explicitly modelling quantile levels or the combination of both features can improve prediction for different parts of the temperature distribution. In addition, we produce forecasts for 2030 considering the intermediate RCP 4.5 scenario: given that no single specification dominates, we account for model uncertainty by considering forecast averaging across all specifications. Our approach allows us to make statements about the probability of temperature levels — for instance, we find that a scenario of $+1.8$ °C will occur with a non-negligible probability under RCP 4.5.

## 1. Introduction

Predicting the evolution of temperatures over the next decades is a crucial component in devising strategies to mitigate climate change. It is therefore of primary importance to develop models that are useful for forecasting not only mean temperatures, but also realizations originating from the tails of the distribution of temperatures. Moreover, it is critical to capture changes in the relationship between climate forcings, in particular greenhouse gases (GHG), and temperatures, as these are likely to determine the degree of urgency of different policy interventions (see Tol, 2005).

This paper contributes to this debate by illustrating the usefulness of a range of models that have the ability to capture variation over time and across different regions of the distribution of global temperature anomalies, and thus have the potential to improve the prediction of "risky" temperatures. As a benchmark specification, we use the distributed lag model of Castruccio et al. (2014), in which temperatures are expressed as a function of current and lagged $CO_2$ levels, and explore how it can be extended along a number of dimensions, namely by introducing additional forcings and allowing their effects to vary over time and across quantiles of the distribution of temperatures.

Our approach is an eminently statistical/econometric one: we assume our model imperfectly captures the dynamics of global temperatures, we use data to fit the parameters of our model according to some loss function and, having estimated the model, we construct data-based, "single-run" forecasts for the variable of interest. This contrasts with the standard practice in climate science, in which models represent the physical understanding of the climate system, which are then simulated to produce projections. Roughly put, global uncertainty arises from both model and parameter uncertainties, but also from initial conditions used in the simulated runs, which may lead to dramatically distinct projections. Hence, models are run several times and the collection ("ensemble") of these runs is taken to represent the statistical properties of the climate system.

Our paper is thus related to the literature on climate sensitivity and transient climate response to cumulative $CO_2$ emissions (TCRE), as it directly relates (linearly) $CO_2$ emissions and changes in global temperatures, which suggests the crucial role emissions have for the development of mitigation policies (see Matthews et al., 2009 and, more recently, Matthews et al., 2018, for example). Several papers try to quantify the uncertainty around the timing and magnitude of

---

* Corresponding author.
*E-mail address:* vgabriel@uvic.ca (V.J. Gabriel).

warming (see Ricke and Caldeira, 2014 or, more recently, Spafford and MacDougall, 2020), with a few papers using observational data to "constrain" those uncertainties (Matthews et al., 2009, Gillett et al., 2013. Our approach is largely data-driven, noting that while our "reduced-form" specifications contain the main ingredients of climate forcings, they lack the layers of theoretical structure that underpin climate models. However, to the best of our knowledge, this literature has not considered time-varying effects of climate forcings across different quantiles of the distribution of temperatures.

Our paper is further grounded on a substantial body of the climate econometrics literature focusing on the analysis of the statistical features of historical time series of temperatures, namely whether or not these contain stochastic trends or are best described as evolving along broken (deterministic) trends (see Kaufmann et al., 2013; Estrada and Perron, 2017 or, more recently, Chang et al., 2020 among many others). The correct determination of the time series properties of global temperatures is key to detect and explain the causes of climate change through appropriate (formal) statistical procedures. Furthermore, there is substantial evidence that the relationship between temperatures and radiative forcing, as well as GHG, may have changed over time. An example of this is the "hiatus" in global warming, a period between 1998 and 2013, in which the warming trend dips while GHG levels continued to increase. Several papers offer a variety of explanations for this phenomenon, see Pretis et al. (2015) and Miller and Nam (2020), for example.

The focus of much of the previous literature has been on the evolution of average global temperatures and, occasionally, their volatility. In this paper, we propose to analyse the evolution of the conditional distribution of global temperatures by specifying quantile regression models with time-varying parameters (TVP-QR) as in Korobilis et al. (2021). This approach is flexible in that it allows us to investigate potential differences in the effects of climate forcings on temperatures across different regions of the distribution, with a particular focus on "tail" temperatures. Moreover, by adopting a TVP framework, we can capture the changes over time in the relationship between global temperatures and their distinct forcings that have been widely documented in the literature.

Estimation of the TVP-QR models is implemented via Bayesian methods: we make use of the efficient Gibbs sampler developed by Korobilis et al. (2021), which overcomes several challenges in implementing such a richly parameterized framework. The aforementioned work, borrows ideas from Korobilis (2021) and transforms the TVP quantile regression into an equivalent high-dimensional regression form, resulting in a computationally simple procedure whereby the TVP-QR model is estimated over a fine grid of quantiles.

Furthermore, generating forecasts in this setup is quite straightforward. Little is known about whether temperature forecasts can be further improved by bringing together the benefits of TVPs with the flexibility of the QR setting. Therefore, we perform a long-range forecasting exercise for temperatures, which suggests that quantile regressions and/or incorporating TVPs lead to good prediction performance, specifically for downside and upside risk for temperature anomalies.

An additional contribution of our work is to propose a forecast averaging approach, whereby we use the predictive likelihood (PL, see Geweke and Amisano, 2010) from the previous forecasting exercise as a criterion to generate weights to combine forecasts from our suite of models. An averaging approach, by taking into account model uncertainty, allows us to produce robust forecasts. Moreover, using the PL measure, which captures the whole distribution, is particularly appropriate in our framework, given that we are interested in forecasting temperatures for a range of different quantiles. In particular, we construct conditional forecasts under the intermediate stabilization scenario given by the Representative Concentration Pathway (RCP) 4.5.[1]

Indeed, both the predictive distributions under the current emission realizations and those imposed by the alternative RCP 4.5 scenario, show that reaching temperatures higher than 1.8 °C will occur with probability greater than 70%.

Note that our approach is distinct from recent studies that consider the spatial heterogeneity of temperatures. Recently, Chang et al. (2020) focus on the time series characteristics of temperature anomaly distributions, concluding that these contain stochastic trends that differ over hemispheres and that the probability of experiencing positive anomalies has increased, consistent with anthropogenically-driven climate change. On the other hand, Gadea Rivas and Gonzalo (2020) study the distributional characteristics of temperatures (such as different moments and quantiles) over time. They detect a trend along both the time series and cross-section dimensions, particularly for lower quantiles, which is also confirmed by a decrease in dispersion, with lower temperatures approaching the median faster than higher temperatures.[2]

Our goal is different: we treat global temperatures as a "target" indicator (in the same way as GDP or inflation rates summarize information about aggregate economic activity) and we focus on the risk of extreme realizations of temperatures, emanating from the tails of its distribution. Therefore, our setup allows us to highlight the importance of asymmetric behaviour in global temperatures, by allowing their determinants to vary across different quantiles and over time. Indeed, we document the increasing impact of several anthropogenic forcings on temperatures: for instance, cumulative $CO_2$ emissions have an asymmetric and stronger impact on the tails of temperature anomalies, while also having a growing effect through time. Meanwhile the magnitude of the impact of greenhouse gases has increased more substantially in the lower tail, rather than in the upper tail. These findings are consistent with recent results from climate models, suggesting that forcings other than $CO_2$ may be driving extreme weather events (e.g., Wang et al., 2023 and Larson and Portmann, 2019). An implication is that reductions in anthropogenic forcings have the potential not only to stymie climate change, but also prevent the persistent (re)occurrence of extreme temperatures.

Our work also complements that of Agliardi et al. (2019), who study the relationship between GHG emissions and global warming with multi-level rolling techniques and model tail events by means of copulas. While they do not find an acceleration of the effects of emissions on temperature anomalies, they observe positive upper tail dependence, i.e., high probabilities of joint extreme large values for temperatures and emission concentrations.

The paper is organized as follows. Section 2 provides an overview of the data and a brief description of our benchmark model and the TVP-QR methodology. In Section 3 we carry out a forecasting comparison exercise using a range of models nested within the TVP-QR framework and proceed to study the implications of time variation across the full distribution. Section 4 concludes.

## 2. Methodology

### 2.1. Data description

Data for global temperatures (anomalies relative to the 1986–2005 base period) comes from Berkeley Earth, which averages raw gridded temperatures and bias-corrected station data.[3] We experiment with

---

[1] RCPs provide consistent projections for radiative forcings to produce alternative warming scenarios — RCP 4.5 is an intermediate scenario, with

"medium levels" of effort to curb emissions, leading to an average increased warming of 1.8 °C relative to 1986–2005.

[2] See also Ballester et al. (2010), Donat and Alexander (2012), Castruccio and Stein (2013), Castruccio et al. (2014) and Leeds et al. (2015) for other spatio-temporal approaches to climate modelling.

[3] We follow the convention in the literature of working with anomalies relative to a base period, which casts into sharp relief ongoing increases in temperatures in recent decades.

**Table 1**
Climate variables.

| Variable | Full description | Unit | Source |
|---|---|---|---|
| Temp | Global temperature anomalies | °C | Berkeley Earth |
| $CO_2$ | Carbon dioxide emissions | ppm | Global Carbon Budget |
| $GHG$ | Greenhouse gases | W/m2 | NASA's GISS |
| $WMGHG$ | Well-mixed greenhouse gases | W/m2 | NASA's GISS |
| $CH_4$ | Methane emissions | ppb | NASA's GISS |
| $N_2O$ | Nitrous Oxide | ppb | NASA's GISS |
| $AN$ | Aero Naturals | W/m2 | NASA's |
| $AS$ | Aerosols | W/m2 | NASA's |
| $AMO$ | Atlantic Multidecadal Oscillation | Index | NOAA |
| $Solar$ | Solar Irradiance | No. of sunspots | Royal Observatory of Belgium |

Notes: °C denotes degrees Celsius, ppp is parts per million, ppb is parts per billion, $W/m^2$ is watts per square metre, NASA is the National Aeronautics and Space Agency, GISS stands for Goddard Institute for Space Studies, NOAA is the National Oceanic and Atmospheric Administration.

both land only temperatures (which runs from 1750 to present), as well as averaged land and HadSST (Hadley Centre Sea Surface Temperature dataset) ocean global temperatures, starting in 1850.[4] As for forcing variables, we consider both annual aggregate and disaggregated GHG emissions concentration data: in addition to carbon dioxide ($CO_2$, ppm) present in our baseline specification, we also take into account the role of methane ($CH_4$, ppb), nitrous oxide ($N_2O$, ppb), as well as other well-mixed greenhouse gases ($WMGHG$, such as CFCs, HFCs and PFCs) – see the CMIP6 (Coupled Model Intercomparisons Project) climate model simulations of Miller et al. (2021) for sources and further details. We also consider the role of anthropogenic and natural aerosols (as used by Miller et al., 2021 in their CMIP6 simulations and sourced from NASA's GISS), as well as Solar Radiance (sunspot numbers from the Royal Observatory of Belgium) and the Atlantic Multidecadal Oscillation (AMO, an index based on the Kaplan Extended SST v2 dataset, available from NOAA).

Although data for several series is available from the 1750s, our sample period starts in 1856, the earliest available data for all the forcings described above, and goes up until 2020, such that the total sample size is $T = 165$. Fig. 1 plots the observed time series for temperatures and $CO_2$ emissions (in a scale of $10^{-10}$), showing the increasing trend of $CO_2$ emissions and growing temperature levels. Cursory inspection also suggests that the range of variation in temperatures has compressed in the last few decades, with the lowest temperatures rising more quickly than higher temperatures, a fact documented in Gadea Rivas and Gonzalo (2020). This pattern indicates that modelling different temperature's quantiles, including a specification where parameters may change over time, may be appropriate. Furthermore, as shown in Fig. 2, a similar behaviour and relationship seems to hold for several other climate forcings, including methane and nitrous oxide.

### 2.2. Benchmark model

As mentioned above, the purpose of this paper is to explore specifications that allow us to capture potential variation in the effects of climate forcings (in particular $GHG$ emissions) on temperature anomalies, both over time and across different regions of its distribution. Let $g_t$ denote the scalar observation of global temperatures in time periods $t = 1, \dots, T$ and $x_t$ a $p$-dimensional vector that includes an intercept, lags of temperatures and a range of climate forcings (described in Section 3), such that $g_t = f(g_t | x_t) + \epsilon_t$, where $f(.)$ is the conditional mean and $\epsilon_t$ is a possibly autocorrelated error term.

