Received: 13 October 2020 | Accepted: 18 June 2021

DOI: 10.1111/rssa.12732

ROYAL STATISTICAL SOCIETY
Series A Statistics in Society

ORIGINAL ARTICLE

Econometric modelling of carbon dioxide emissions and concentrations, ambient temperatures and ocean deoxygenation

Alok Bhargava

University of Maryland, School of Public Policy, College Park, MD, USA

Correspondence
Alok Bhargava, University of Maryland School of Public Policy, College Park, MD 20742, USA
Email: Bhargava@umd.edu

Abstract

This paper analysed several longitudinal data sets for investigating the dynamic inter-relationships between CO_2 emissions and atmospheric concentrations, ambient temperatures and ocean acidification and deoxygenation. The methodological framework addressed issues such as the use of temperature 'anomalies', diffusion of CO_2 to atmospheric stations, distributional misspecification and non-stationarity of errors affecting empirical models, and use of spline functions for modelling trends in temperatures. Longitudinal data on CO_2 emissions for 163 countries and atmospheric CO_2 concentrations at 10 stations, ambient temperatures from over 8,500 weather stations and seawater composition from over 380,000 oceanographic stations were analysed for 1985–2018 by estimating dynamic random effects models using maximum likelihood methods. The main findings were that CO_2 emissions exhibited rapid upward trends at the country level, while minimum and maximum temperatures showed cyclical patterns; economic activity and population levels were associated with higher CO_2 emissions. Second, there were gradual upward trends in annual and seasonal temperatures compiled at weather stations, and atmospheric CO_2 concentrations were significantly associated with higher temperatures in the hemispheres. Third, there was a steady decline in dissolved oxygen levels, and the interactive effects of water temperatures and pH levels were significant. Overall, the

results underscore the benefits of reducing CO_2 emissions for ambient temperatures and for ocean deoxygenation. Synergies between CO_2 emissions, ambient temperatures and ocean acidification are likely to exacerbate the melting of polar ice.

KEYWORDS

ambient temperatures, climate change, CO_2 emissions, maximum likelihood estimation, mitigation policies, ocean acidification, ocean deoxygenation, socioeconomic factors

1 | INTRODUCTION

The emissions of greenhouse gases such as carbon dioxide, methane and nitrous oxide due to anthropogenic activities are likely to increase ambient temperatures and ocean acidification via the absorption of CO_2 (Intergovernmental Panel on Climate Change, 2014). Moreover, increases in ambient and ocean temperatures can accelerate the melting of polar ice and in turn raise sea levels. Historically, the inter-relationships between atmospheric CO_2 concentrations, ambient temperatures and ocean acidification were recognized by Arrhenius (1896) and simple interpolations were presented. With the availability of climatic data, it is useful to analyse such inter-relationships although the analyses present numerous methodological challenges. For example, it is common in climatic research to interpret positive annual temperature 'anomalies' (i.e. deviations of temperatures from a sample mean computed for an earlier time interval) as evidence of global warming (IPCC, 2014). However, the trends underlying global temperatures are complex and exhibit cyclical and seasonal fluctuations that are likely to be different in the Northern and Southern hemispheres. Moreover, while the use of temperature anomalies is suitable for illustrative purposes, it can hamper the testing of statistical hypotheses (Lehmann, 1985; Wiener, 1949) regarding the dynamics of temperatures. The stark seasonal patterns in the hemispheres and the diffusion of CO_2 from vegetation and anthropogenic activity to the atmosphere underscore the need for modelling the effects of atmospheric CO_2 concentrations on ambient temperatures compiled at weather stations using the appropriate statistical methods for longitudinal data.

Furthermore, it is estimated that up to two thirds of CO_2 emissions are ultimately absorbed by oceans and that CO_2 can remain in the atmosphere for several decades. Thus, empirical models assessing the effects of CO_2 emissions and atmospheric concentrations on ambient temperatures using country and station level temperature data can provide useful insights. While 'negative externalities' such as those due to CO_2 diffusion have been studied in the economics literature (Pigou, 1932) and indirect taxes have been proposed (e.g. Stern, 2006), it is important to take into account air–sea fluxes for ocean acidification (Breitburg et al., 2018; Keeling et al., 2010). The data on atmospheric CO_2 concentrations are available at regular intervals for six and four stations, respectively, in the Northern and Southern hemispheres (Scripps Institute of Oceanography, 2019). In contrast, longitudinal data on temperatures are available from over 8,500 weather stations. As discussed below, CO_2 concentrations in the Southern hemisphere tend to catch up with those in the Northern hemisphere where most countries are located. It is important to develop dynamic models for ambient temperatures incorporating geodesic variables such as latitudes, longitudes and elevation (Arrhenius, 1896; Linacre, 1992; Torge, 2001) for investigating the effects of atmospheric CO_2 concentrations on temperatures. Moreover,

there are likely to be interactive effects of pH variable reflecting acidification and water temperatures for ocean deoxygenation.

From the standpoint of statistical analyses of climatic data, stochastic properties of variables are important for the formulation of empirical models. For example, there are upward trends in annual CO_2 emissions by countries and in annual and quarterly atmospheric CO_2 concentrations during 2000–2016. By contrast, annual and quarterly ambient temperatures from weather stations exhibit cyclical patterns so that it is helpful to analyse 3- or 5-yearly averages and consider the use of spline functions (Wongsai et al., 2017) when modelling the effects CO_2 concentrations for ambient temperatures. Furthermore, there are difficulties in extrapolating the limited numbers of observations from oceanographic stations for the construction of high resolution grids (Robeson, 1997). While the use of averages for oceans for $5^o \times 5^o$ grids is appealing (National Centers for Environmental Information, 2019), there are likely to be few repeated observations available for the grids. The choice between interpolations for creating finer grids versus analyses of direct observations from oceanographic stations is facilitated by the fact that stochastic properties of *in situ* measurements are not 'contaminated' by mathematical transformations (Wiener, 1949). Moreover, even if the stations are non-randomly located near populous regions, mainly the estimated coefficients of constant terms in empirical models are affected in large samples (Wang et al., 1997). Lastly, longitudinal data analyses using maximum likelihood methods that are capable of tackling distributional misspecification of errors affecting empirical models (Koopmans et al., 1950; Mann & Wald, 1943) and non-stationary processes (Bhargava & Sargan, 1983), and can provide insights into the effects of CO_2 concentrations on ambient temperatures and on ocean acidification.

The structure of this paper is as follows: Section 