To illustrate our point, we take the empirical model of Castruccio et al. (2014) as our baseline, in which temperatures are driven by past and present (log) $CO_2$ levels, as well as cumulative emissions:

$$g_t = \beta_0 + \beta_1 \frac{1}{2}(\log CO_{2_t} + \log CO_{2_{t-1}}) + \beta_2 \sum_{i=2}^{\infty} w_i \log CO_{2_{t-i}} + \epsilon_t, \quad (1)$$



**Fig. 1.** Temperatures and $CO_2$ emissions over the sample period.

where $w_i$ denotes exponentially decaying weights such that $w_i = \omega^{-2}(1 - \omega)\omega^i$, with $0 < \omega < 1$.[5] Our implementation of (1) is flexible and differs from that of Agliardi et al. (2019): rather than specifying an ARMA process for $\epsilon_t$, we opt to capture any omitted dynamics by including lags of temperatures — we find that in most instances adding two relevant lags of temperature delivers uncorrelated residuals. Therefore, here on out the benchmark model refers to the following specification:

$$g_t = \beta_0 + \beta_1 g_{t-1} + \beta_2 g_{t-4} + \beta_3 \frac{1}{2}(\log CO_{2_t} + \log CO_{2_{t-1}})$$
$$+ \beta_4 \sum_{i=2}^{\infty} w_i \log CO_{2_{t-i}} + \epsilon_t,$$
$$g_t = f(g_{t-1}, CO_{2_{t-1}}). \quad (2)$$

### 2.3. Quantile regressions with time-varying parameters

The specification in (1) is concerned with the mean of temperature anomalies, and in that case the dependence between $g_t$ and $x_t$ would be modelled using the following equation:

$$g_t = f(g_t | x_t) + \epsilon_t, \quad (3)$$

---

[4] Results are qualitatively similar, here we report estimates based on the first dataset (see Table 1). We have also experimented with NASA's GISTEMP v4 (GISS Temperature Analysis), but again the results were very similar.
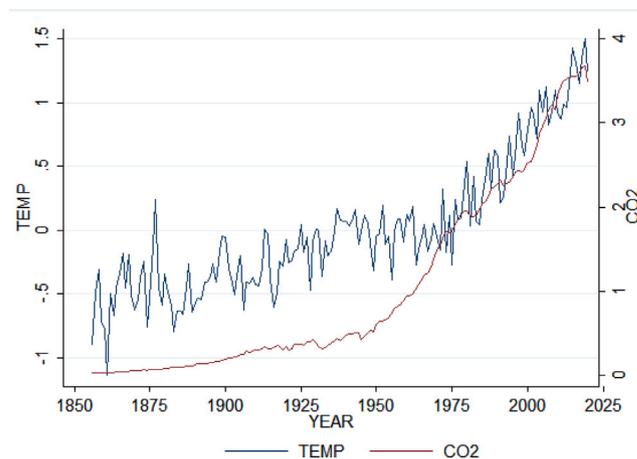
[5] While it would be interesting to allow the decay factor $\omega$ to be estimated, for the sake of simplicity we follow (Castruccio et al., 2014) and Agliardi et al. (2019) in calibrating $\omega$ to sensible values over the interval $(0, 1)$, with little overall impact in the results.
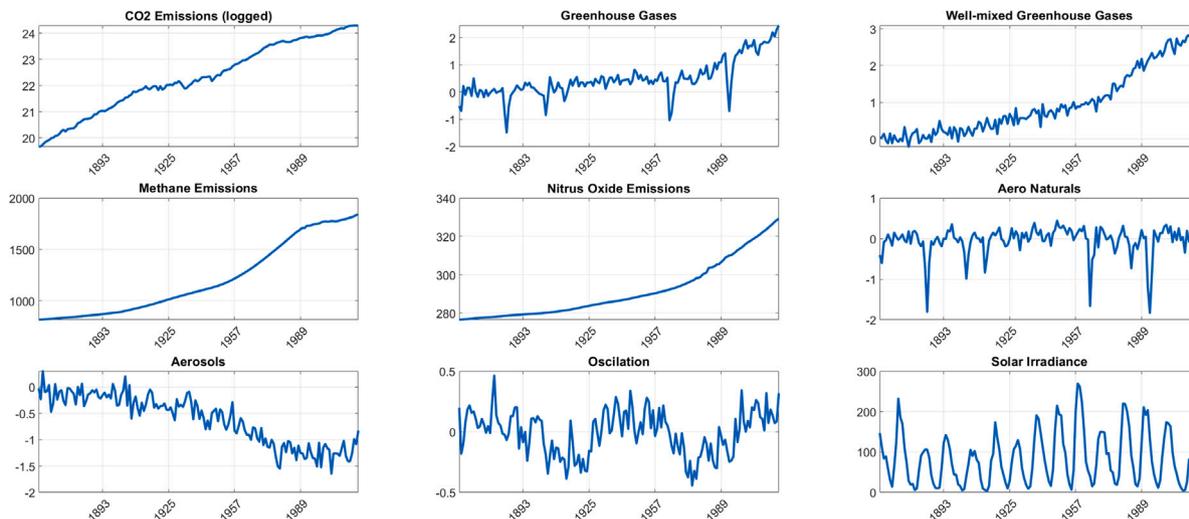
**Fig. 2.** Climate forcings over the sample period.

where the function $f(y_t|\mathbf{x}_t) = \mathbb{E}(y_t|\mathbf{x}_t) = \mathbf{x}_t\beta$ and the solution is given by:

$$\widehat{\beta} = \min_{\beta} \mathbb{E} \sum_{t=1}^{T} (y_t - \mathbf{x}_t\beta)^2, \tag{4}$$

However, it would also be of interest to consider the behaviour of different regions of the distribution of $g_t$ conditional on the forcings in $\mathbf{x}_t$. Therefore, in this case $f(y_t|\mathbf{x}_t) = \mathcal{Q}_\tau(y_t|\mathbf{x}_t) = \mathbf{x}_t\beta(\tau)$ gives the linear quantile regression (QR) model, $\tau = \tau_1, \tau_2, \ldots, \tau_n$, with solution

$$\widehat{\beta}(\tau) = \min_{\beta(\tau)} \mathbb{E} \sum_{t=1}^{T} L_\tau(y_t - \mathbf{x}_t\beta(\tau)), \tag{5}$$

where $L_\tau(u) = (\tau - \mathbb{I}(u < 0))u$ is an asymmetric loss function.

In addition to attempting to analyse the complete distribution, one would also wish to account for the fact that the effects of the climate forcings may not be constant over time and therefore we consider quantile regression models with possibly time-varying parameters — here we follow closely the exposition in Korobilis et al. (2021) which make use of Bayesian methods to estimate such parameters. The goal is to trace the full conditional distribution of $g_t$, which is achieved through the following model for each of its quantiles, say $\tau = \{0.05, 0.10, \ldots, 0.90, 0.95\}$,

$$g_t = \mathcal{Q}_\tau\left(g_t|\mathbf{x}_t\right) + \varepsilon_t, \tag{6}$$

with $\mathcal{Q}_\tau$ denoting the conditional quantile function of the $\tau$-th quantile of $g_t$. In particular, we focus on the function

$$\mathcal{Q}_\tau\left(g_t|\mathbf{x}_t\right) = \mathbf{x}_t\boldsymbol{\beta}_t(\tau), \tag{7}$$

$$\boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_{t-1}(\tau) + \boldsymbol{v}_t, \tag{8}$$

such that $\boldsymbol{v}_t \sim N_p\left(\mathbf{0}, \boldsymbol{V}_t(\tau)\right)$ is a state error with covariance matrix $\boldsymbol{V}_t(\tau)$. In this specification, parameters evolve as random walks, which allows for flexibility in that the evolution of $\boldsymbol{\beta}$ can be smooth for small $\boldsymbol{V}_t(\tau)$, or it can capture sudden shifts for large values of $\boldsymbol{V}_t(\tau)$. Details regarding the proposed methodology can be found in the technical appendix, Section 5.

The flexible specification introduced above will be used in order to produce $h$-step ahead forecasts in a rolling estimation window for the temperature anomalies, with a variety of competing models. These will include mean regression and quantile regression models by minimizing

the appropriate loss function mentioned before, as well as constant and time variant parameter alternatives, as outlined in Section 3.[6,7,8,9]

## 3. Empirical analysis

### 3.1. TVP-QR estimation results

As mentioned in Section 2, visual inspection of how temperatures and climate forcings are evolving hint at a non-constant distribution of temperatures across the sample. Such a suspicion is in fact supported by Fig. 3, which illustrates the 5th, 50th and 95th conditional quantiles estimated by the quantile model variation of the benchmark model, with constant parameters. It is clear that the distribution has become narrower over time, with the lower tail of the distribution exhibiting a larger increase than the upper tail. Furthermore, the distribution appears to have shifted from being left-skewed to a more symmetric, and now rather right-skewed, distribution. Indeed, comparing quantile estimates with mean estimates as shown in Table 2 for the benchmark model, it is apparent that different forcings exhibit asymmetries in their impact on temperature anomalies. For example, even though past temperatures have an increasing impact as we move to the upper tails of the distribution (i.e., higher persistence in conditionally higher versus lower temperature anomalies), the impact of cumulative $CO_2$ emissions is considerably higher on the tails than around the centre of the temperature distribution. This is a fact that would not be captured by conventional mean regression models and, therefore, focusing only on the study of the average relationship between emissions and temperatures may be neglecting differential impacts which can be highly relevant for policy purposes.

Furthermore, if one focuses on the upper tail of the distribution (i.e., $\tau = 0.95$), which is often thought to be of an increased public interest, estimation of the TVP-QR specification of the benchmark

---

[6] Mean regression models refer to solving Eq. (4), instead of Eq. (5) and therefore $\beta_t(\tau) = \beta_t$.

[7] The solution $\boldsymbol{V}_t(\tau) = 0$ is what allows the constant parameter case to be a special case of the TVP specification and implies that $\boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_{t-1}(\tau)$.

[8] An alternative model, with a time-varying intercept but constant parameters for the remainder of regressors (TVI) can also be obtained under the general specification.

[9] We conduct a small Monte Carlo exercise that confirms the ability of the TVP specification to capture effectively both the time-varying and the constant parameter cases — results are available upon request.
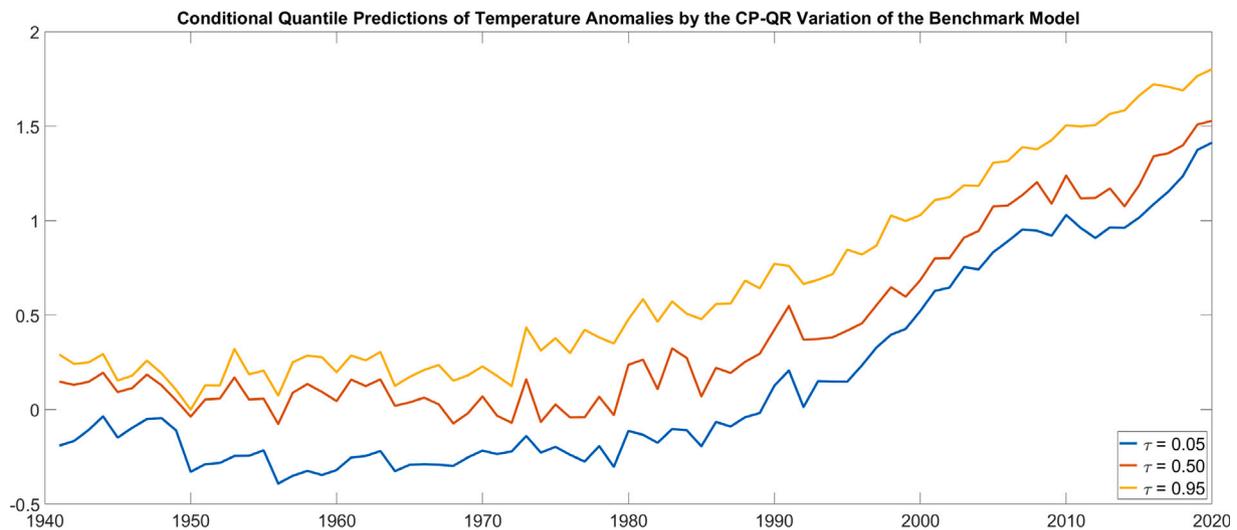
**Fig. 3.** Conditional quantiles of temperature anomalies based on the CP-QR variation of the benchmark model.
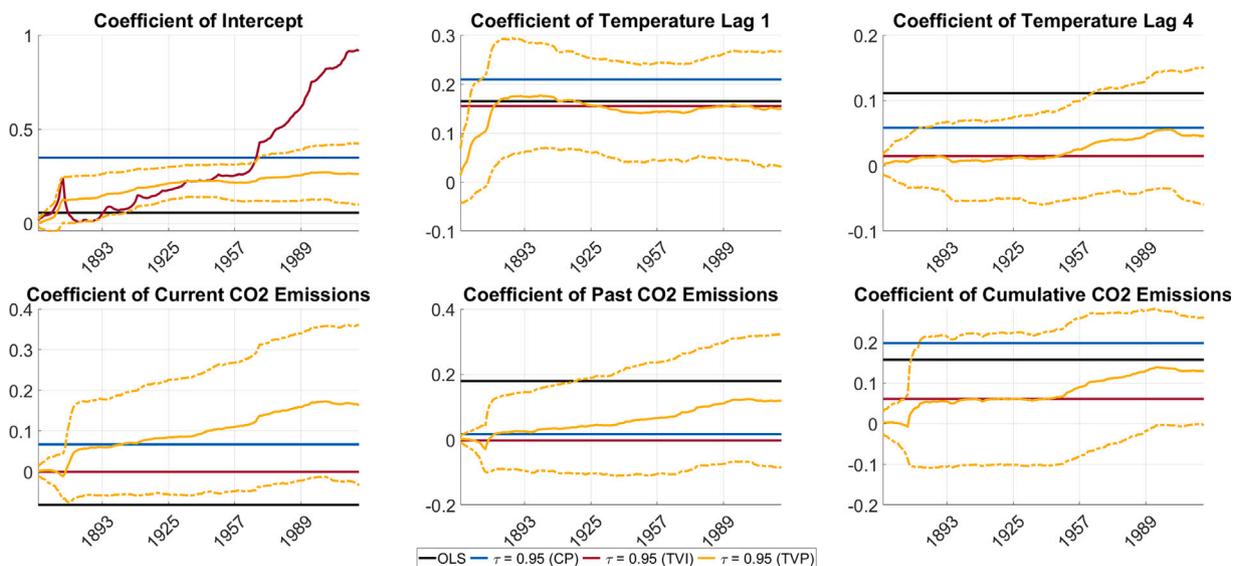


**Fig. 4.** Posterior estimates of the benchmark model for the 95th quantile, estimated by a constant parameter quantile regression, a time-varying intercept only alternative and a full time-varying parameter specification (including one standard deviation confidence bands), along with the OLS estimate.

**Table 2**
OLS versus quantile estimates for the benchmark model.

|  | OLS | CP-QR ($\tau = 0.05$) | CP-QR ($\tau = 0.50$) | CP-QR ($\tau = 0.95$) |
|---|---|---|---|---|
| Intercept | 0.0559 | −0.2638 | 0.0470 | 0.3475 |
| Temperature Lag 1 | 0.1653 | 0.0812 | 0.1392 | 0.2134 |
| Temperature Lag 4 | 0.1112 | 0.1479 | 0.1407 | 0.0581 |
| Current $CO_2$ Emissions | −0.0820 | 0.0121 | 0.0116 | 0.0558 |
| Past $CO_2$ Emissions | 0.1794 | 0.0886 | 0.0998 | 0.0241 |
| Cumulative $CO_2$ Emissions | 0.1574 | 0.2062 | 0.1391 | 0.2000 |

Notes: QR denotes "quantile regression", while CP indicates "constant parameters". Benchmark model refers to the specification shown in Eq. (2).

model in (12) enables us to explore not only the non-linearities present, but also the evolution of such effects. As it is evident in Fig. 4, we see that the effect of each of those variable changes over time, with distinctively different behaviour particularly over the end of the sample. Of great relevance is the fact that the effect of anthropogenic forcings such as $CO_2$, in cumulative form, has seen a structural break around the 1950s and has then steadily increased over the sample period, while it has now appeared to have reached a plateau. On the other hand, if one takes a look at a model that includes all climate forcings (e.g., see Fig. 10 in Appendix), well-mixed greenhouse gases

also appear to exhibit similar behaviour (albeit with different timings), while the impact of solar radiance seems to be driven by a very low-frequency cycle. Naturally, such features cannot be captured by models which do not allow for time-varying parameters. It is natural therefore to incorporate such features in forecasting exercises, as we attempt to do in the next section.

Our results, based on statistical data-driven models, are nevertheless in line with those found in the literature on transient climate response to cumulative $CO_2$ emissions (Matthews et al., 2009, Gillett et al., 2013 and, more recently, Spafford and MacDougall, 2020). Using simulations

from coupled climate–carbon models, they show a nearly linear relationship between temperature change and cumulative emissions. This almost constant climate carbon response (CCR) follows from a temperature change per unit change in atmospheric carbon that increased and an airborne fraction of cumulative emissions that decreased over time. Even though the two methodologies are not strictly comparable, the results are broadly consistent, aside from a gradual level shift around the middle of the sample. Indeed, from Figure 6 and for the first half of the sample, we observe a fairly constant coefficient of cumulative $CO_2$ emissions and a slightly increasing coefficients of $CO_2$ emissions to explain temperatures. After the middle of the 20th century, both magnitudes increase, but they remain fairly constant in the last few decades. It could also be argued that the time-varying effects detected by our approach are somehow captured by the uncertainties observed in climate models (see, for example the recent of Spafford and MacDougall, 2020 documenting the distributional shape of uncertainty of the TCRE).

On the other hand, combining carbon-cycle and physical-climate models, Ricke and Caldeira (2014) find a median time between an emission and maximum warming of 10.1 years (90 per cent probability range of 6.6–30.7 years). In our model, we cannot estimate how long it takes to observe the consequences of a single impulse in $CO_2$ emissions. Nevertheless, by noting in Fig. 6 the recent increase in the coefficients of emissions and cumulative emissions in the model for the 95th quantile, we conjecture that a response of one decade (or three at most) probably falls short.

### 3.2. Forecasting global temperatures

The flexibility present in TVP-QR models is desirable in terms of estimation and impact analysis. To explore the performance of the TVP-QR framework in a climate scenario we run a small forecasting performance exercise where 10-year ahead conditional forecasts of multiple quantiles of global temperatures are produced, using different competing models. A 10-year horizon seems a sensible choice, given that research shows that this is the median time between a given $CO_2$ emission and the corresponding maximum warming (see Ricke and Caldeira, 2014). Given the forward-looking nature of the debate around mitigation policies, such an horizon potentially allows for some of these policies to show their full effect. On the other hand, given the sample period in our analysis, this horizon enables us to produce real-time forecasts for temperature anomalies for the end of the current decade.[10]

As before, we employ (1) as our baseline specification estimated by OLS, so that the competing models then differ in terms of model specification (mean regressions, MR, vs. quantile regressions, QR), parameter alternatives (constant parameter, CP, vs. time-varying intercept only, TVI, vs. time-varying parameters, TVP), as well as in terms of forcings (e.g., $GHG$, $SolarPower$, $Oscillation$, etc.).

The following equation

$$g_{t+h} = \beta_t(\tau)\boldsymbol{x}_t + \varepsilon_{t+h}, \quad \varepsilon_{t+h} \sim ALD(\sigma(\tau)), \tag{9}$$

is the general form of the specifications used to forecast $h$-steps ahead.[11] When the state variance in Eq. (14) is set equal to zero, one can obtain the constant parameter case: a full list of the models and their corresponding performance can be seen in Table A.1. With the exception of the benchmark model, which is estimated by least squares, all other models are estimated using a variant of the automated horseshoe prior of Carvalho et al. (2010).[12] For models with additional forcings as predictors, we estimate each model with all the regressors from the

benchmark model in Eq. (2) and one additional forcing at a time, or a model where all forcings are included.[13,14]

In this pseudo out-of-sample forecasting exercise, each model is originally estimated on the first half of the sample (1856–1939) and then used to produce ten-year ahead forecasts for 19 different quantiles ($\tau = 0.05, 0.10, \dots, 0.90, 0.95$), using a rolling forecasting scheme, where at each period in time thereafter only the last $\frac{T}{2}$ periods of available information is used in the parameter estimation, before continuing to a 10-year ahead prediction. Although we estimate a range of conditional quantiles, for the sake of brevity and due to their importance, we focus on the relative quantile scores for the lower and upper tails of the distribution (5th and & 95th percentiles respectively). As a measure of performance, we use the following quantile score for each competing model $j$, which is taken as an average across all the forecasting periods,

$$QScore_h^j(\tau) = \frac{1}{R_h} \sum_{s=1}^{R_h} [g_{s+h} - \hat{Q}_\tau(g_{s+h}|\boldsymbol{x}_s)][\tau - \mathbb{I}\{g_{s+h} \leq \hat{Q}_\tau(g_{s+h}|\boldsymbol{x}_s)\}], \tag{10}$$

where $R_h$ is the length of the forecast evaluation sample. Complementary to this measure, which focuses on each quantile level specifically, we follow (Geweke and Amisano, 2010) in using the predictive likelihood (PL) as a measure of the general performance of competing models to capture the whole predictive density of global temperatures. Indeed, each model will produce a predictive distribution for temperatures *ex-ante* and consequently a predictive likelihood *ex-post*.

Thus, the PL is obtained as the 10-year ahead predictive density evaluated at the 10-year ahead realization of temperatures. In this case, a value greater than 1 implies a superior performance in capturing the complete distribution of temperatures. In order to examine if there is a gain in using any of the competing models, we proceed by examining their relative performance compared to the QScore or PL of the benchmark model as a simple ratio measure. Therefore, in terms of the quantile scores, a smaller value indicates a better performance, so that a value lower than 1 implies that the competing model outperforms the benchmark, while for the PL, a value greater than 1 implies a better performance in capturing the complete distribution of temperature anomalies.

Table 3 summarizes the forecasting performance of the different class of models across the specific tails of the distribution, as captured by the relevant quantile score and across the overall distribution, as measured by the predictive likelihood. Given these, we observe that for the lower tail most of our competing model classes outperform the benchmark specification. In particular, mean models with some time variation (by allowing only for the intercept coefficient to vary or by allowing all coefficients to change over time) are among the best performing ones (i.e., TVI-MR or TVP-MR outperforms CP-MR).

On the other hand, for the upper tail of the distribution, although time variation may not appear to be necessary, the best performing class still models the quantile levels explicitly. In terms of the overall density, both a quantile specification and the flexibility of time-varying coefficients are needed in order to maximize the forecasting gains over the benchmark model. By examining the forecasting performance of each specific model individually (e.g., see Table A.1 in Appendix) we can also see that methane emissions ($CH_4$) appear to be a good predictor of upside risk in temperature anomalies, while when attempting

---

[10] Other horizons are also feasible: in Table A.2 we present results for $h = 1$.

[11] Recall that $\boldsymbol{x}_t$ contains forcings (including the baseline forcing $CO_2$ in logged form), but also intercepts and lags of temperatures.

[12] Additional details regarding the specification can be found in Appendix.

[13] $\boldsymbol{x}_t = (f(g_{t-1}, CO_{2_{t-1}}), GHG_{t-1}, WMGHG_{t-1}, CH_{t-1}^4, N_2O_{t-1}, AN_{t-1}, AS_{t-1}, AMO_{t-1}, Solar_{t-1})$ where $f(g_{t-1}, CO_{2_{t-1}})$ is defined as in Eq. (2).

[14] One could easily expand the set of competing models to include different combinations of forcings (or principal components of such predictors), but in this forecasting exercise the main focus is to identify whether the TVP-QR approach is suitable in the current context and identify which predictors might be the main drivers of the evolution of temperature distributions. Thus, the approach of including predictors one by one has been deemed more appropriate.

**Fig. 5.** Quantile Score evolution (left panel) and the predicted conditional quantile (right panel) for the 5th conditional quantile of global temperatures for specific models.
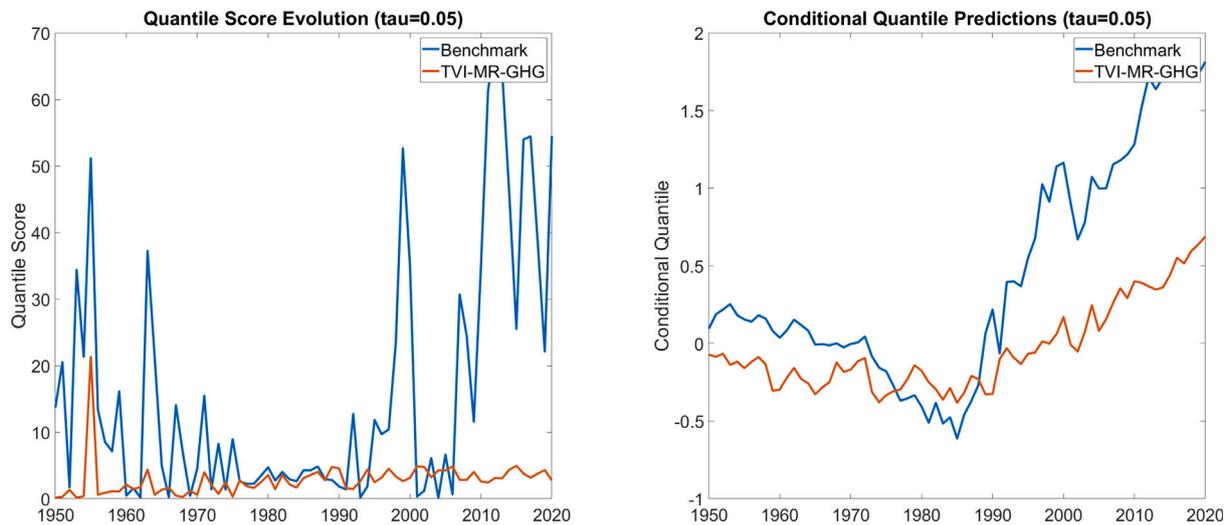


**Fig. 6.** Quantile Score evolution (left panel) and the predicted conditional quantile (right panel) for the 95th conditional quantile of global temperatures for specific models.
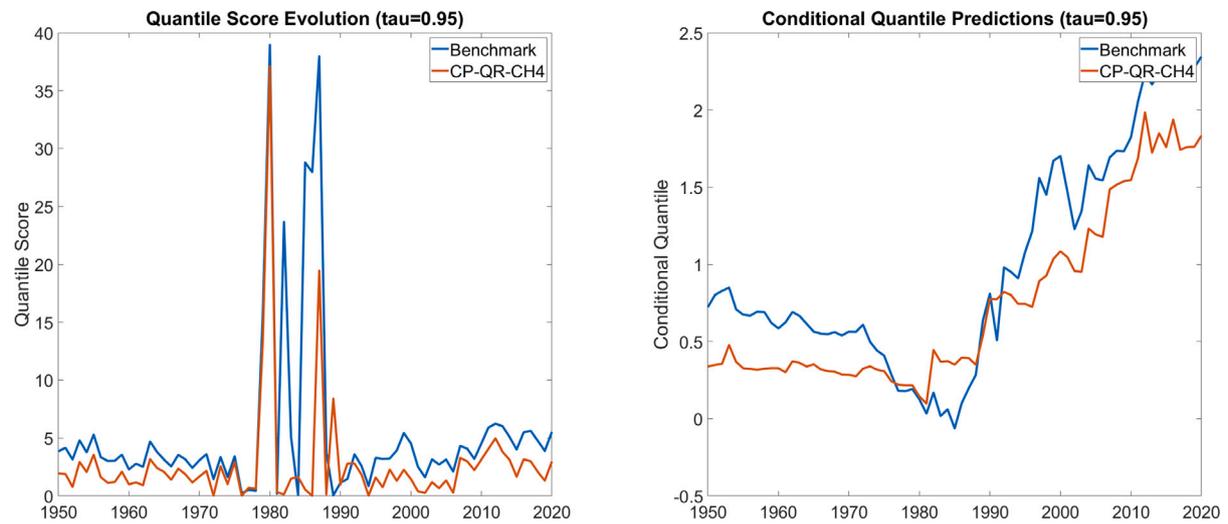
**Table 3**
Summary of forecasting performance for 10-year ahead prediction.

| Model Class | QScore5 | QScore95 | PL |
|---|---|---|---|
| CP-MR | 0.402 | 0.609 | 1.640 |
| TVI-MR | **0.181** | 0.616 | 2.142 |
| TVP-MR | 0.209 | 0.578 | 1.821 |
| CP-QR | 0.602 | **0.536** | 1.670 |
| TVI-QR | 0.312 | 1.417 | 2.151 |
| TVP-QR | 0.469 | 0.998 | **2.425** |

Notes: MR denotes "mean regression", QR denotes "quantile regression", while CP indicates "constant parameters", TVI denotes "time-varying intercept only" and TVP denotes "time-varying parameters"; QScore refers to the quantile score for each respective quantile $\tau$ and a value of less than 1 implies that the model is outperforming the benchmark model. PL refers to the predictive likelihood measure – a value greater than 1 implies that the respective competing model is outperforming the benchmark. The value indicated for each specific model class is the average performance measure of all the competing models within that class and bold values indicate the best performing class.

to predict the complete distribution, information from all the climate forcings seems to be useful.

Figs. 5 &6 illustrate the evolution of the QScore for the 5th and 95th quantile for the best performing model and the QScore of the benchmark model in the left panel, while the right panel indicates the predicted conditional quantile value for each of those models. In the case of the lower tail, the forecasting performance of the best competing model (TVI-MR specification, augmented with GHG) and the benchmark model varies over time — however, the best specification outperforms the benchmark model across most of the periods under consideration.

The shortcomings of the benchmark specification are even more apparent in the latter part of the sample, with the predictions of the two models diverging after the 1990s. This coincides with an increasing trend in temperatures overall, but particularly an accelerated trend in the lower temperatures. In this case, due to the ability of the intercept coefficient to take distinct values in each period, time-varying models may be able to pick the stronger trend present in the lower tail of the distribution much better. A similar picture can also be observed in Fig. 6, where the CP-QR specification augmented with methane emissions also outperforms the benchmark across most of the sample periods. In this case, time variation may not provide significant gains, but explicitly modelling the upper tail of the temperature distribution is relevant.

Although overall there is no single model or specification which could be deemed as the best predictive model across the whole distribution of average global temperatures, a clear fact is that the benchmark

**Table 4**
Models in the superior set for 10-year ahead prediction.

| Groups | QScore5 | QScore95 | PL |
|---|---|---|---|
| All competing models (61 models in total) | All models except: Benchmark (by OLS), TVP-QR-N$_2$O | All models except: TVI-QR-$GHG$, TVI-QR-CH$_4$, TVI-QR-$AN$, TVI-QR-$AS$, TVI-QR-$AMO$, TVI-QR-Full, TVP-QR-Full | All models except: Benchmark (by OLS), CP-MR-CH$_4$, CP-MR-N$_2$O, CP-MR-$AS$, CP-MR-$AMO$, CP-MR-$Solar$, CP-MR-Full |
| Benchmark (by OLS) + all CP-MR models (11 models in total) | All models except: Benchmark (by OLS), $AS$ | All 11 models | All models except: Benchmark (by OLS) |
| Benchmark (by OLS) + all TVI-MR models (11 models in total) | All models except: Benchmark (by OLS), $AN$ | All 11 models | All models except: Benchmark (by OLS), CH$_4$ |
| Benchmark (by OLS) + all TVP-MR models (11 models in total) | All models except: Benchmark by OLS | All 11 models | All models except: Benchmark (by OLS), $Solar$ |
| Benchmark (by OLS) + all CP-QR models (11 models in total) | Full model | All models except: Benchmark (by OLS), $AMO$ | Full model |
| Benchmark (by OLS) + all TVI-QR models (11 models in total) | All models except: Benchmark (by OLS) | All models except: $WMGHG$ | All models except: Benchmark (by OLS), $WMGHG$, N$_2$O, Full |
| Benchmark (by OLS) + all TVP-QR models (11 models in total) | Full model | All 11 models | All models except: Benchmark (by OLS), N$_2$O, $AN$, $AS$ |
| Greenhouse Gases (6 models in total) | All models except: TVP-QR | All models except: TVI-QR, TVP-QR | TVI-MR, TVI-QR, TVP-QR |
| Well-mixed Greenhouse Gases (6 models in total) | All 6 models | All models except: TVI-QR, TVP-QR | All models except: CP-MR |
| Methane Emissions (6 models) | All models except: CP-QR, TVP-QR | All models except: TVI-QR, TVP-QR | All models except: CP-MR |
| Nitrous Oxide (6 models in total) | All models except: TVP-QR | All models except: TVI-QR, TVP-QR | All models except: CP-MR |
| Aero Naturals (6 models in total) | All 6 models | All models except: TVI-QR, TVP-QR | All models except: CP-MR |
| Aerosols (6 models in total) | All 6 models | All models except: TVI-QR | All models except: CP-MR, CP-QR |
| Atlantic Multidecadal Oscillation (6 models in total) | All 6 models | All models except: TVI-QR | TVI-MR, TVI-QR, TVP-QR |
| Solar Irradiance (6 models in total) | All 6 models | All models except: TVI-QR | TVI-QR, TVP-QR |

Notes: results obtained with a confidence level of 80% and 5000 bootstrap replications.

model is often outperformed by several of the proposed models and may no longer be sufficient for accurately forecasting the distribution of global temperatures. In fact, these results suggest one should not be solely dependent on a single competing model for such predictions.[15]

Furthermore, and given that we have a large number of competing models (in excess of 60), it is worth exploring the model confidence set (MCS) approach of Hansen et al. (2011). The MCS contains a set of models such that the best model is included with a given confidence level (very much like a confidence interval in the case of a parameter). The more informative the data is, the tighter (i.e., containing fewer models) the set is going to be. The procedure is implemented by bootstrap in a sequential manner by means of an equivalence test and an elimination rule — if the equivalence test is rejected (i.e., a model is significantly inferior compared to another one), then models are gradually discarded until the equivalence test is not rejected, with the MCS containing the "surviving" models. Thus, contrary to standard

model selection criteria, which choose a single model, the MCS allows for the possibility that more than one model can be the best.

In our study, we determine the MCSs in terms of the QScore5, QScore95, and PL criteria applied to the selection of models in Tables A.1–A.2 in Appendix. We identify the best overall performing models and by sub-classes of type of models' specification (c.f. Table 3) and climate driver (c.f. Table 1). The results are based on 5000 resamples for the bootstrap implementation, employing the Tmax statistic described in Hansen et al. (2011), with a confidence level of 80%, thus aiming for relatively tight confidence sets. The Tmax statistic is based on multiple $t$-statistics where the sample loss of the $i$th model is compared to the average across models in set.

The MCSs constructed in this way are presented in Tables 4 for $h = 10$ and A.3 for $h = 1$. From Table 4, we observe that when all the 61 different models are considered, the data does not provide sufficient information to distinguish among those that best forecast the temperature's lower tail (QScore5). It is also difficult to find a pattern for the upper tail (QScore95), but at least we rule out the class of the TVI-QR models. For the PL criterion, the TVI-MR class is excluded from the superior set. Noticeably, the benchmark model is frequently excluded from the MCS.

Once we consider this analysis by model sub-classes, we again confirm that the data is largely uninformative for QScore5 and QScore95.

---

[15] Please note that we conducted similar exercises for different forecasting horizons (see Appendix) and utilizing different priors (i.e., student-t prior etc.), but the results remain qualitatively the same and in fact suggest that the performance of the benchmark model deteriorates as the horizon increases.

Nevertheless, we find that $GHG$, $Aerosols$ and $SolarIrradiance$ seem not to be good predictors of lower tail temperatures for the class of TVP-MR models. For the upper tail, the same happens with $AMO$ for CP-QR models and $WMGHG$ for TVI-QR models. When we fix the climate driver, we conclude that the TVI-QR model is never selected for QScore95, but always features in the MCS if we consider the broader PL measure. Under this criterion, CP-MR are always excluded and we have sharper choices in some cases, with TVI-QR and TVP-QR performing generally well.[16]

On the one hand, the results from the MCS analysis suggest that models allowing for time-varying coefficients and quantile heterogeneity seem to be quite robust, as we initially conjectured. On the other hand, there is no single model or class of models that always dominates, which leads us to consider a forecast averaging framework next.

### 3.3. Temperature distributions for 2030

An additional advantage of our framework is the ability to produce in a straightforward manner (without resorting to extensive simulations) a distributional snapshot of future temperatures. To further ensure robustness of our results, we follow a forecast averaging approach, a method with a long history in econometrics, which in general improves forecast accuracy by reducing forecast variances and offsetting individual model biases — see the seminal works of Bates and Granger (1969) and Granger (1989) on forecast combination, Hoeting et al. (1999) for a survey on Bayesian model averaging, and Raftery et al. (2010) and McAlinn and West (2019) for recent developments in Bayesian dynamic model averaging.

Thus, in order to guard against the model uncertainty outlined in the previous section, given the distinct performance of different models and predictors across different parts of the distribution, we propose to weigh each model's predictions according to the PL measure of the previous exercise. We believe this to be an appropriate criterion given that it is a measure which focuses on the whole distribution and not a particular tail. One could alternatively base this averaging on probabilities determined on other information criteria or QScores for a particular tail of the distribution.

Following Kapetanios et al. (2008), which has been found to work equally well or better than Bayesian averaging, we can now obtain the conditional quantile forecast of variable $g_{t+h}$ as:

$$\hat{Q}_\tau(g_{t+h}|x_t) = \sum_{j=1}^m w_j * \hat{Q}_\tau^j(g_{t+h}|x_t) \quad for \quad j = 1, \dots, m \quad (11)$$

where $w_j = \frac{exp(PL_j)}{\sum_{j=1}^m exp(PL_j)}$ and $m$ is the total number of alternative models.[17]

Utilizing this forecast combination approach, we produce *i*) 10-year ahead forecasts for the average global temperatures in 2030, given the emission realizations that occurred up until 2020, and *ii*) an alternative counterfactual scenario of emissions between 2005–2020, as that given by RCP scenarios. Different RCP scenarios make assumptions about anthropogenic-led changes in concentrations of $GHG$ in the atmosphere. This includes conjectures about efforts to curb emissions, such as changes in the share of renewables in energy production, carbon-capture measures and changes in the transport mix (more public transport, cycling, electric vehicles, etc.). For brevity, we focus on the intermediate pathway, RCP 4.5, which is a medium stabilization

scenario.[18] Given those two scenarios of emission levels (realized and RCP 4.5), we produce conditional forecasts for a range of quantiles for the average global temperature in 2030 and then proceed to obtain the predictive density via a non-parametric kernel.[19]

In Fig. 7, we can see that temperatures associated with a medium stabilization pathway (RCP 4.5, corresponding to lower emission levels if implemented) are on average lower, with a probability density function located to the left of the one associated with the realized emission levels that occurred between 2005 and 2020.[20] Specifically, focusing on the left panel we can see that the medium stabilization pathway scenario is associated with lower temperature anomalies in the upper part of the distribution. This implies that the largest temperature anomalies (i.e., 85th percentile) under the realized emissions is almost 0.2 degrees higher than under the reduced emissions scenario.

Note that one could expect the two conditional distributions to be more dissimilar to each other, particularly given the policy focus on significantly reducing several forcing emissions. However, one should point out that, in this particular instance, both the prediction given the actual emission realizations, as well as the one using the alternative counterfactual scenario of emissions, utilize the same temperature realizations. If the counterfactual emissions scenario was also associated with distinctively different temperatures (or temperature projections) up until 2020 and those were used instead in the counterfactual case, then the two distributions would be further apart.

## 4. Conclusion

In this paper we considered a range of models that have the potential to improve the prediction of temperatures. To do so, we employed a time-series quantile regression approach with TVP that allows us to *i*) quantitatively differentiate the impact of the distinct drivers across the distribution of temperatures, and *ii*) to document the extent to which these effects have changed over time. This is important as, given the substantial changes in the dynamics and volatility of temperatures in the last two centuries, our setup allows us to model these crucial features in a flexible, yet richly parameterized way. This, in turn, proves to be very convenient for forecasting, particularly tail risks in global temperatures, something that traditional, "mean-centric" approaches are not designed to achieve.

Our results show that in terms of "risky" temperatures, the benchmark model predominantly used in the literature may no longer be sufficient. In particular, mean models with some time variation, outperform their constant-parameter alternatives when concerned with the left tail of the distribution. On the other hand, for the upper tail of the distribution, though time variation may not appear to be necessary, explicitly modelling the quantile levels of temperature anomalies can offer significant forecasting gains. Furthermore, in terms of the overall density, both a quantile specification and the flexibility of time-varying coefficients are needed. Our results also underline the escalating role of additional anthropogenic forcings such as methane emissions ($CH_4$). In particular they appear to be a good predictor of upside risk in temperature anomalies, while when concerned with the overall distribution of temperature anomalies, all forcings may prove relevant.

In terms of forecasting, we also show the relevance of forecast averaging as a way of safeguarding against model uncertainty. This is particularly important when no single specification seems to dominate, as it is the case for temperatures forecasting. Moreover, we illustrate

---

[16] For $h = 1$, results in Table A.3 are similar, with the MCS always containing TVI-QR model, and the TVP-QR specification featuring in all but one cases under the PL criterion.

[17] In the case of the PL measure, the associated weights across the competing models used are not very dissimilar but rather close to a simple average. This, however, may change if one focuses on a particular tail of the distribution and uses quantile scores instead, which exhibit greater variance across models.

[18] We experimented with the original RCP pathways (2.6, 4.5, 6 and 8.5), but the results were not substantially different.

[19] One could alternatively fit a parametric function, e.g., an asymmetric *t*-distribution, as this has been proposed by Azzalini and Capitanio (2003) and used extensively in the macro-at-risk literature.

[20] RCP scenario emissions are available from 2005, and 2020 is the last observational period used to produce the conditional distribution forecasts for 2030 temperature anomalies.
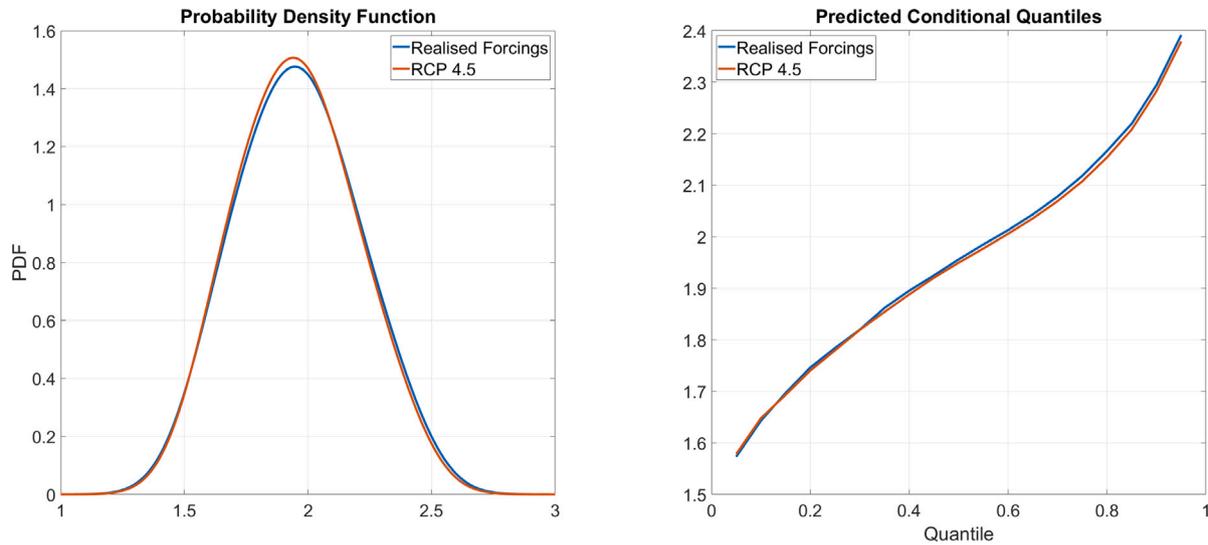
**Fig. 7.** Predicted probability density function (left panel) and predicted conditional quantiles (right panel) for 2030 average global temperatures, given realized emissions and target emissions under RCP 4.5.

the usefulness of producing forecasts that offer us a fuller picture of the distribution. Our implementation offers only a glimpse of the possibilities — indeed, it would be worthwhile to consider different scenarios for both forcings and temperatures and obtain probabilistic statements about specific targets.

It would also be interesting to compare our findings (obtained for a post-industrial revolution sample period) with a (time-varying) quantile analysis of paleoclimate data. Indeed, ascertaining the long-run contribution of orbital variables, in conjunction with GHG concentration levels, in explaining different parts of the distribution of temperatures would help us to establish a long-term baseline for its evolution and therefore provide a useful contrast with recent anthropogenic-driven climate change. This is left for future research.

## 5. Technical appendix

### 5.1. Quantile regressions with time-varying parameters

We consider here quantile regression models with possibly time-varying parameters following closely the exposition in Korobilis et al. (2021). We trace the full conditional distribution of $g_t$, by modelling each of its quantiles, say $\tau = \{0.05, 0.10, \ldots, 0.90, 0.95\}$ in the following manner:

$$g_t = Q_\tau\left(g_t | \boldsymbol{x}_t\right) + \varepsilon_t, \tag{12}$$

with $Q_\tau$ denoting the conditional quantile function of the $\tau$-th quantile of $g_t$. In particular, we focus on the function

$$Q_\tau\left(g_t | \boldsymbol{x}_t\right) = \boldsymbol{x}_t \boldsymbol{\beta}_t(\tau), \tag{13}$$

$$\boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}_{t-1}(\tau) + \boldsymbol{v}_t, \tag{14}$$

such that $\boldsymbol{v}_t \sim N_p\left(\boldsymbol{0}, \boldsymbol{V}_t(\tau)\right)$ is a state error with covariance matrix $\boldsymbol{V}_t(\tau)$. In this specification, parameters evolve as random walks, which allows for flexibility, in that the evolution of $\boldsymbol{\beta}$ can be smooth for small $\boldsymbol{V}_t(\tau)$, or it can capture sudden shifts for large values of $\boldsymbol{V}_t(\tau)$.

For the case of constant parameters, $\boldsymbol{\beta}_t(\tau) = \boldsymbol{\beta}(\tau)$, univariate conditional quantiles can be obtained by solving

$$\widehat{\beta}(\tau) = \arg\min_b \sum_{t=1}^T L_\tau(g_t - \boldsymbol{x}_t b), \tag{15}$$

where $L_\tau(u) = (\tau - \mathbb{I}(u < 0))u$ is a loss function, which is equivalent to maximizing an asymmetric Laplace likelihood (see Yu and Moyeed,

2001 for details) when $\varepsilon_t$ has density given by

$$p(\varepsilon_t; \tau, \sigma) \sim \frac{\tau(1 - \tau)}{\sigma(\tau)^2} \left[e^{(1-\tau)\frac{\varepsilon_t}{\sigma(\tau)^2}} \mathbb{I}(\varepsilon_t \leq 0) + e^{-\tau \frac{\varepsilon_t}{\sigma(\tau)^2}} \mathbb{I}(\varepsilon_t > 0)\right], \tag{16}$$

with $\sigma(\tau)^2$ a scale parameter. In turn, the asymmetric Laplace distribution (ALD) can be written as a Gaussian-Exponential scale mixture of the form

$$(\varepsilon_t | u_t, z_t) \sim \theta(\tau) z_t + \sqrt{\sigma(\tau)^2 \kappa(\tau)^2 z_t(\tau)} u_t, \tag{17}$$

where $z_t(\tau) \sim Exp(\sigma^2(\tau))$ and $u_t \sim N(0, 1)$, with $\theta(\tau), \kappa(\tau)^2$ being parameters defined as $\theta(\tau) = \frac{1-2\tau}{\tau(1-\tau)}$, $\kappa(\tau)^2 = \frac{2}{\tau(1-\tau)}$.[21]

The error distribution in Eq. (12) is assumed to follow the mixture distribution in Eq. (17). The advantages of this assumption are clear: the conditional parameter posteriors will be identical to standard expressions from linear Gaussian regression models, given that the likelihood is conditionally (on $z_t$) Gaussian.

Regarding time variation, Korobilis (2021) show that the model in Eqs. (13)–(14) can be rewritten as a high-dimensional regression. Indeed, stacking all $T$ observations, we obtain

$$Q_\tau\left(\boldsymbol{g} | \mathcal{X}\right) = \mathcal{X} \boldsymbol{\beta}^\Delta(\tau), \tag{18}$$

$$\boldsymbol{\beta}^\Delta(\tau) = \boldsymbol{v}, \tag{19}$$

where $\boldsymbol{g} = [g_1, \ldots, g_T]'$, $\boldsymbol{v} = \left[\boldsymbol{v}_1', \ldots, \boldsymbol{v}_T'\right]'$ and

$$\mathcal{X} = \begin{bmatrix} x_1 & 0 & \vdots & 0 & 0 \\ x_2 & x_2 & \vdots & 0 & 0 \\ \ldots & \ldots & \ddots & \ldots & \ldots \\ x_{T-1} & x_{T-1} & \vdots & x_{T-1} & 0 \\ x_T & x_T & \vdots & x_T & x_T \end{bmatrix}, \quad \text{and} \\ \phantom{xxxxxxxxx} T \times Tp$$

$$\boldsymbol{\beta}^\Delta(\tau) = \begin{bmatrix} \boldsymbol{\beta}_1(\tau) \\ \Delta\boldsymbol{\beta}_2(\tau) \\ \ldots \\ \Delta\boldsymbol{\beta}_{T-1}(\tau) \\ \Delta\boldsymbol{\beta}_T(\tau) \\ Tp \times 1 \end{bmatrix}. \tag{20}$$

---

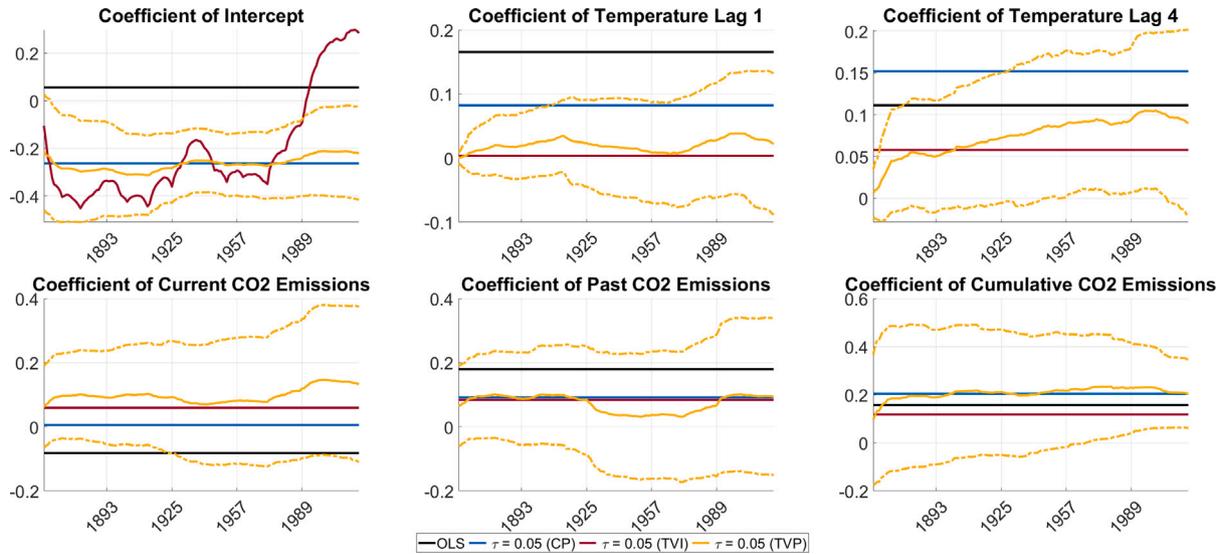[21] The result can be obtained following Kozumi & Kobayashi (2011).

**Fig. 8.** Posterior estimates of the benchmark model for the 5th quantile, estimated by a constant parameter quantile regression, a time-varying intercept only alternative and a full time-varying parameter specification (including one standard deviation confidence bands), along with the OLS estimate.

Eq. (19) may be seen as a prior for $\beta_\Delta(\tau)$, such that Eq. (18) is a linear regression model and thus estimated by standard procedures for constant parameter models. Note that we can easily recover the original vector of TVPs, $\beta = [\beta_1(\tau)', \ldots, \beta_T(\tau)']'$ from the cumulative sum of the vector of first differences, $\beta^\Delta(\tau)$.

This new parameterization combines Eq. (12) with the assumptions for the error term in Eq. (17), alongside the reparameterized TVP model in Eqs. (18) and (19), such that

$$g = \mathcal{X}\beta^\Delta(\tau) + \theta(\tau)z(\tau) + \widetilde{S}u, \tag{21}$$

where $\widetilde{S}$ is a $T \times T$ diagonal matrix with $t$th diagonal element $\sqrt{\sigma(\tau)^2\kappa(\tau)^2 z_t(\tau)}$.

Employing the following priors, $z(\tau) \sim Exp(\sigma(\tau)^2)$, $\sigma(\tau)^2 \sim inv - Gamma(\rho_1, \rho_2)$ and $\beta^\Delta(\tau) \sim N(0, V_t(\tau))$, combined with the likelihood from (21), produces the conditional posteriors

$$\beta^\Delta(\tau)|\bullet \sim N\left(Q \times \left(\mathcal{X}'U^{-1}\widetilde{y}\right), Q\right), \tag{22}$$

$$\sigma(\tau)^2|\bullet \sim inv - Gamma\left(\rho_1 + \frac{3T}{2}, \rho_2 + \sum_{t=1}^T \frac{\left(y_t^\star\right)^2}{2z_t(\tau)\kappa(\tau)^2} + \sum_{t=1}^T z_t(\tau)\right), \tag{23}$$

$$z_t(\tau)|\bullet \sim IG\left(\frac{\sqrt{\theta(\tau)^2 + 2\kappa(\tau)^2}}{|y_t - \mathcal{X}_t\beta^\Delta(\tau)|}, \frac{\theta(\tau)^2 + 2\kappa(\tau)^2}{\sigma(\tau)^2\kappa(\tau)^2}\right), \tag{24}$$

where the notation $|\bullet$ means "conditioning on other parameters and data", with $Q = \left(\mathcal{X}'U^{-1}\mathcal{X} + V_t(\tau)^{-1}\right)^{-1}$, $U = \left(\sigma(\tau)^2\kappa(\tau)^2\right) \times diag\left(z_1(\tau), \ldots, z_T(\tau)\right)$, $\widetilde{y} = (y - \theta(\tau)z(\tau))$, $y_t^\star = \left(y_t - \mathcal{X}_t\beta^\Delta(\tau) - \theta(\tau)z_t(\tau)\right)$, and $IG$ denotes the inverse Gaussian distribution.

Although a variety of priors has been considered, the results reported here consider the Makalic and Schmidt (2016) formulation of the horseshoe prior of Carvalho et al. (2010), for $\beta^\Delta(\tau)$:[22]

$$\beta^\Delta(\tau)|\lambda(\tau)^2, \{\psi_i(\tau)^2\}_{i=1}^{Tp} \sim N(0, V_t(\tau)), \tag{25}$$

$$V_{i,i}(\tau) = \lambda(\tau)^2\psi_i(\tau)^2, \quad i = 1, \ldots, Tp,$$

$$\lambda(\tau)^2|\xi(\tau) \sim inv - Gamma\left(1/2, 1/\xi(\tau)\right), \tag{26}$$

$$\xi(\tau) \sim inv - Gamma\left(1/2, 1\right), \tag{27}$$

$$\psi_i(\tau)^2|\zeta_i(\tau) \sim inv - Gamma\left(1/2, 1/\zeta_i(\tau)\right), \tag{28}$$

$$\zeta_i(\tau) \sim inv - Gamma\left(1/2, 1\right), \tag{29}$$

The vector of hyperparameters $V(\tau)$ has its own prior and is updated by the data, while $\lambda(\tau)$ and $\psi_i(\tau)$ do not require any tuning, so that the horseshoe becomes fully automatic and equally suited to low-dimensional and high-dimensional problems.[23],[24]

Given that the proposed methodology, estimates each quantile level independently, there is no guarantee that the monotonicity assumption for estimated quantiles holds (i.e., the conditional quantile value for temperature anomalies $\tau = 0.05$ should be smaller than the conditional quantile value for temperature anomalies $\tau = 0.10$). Therefore, similarly to Korobilis et al. (2021), we also employ the algorithm of Rodrigues and Fan (2017), which is specifically targeted for Bayesian quantile regressions. In essence, this algorithm smooths out the model quantile estimates, by allowing for information from neighbouring quantiles $\tau^\star$ to be used in the estimation of a relevant quantile level $\tau$. The closer the neighbouring quantile $\tau^\star$ to $\tau$ the more information its quantile curve can provide for estimation of the quantile curve at $\tau$. When $\tau = \tau^\star$ then the induced quantile is equivalent to the estimated quantile from the model.

---

[22] The horseshoe prior defined in Carvalho et al. (2010) makes use of half-Cauchy priors, which may complicate the derivation of conditional posteriors. Makalic and Schmidt (2016) show that the half-Cauchy distribution can be expressed as a mixture of inverse Gamma distributions, which is the formulation adopted here — note also that it is straightforward to embed the formulas for the conditionals posteriors of $\lambda(\tau)^2$, $\xi$, $\psi_i(\tau)^2$ and $\zeta_i(\tau)$ to the Gibbs-Sampler.

[23] See Korobilis et al. (2021) for details on the implementation of the Gibbs sampler for our TVP-QR case.

[24] It is worth noting that one could allow for variable specific scalings such that $V_{ik,ik}(\tau) = \lambda(\tau)^2\psi_{ik}(\tau)^2, i = 1, \ldots, Tp$. We refer the interested reader to Pfarrhofer (2022), where such a setup has been shown to work well for a dynamic shrinkage prior case.

[25] The specification includes all climate forcings considered along with two relevant lags of temperatures (i.e., $x_t = (f(g_{t-1}, CO_{2_{t-1}}), GHG_{t-1}, WMGHG_{t-1}, CH_{t-1}^4, N_2O_{t-1}, AN_{t-1}, AS_{t-1}, AMO_{t-1}, Solar_{t-1})$.

[26] The specification includes all climate forcings considered along with two relevant lags of temperatures (i.e., $x_t = (f(g_{t-1}, CO_{2_{t-1}}), GHG_{t-1}, WMGHG_{t-1}, CH_{t-1}^4, N_2O_{t-1}, AN_{t-1}, AS_{t-1}, AMO_{t-1}, Solar_{t-1})$.

**Fig. 9.** Posterior estimates of the model with all climate forcings for the 5th quantile, estimated by a constant parameter quantile regression, a time-varying intercept only alternative and a full time-varying parameter specification (including one standard deviation confidence bands), along with the OLS estimate. [25]
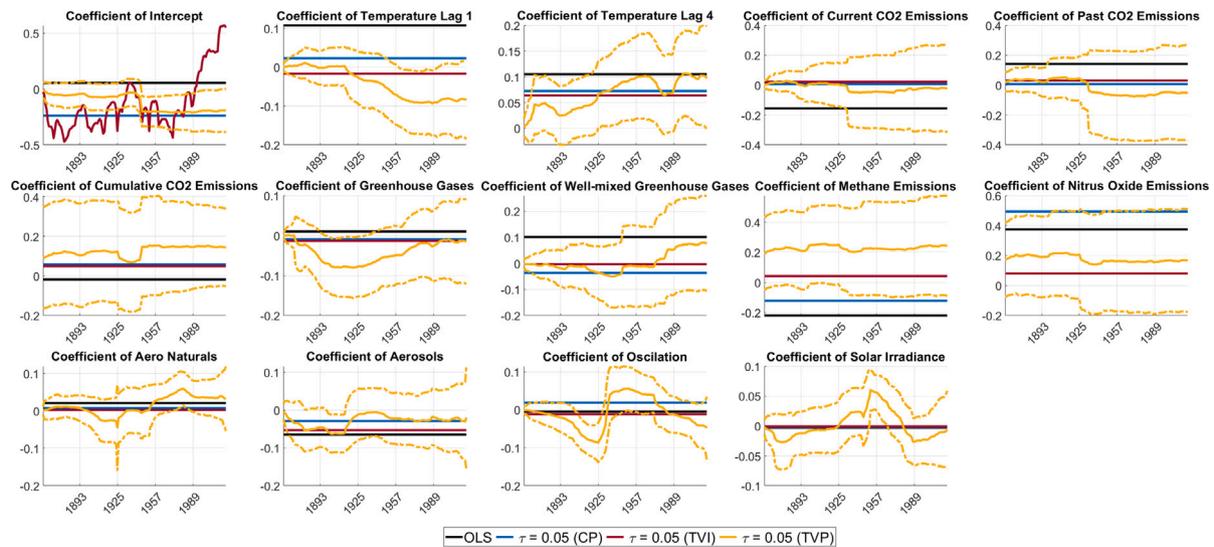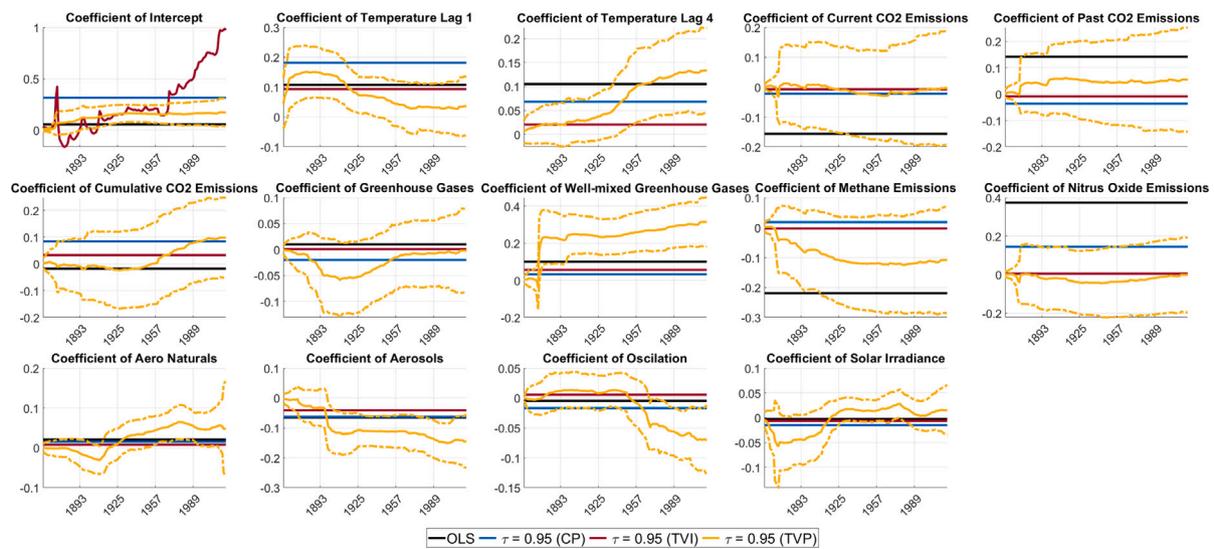


**Fig. 10.** Posterior estimates of the model with all climate forcings for the 95th quantile, estimated by a constant parameter quantile regression, a time-varying intercept only alternative and a full time-varying parameter specification (including one standard deviation confidence bands), along with the OLS estimate. [26]

## CRediT authorship contribution statement

**Anthoulla Phella:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Vasco J. Gabriel:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Luis F. Martins:** Writing – review & editing, Writing – original draft, Methodology, Investigation.

## Acknowledgments

## Appendix

See Figs. 8 and 9.

**Table A.1**

Forecasting Performance for 10-year ahead conditional quantile forecasts of global average temperatures.

| Model specification | Mean models | | | Model specification | Quantile models | | |
|---|---|---|---|---|---|---|---|
| | QScore5 | QScore95 | PL | | QScore5 | QScore95 | PL |
| CP-MR-Benchmark | 0.461 | 0.621 | 1.578 | CP-QR-Benchmark | 0.698 | 0.532 | 1.500 |
| TVI-MR-Benchmark | 0.180 | 0.611 | 2.119 | TVI-QR-Benchmark | 0.295 | 1.415 | 2.168 |
| TVP-MR-Benchmark | 0.214 | 0.561 | 1.786 | TVP-QR-Benchmark | 0.524 | 0.956 | 2.651 |
| | | | | | | | |
| CP-MR-$GHG$ | 0.441 | 0.610 | 1.618 | CP-QR-$GHG$ | 0.608 | 0.553 | 1.561 |
| CP-MR-$WMGHG$ | 0.363 | 0.579 | 1.726 | CP-QR-$WMGHG$ | 0.580 | 0.525 | 1.678 |
| CP-MR-$CH_4$ | 0.307 | 0.599 | 1.669 | CP-QR-$CH_4$ | 0.448 | 0.505 | 1.875 |
| CP-MR-$N_2O$ | 0.326 | 0.613 | 1.683 | CP-QR-$N_2O$ | 0.488 | 0.532 | 1.754 |
| CP-MR-$AeroNaturals$ | 0.459 | 0.632 | 1.571 | CP-QR-$AeroNaturals$ | 0.783 | 0.526 | 1.512 |
| CP-MR-$Aerosols$ | 0.453 | 0.630 | 1.566 | CP-QR-$Aerosols$ | 0.671 | 0.528 | 1.552 |
| CP-MR-$AMO$ | 0.436 | 0.594 | 1.687 | CP-QR-$AMO$ | 0.771 | 0.571 | 1.462 |
| CP-MR-$SolarIrradiance$ | 0.457 | 0.619 | 1.570 | CP-QR-$SolarIrradiance$ | 0.745 | 0.538 | 1.479 |
| | | | | | | | |
| TVI-MR-$GHG$ | 0.177 | 0.626 | 2.174 | TVI-QR-$GHG$ | 0.308 | 1.390 | 2.293 |
| TVI-MR-$WMGHG$ | 0.178 | 0.595 | 2.164 | TVI-QR-$WMGHG$ | 0.299 | 1.445 | 2.066 |
| TVI-MR-$CH_4$ | 0.182 | 0.607 | 2.091 | TVI-QR-$CH_4$ | 0.324 | 1.389 | 1.981 |
| TVI-MR-$N_2O$ | 0.180 | 0.603 | 2.106 | TVI-QR-$N_2O$ | 0.329 | 1.425 | 2.093 |
| TVI-MR-$AeroNaturals$ | 0.184 | 0.634 | 2.140 | TVI-QR-$AeroNaturals$ | 0.321 | 1.479 | 2.229 |
| TVI-MR-$Aerosols$ | 0.180 | 0.626 | 2.143 | TVI-QR-$Aerosols$ | 0.305 | 1.435 | 2.179 |
| TVI-MR-$AMO$ | 0.182 | 0.614 | 2.171 | TVI-QR-$AMO$ | 0.308 | 1.406 | 2.414 |
| TVI-MR-$SolarIrradiance$ | 0.180 | 0.646 | 2.147 | TVI-QR-$SolarIrradiance$ | 0.306 | 1.404 | 2.332 |
| | | | | | | | |
| TVP-MR-$GHG$ | 0.217 | 0.599 | 1.791 | TVP-QR-$GHG$ | 0.525 | 1.040 | 2.459 |
| TVP-MR-$WMGHG$ | 0.205 | 0.611 | 1.842 | TVP-QR-$WMGHG$ | 0.444 | 0.988 | 2.434 |
| TVP-MR-$CH_4$ | 0.192 | 0.578 | 1.858 | TVP-QR-$CH_4$ | 0.417 | 1.054 | 2.378 |
| TVP-MR-$N_2O$ | 0.194 | 0.566 | 1.867 | TVP-QR-$N_2O$ | 0.466 | 1.075 | 2.089 |
| TVP-MR-$AeroNaturals$ | 0.226 | 0.562 | 1.770 | TVP-QR-$AeroNaturals$ | 0.498 | 1.085 | 2.131 |
| TVP-MR-$Aerosols$ | 0.218 | 0.603 | 1.777 | TVP-QR-$Aerosols$ | 0.549 | 0.933 | 2.022 |
| TVP-MR-$AMO$ | 0.204 | 0.564 | 1.818 | TVP-QR-$AMO$ | 0.460 | 0.924 | 2.486 |
| TVP-MR-$SolarIrradiance$ | 0.216 | 0.608 | 1.761 | TVP-QR-$SolarIrradiance$ | 0.530 | 0.925 | 2.841 |
| | | | | | | | |
| CP-MR-Full | 0.314 | 0.598 | 1.737 | CP-QR-Full | 0.216 | 0.556 | 2.223 |
| TVI-MR-Full | 0.187 | 0.596 | 2.167 | TVI-QR-Full | 0.325 | 1.379 | 1.750 |
| TVP-MR-Full | 0.205 | 0.528 | 1.936 | TVP-QR-Full | 0.278 | 1.004 | 2.755 |

Notes: MR denotes "mean regression", QR denotes "quantile regression", while CP indicates "constant parameters", TVI denotes "time-varying intercept only" and TVP denotes "time-varying parameters"; "Full" implies that all possible regressors have been used; QScore refers to the quantile score for each respective quantile $\tau$ and a value of less than 1 implies that the model is outperforming the benchmark model. PL refers to the predictive likelihood measure – a value greater than 1 implies that the respective competing model is outperforming the benchmark.

**Table A.2**

Forecasting Performance for 1-year ahead conditional quantile forecasts of global average temperatures.

| Model specification | Mean models | | | Model specification | Quantile models | | |
|---|---|---|---|---|---|---|---|
| | QScore5 | QScore95 | PL | | QScore5 | QScore95 | PL |
| CP-MR-Benchmark | 0.947 | 0.922 | 0.986 | CP-QR-Benchmark | 1.188 | 0.853 | 1.063 |
| TVI-MR-Benchmark | 1.014 | 0.854 | 1.027 | TVI-QR-Benchmark | 1.211 | 1.223 | 1.185 |
| TVP-MR-Benchmark | 1.023 | 0.926 | 0.983 | TVP-QR-Benchmark | 1.330 | 1.230 | 1.094 |
| | | | | | | | |
| CP-MR-$GHG$ | 0.933 | 0.918 | 0.987 | CP-QR-$GHG$ | 1.388 | 0.833 | 1.073 |
| CP-MR-$WMGHG$ | 0.950 | 0.859 | 1.025 | CP-QR-$WMGHG$ | 1.225 | 0.851 | 1.082 |
| CP-MR-$CH_4$ | 0.979 | 0.904 | 0.995 | CP-QR-$CH_4$ | 1.261 | 0.857 | 1.055 |
| CP-MR-$N_2O$ | 0.926 | 0.860 | 1.010 | CP-QR-$N_2O$ | 1.057 | 0.808 | 1.081 |
| CP-MR-$AeroNaturals$ | 0.942 | 0.934 | 0.985 | CP-QR-$AeroNaturals$ | 1.334 | 0.868 | 1.093 |
| CP-MR-$Aerosols$ | 0.958 | 0.899 | 0.989 | CP-QR-$Aerosols$ | 1.210 | 0.831 | 1.076 |
| CP-MR-$AMO$ | 0.914 | 0.936 | 0.981 | CP-QR-$AMO$ | 1.302 | 0.892 | 1.069 |
| CP-MR-$SolarIrradiance$ | 0.930 | 0.952 | 0.981 | CP-QR-$SolarIrradiance$ | 1.313 | 0.873 | 1.076 |
| | | | | | | | |
| TVI-MR-$GHG$ | 1.039 | 0.854 | 1.010 | TVI-QR-$GHG$ | 1.584 | 1.249 | 1.183 |
| TVI-MR-$WMGHG$ | 1.036 | 0.855 | 1.029 | TVI-QR-$WMGHG$ | 1.524 | 1.232 | 1.202 |
| TVI-MR-$CH_4$ | 1.032 | 0.867 | 1.028 | TVI-QR-$CH_4$ | 1.558 | 1.247 | 1.252 |
| TVI-MR-$N_2O$ | 1.020 | 0.873 | 1.031 | TVI-QR-$N_2O$ | 1.583 | 1.226 | 1.214 |
| TVI-MR-$AeroNaturals$ | 1.028 | 0.834 | 1.029 | TVI-QR-$AeroNaturals$ | 1.537 | 1.244 | 1.204 |
| TVI-MR-$Aerosols$ | 1.064 | 0.843 | 1.087 | TVI-QR-$Aerosols$ | 1.526 | 1.298 | 1.193 |
| TVI-MR-$AMO$ | 1.030 | 0.853 | 1.029 | TVI-QR-$AMO$ | 1.552 | 1.286 | 1.182 |
| TVI-MR-$SolarIrradiance$ | 1.031 | 0.859 | 1.027 | TVI-QR-$SolarIrradiance$ | 1.495 | 1.260 | 1.195 |

*(continued on next page)*

**Table A.2** (*continued*).

| Model specification | Mean models | | | Model specification | Quantile models | | |
|---|---|---|---|---|---|---|---|
| | QScore5 | QScore95 | PL | | QScore5 | QScore95 | PL |
| TVP-MR-*GHG* | 1.067 | 0.952 | 0.921 | TVP-QR-*GHG* | 1.723 | 1.471 | 1.096 |
| TVP-MR-*WMGHG* | 1.033 | 0.902 | 0.959 | TVP-QR-*WMGHG* | 1.619 | 1.361 | 1.132 |
| TVP-MR-CH$_4$ | 1.012 | 0.909 | 0.971 | TVP-QR-CH$_4$ | 1.692 | 1.315 | 1.133 |
| TVP-MR-N$_2$O | 1.025 | 0.908 | 0.977 | TVP-QR-N$_2$O | 1.673 | 1.285 | 1.131 |
| TVP-MR-*AeroNaturals* | 1.164 | 1.067 | 0.989 | TVP-QR-*AeroNaturals* | 1.602 | 1.581 | 1.204 |
| TVP-MR-*Aerosols* | 1.122 | 0.888 | 0.973 | TVP-QR-*Aerosols* | 1.662 | 1.425 | 1.082 |
| TVP-MR-*AMO* | 1.015 | 0.923 | 0.981 | TVP-QR-*AMO* | 2.009 | 1.284 | 1.112 |
| TVP-MR-*SolarIrradiance* | 0.957 | 0.953 | 0.956 | TVP-QR-*SolarIrradiance* | 1.820 | 1.322 | 1.077 |
| CP-MR-Full | 1.031 | 0.889 | 0.992 | CP-QR-Full | 1.216 | 0.954 | 1.049 |
| TVI-MR-Full | 1.089 | 0.871 | 1.038 | TVI-QR-Full | 1.697 | 1.278 | 1.166 |
| TVP-MR-Full | 1.143 | 0.951 | 0.909 | TVP-QR-Full | 1.989 | 1.530 | 1.076 |

Notes: MR denotes "mean regression", QR denotes "quantile regression", while CP indicates "constant parameters", TVI denotes "time-varying intercept only" and TVP denotes "time-varying parameters"; "Full" implies that all possible regressors have been used; QScore refers to the quantile score for each respective quantile $\tau$ and a value of less than 1 implies that the model is outperforming the benchmark model. PL refers to the predictive likelihood measure – a value greater than 1 implies that the respective competing model is outperforming the benchmark.

**Table A.3**

Models in the superior set for 1-year ahead prediction.

| Groups | QScore5 | QScore95 | PL |
|---|---|---|---|
| All competing models (61 models in total) | All models except: all TVI-QR models but TVI-QR-Benchmark, TVI-QR-*Solar*, all TVP-QR models | All models except: TVI-QR-*GHG*, TVI-QR-CH$_4$, TVI-QR-*AN*, TVI-QR-*AN*, TVI-QR-*AMO*, TVI-QR-Full, TVP-QR-Full | All models except: all CP-MR models, all TVP-MR models, TVI-MR-Benchmark, TVI-MR-*GHG*, TVI-MR-CH$_4$, TVI-MR-*AMO* |
| Benchmark (by OLS) + all CP-MR models (11 models in total) | All 11 models | All models except: *Solar* | All 11 models |
| Benchmark (by OLS) + all TVI-MR models (11 models in total) | All models except: Full model | All 11 models | All models except: *GHG* |
| Benchmark (by OLS) + all TVP-MR models (11 models in total) | All models except: *AN*, *AS* | All models except: *AN* | All models except: *GHG*, *Solar*, Full model |
| Benchmark (by OLS) + all CP-QR models (11 models in total) | All models except: *GHG*, *AN* | All 11 models | All 11 models |
| Benchmark (by OLS) + all TVI-QR models (11 models in total) | Benchmark (by OLS) | Benchmark (by OLS), Benchmark variation | All 11 models |
| Benchmark (by OLS) + all TVP-QR models (11 models in total) | Benchmark (by OLS) | All models except: *AN*, Full model | All models except: Benchmark (by OLS) |
| Greenhouse Gases (6 models in total) | CP-MR | TVI-MR, CP-QR | CP-QR, TVI-QR, TVP-QR |
| Well-mixed Greenhouse Gases (6 models in total) | CP-MR | All models except: TVI-QR, TVP-QR | CP-QR, TVI-QR, TVP-QR |
| Methane Emissions (6 models) | CP-MR, TVI-MR, TVP-MR | All models except: TVI-QR, TVP-QR | TVI-QR |
| Nitrous Oxide (6 models in total) | All models except: TVI-QR, TVP-QR | All models except: TVI-QR, TVP-QR | TVI-QR, TVP-QR |
| Aero Naturals (6 models in total) | CP-MR | TVI-MR, CP-QR | TVI-QR, TVP-QR |
| Aerosols (6 models in total) | All models except: TVI-QR, TVP-QR | All models except: TVI-QR, TVP-QR | All models except: CP-MR, TVP-MR |
| Atlantic Multidecadal Oscillation (6 models in total) | CP-MR | All models except: TVI-QR, TVP-QR | CP-QR, TVI-QR, TVP-QR |
| Solar Irradiance (6 models in total) | CP-MR, TVP-MR | TVI-MR, CP-QR | CP-QR, TVI-QR, TVP-QR |

Notes: results obtained with a confidence level of 80% and 5000 bootstrap replications.

## References

Agliardi, E., Alexopoulos, T., Cech, C., 2019. On the relationship between GHGs and global temperature anomalies: Multi-level rolling analysis and copula calibration. Environ. Res. Econ. 72, 109–133.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J. R. Stat. Soc. Ser. B Stat. Methodol. 65, 367–389.

Ballester, J., Giorgi, F., Rodó, X., 2010. Changes in European temperature extremes can be predicted from changes in PDF central statistics: A letter. Clim. Change 98, 277–284.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Oper. Res. Q. 20, 451–468.

Carvalho, C.M., Polson, N.G., Scott, J.G., 2010. The horseshoe estimator for sparse signals. Biometrika 97, 465–480.

Castruccio, S., McInerney, D.J., Stein, M.L., Liu Crouch, F., Jacob, R.L., Moyer, E.J., 2014. Statistical emulation of climate model projections based on precomputed GCM runs*. J. Clim. 27, 1829–1844.

Castruccio, S., Stein, M.L., 2013. Global space-time models for climate. Ann. Appl. Stat. 7, 1593–1611.

Chang, Y., Kaufmann, R.K., Kim, C.S., Miller, J.I., Park, J.Y., Park, S., 2020. Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. J. Econometrics 214, 274–294.

Donat, M.G., Alexander, L.V., 2012. The shifting probability distribution of global daytime and night-time temperatures. Geophys. Res. Lett. 39, L14707.

Estrada, F., Perron, P., 2017. Extracting and analyzing the warming trend in global and hemispheric temperatures. J. Time Series Anal. 38, 711–732.

Gadea Rivas, M.D., Gonzalo, J., 2020. Trends in distributional characteristics: Existence of global warming. J. Econometrics 214, 153–174.

Geweke, J., Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. Int. J. Forecast. 26, 216–230.

Gillett, N.P., Arora, V.K., Matthews, D., Allen, M.R., 2013. Constraining the ratio of global warming to cumulative CO2 emissions using CMIP5 simulations. J. Clim. 26, 6844–6858.

Granger, C.W.J., 1989. Combining forecasts - twenty years later. J. Forecast. 8, 167–173.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. Statist. Sci. 14, 382–417.

Kapetanios, G., Labhard, V., Price, S., 2008. Forecast combination and the Bank of England's suite of statistical forecasting models. Econ. Model. 25, 772–792.

Kaufmann, R.K., Kauppi, H., Mann, M.L., Stock, J.H., 2013. Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. Clim. Change 118, 729–743.

Korobilis, D., 2021. High-dimensional macroeconomic forecasting Using message passing algorithms. J. Bus. Econom. Statist. 39, 493–504.

Korobilis, D., Landau, B., Musso, A., Phella, A., 2021. The time-varying evolution of inflation risks. SSRN Electr. J.

Larson, E.J.L., Portmann, R.W., 2019. Anthropogenic aerosol drives uncertainty in future climate mitigation efforts. Sci. Rep. 9, 16538.

Leeds, W.B., Moyer, E.J., Stein, M.L., 2015. Simulation of future climate under changing temporal covariance structures. Adv. Statist. Climatol. Meteorol. Oceanogra. 1, 1–14.

Makalic, E., Schmidt, D.F., 2016. A simple sampler for the horseshoe estimator. IEEE Signal Process. Lett. 23, 179–182.

Matthews, H.D., Gillett, N.P., Stott, P.A., Zickfeld, K., 2009. The proportionality of global warming to cumulative carbon emissions. Nature 459, 829–832.

Matthews, H.D., Zickfeld, K., Knutti, R., Allen, M.R., 2018. Focus on cumulative emissions, global carbon budgets and the implications for climate mitigation targets. Environ. Res. Lett. 13, 010201.

McAlinn, K., West, M., 2019. Dynamic Bayesian predictive synthesis in time series forecasting. J. Econometrics 210, 155–169.

Miller, J.I., Nam, K., 2020. Dating hiatuses: a statistical model of the recent slowdown in global warming and the next one. Earth Syst. Dyn. 11, 1123–1132.

Miller, R.L., Schmidt, G.A., Nazarenko, L.S., Bauer, S.E., Kelley, M., Ruedy, R., Russell, G.L., Ackerman, A.S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T.L., Cook, B.I., Cruz, C.A., Del Genio, A.D., Elsaesser, G.S., Faluvegi, G., Kiang, N.Y., Kim, D., Lacis, A.A., Leboissetier, A., LeGrande, A.N., Lo, K.K., Marshall, J., Matthews, E.E., McDermid, S., Mezuman, K., Murray, L.T., Oinas, V., Orbe, C., Pérez García-Pando, C., Perlwitz, J.P., Puma, M.J., Rind, D., Romanou, A., Shindell, D.T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., Yao, M.-S., 2021. CMIP6 Historical Simulations (1850–2014) With GISS-E2.1. J. Adv. Modelling Earth Syst. 13, 1–35.

Pfarrhofer, M., 2022. Modeling tail risks of inflation using unobserved component quantile regressions. J. Econom. Dynam. Control 143, 104493.

Pretis, F., Mann, M.L., Kaufmann, R.K., 2015. Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. Clim. Change 131, 705–718.

Raftery, A.E., Kárný, M., Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics 52, 52–66.

Ricke, K.L., Caldeira, K., 2014. Maximum warming occurs about one decade after a carbon dioxide emission. Environ. Res. Lett. 9, 124002.

Rodrigues, T., Fan, Y., 2017. Regression adjustment for noncrossing Bayesian quantile regression. J. Comput. Graph. Statist. 26, 275–284.

Spafford, L., MacDougall, A.H., 2020. Quantifying the probability distribution function of the transient climate response to cumulative CO2 emissions. Environ. Res. Lett. 15, 034044.

Tol, R.S.J., 2005. Adaptation and mitigation: trade-offs in substance and methods. Environ. Sci. Policy 8, 572–578.

Wang, P., Yang, Y., Xue, D., Ren, L., Tang, J., Leung, L.R., Liao, H., 2023. Aerosols overtake greenhouse gases causing a warmer climate and more weather extremes toward carbon neutrality. Nature Commun. 14, 7257.

Yu, K., Moyeed, R.A., 2001. Bayesian quantile regression. Statist. Probab. Lett. 54, 437–447.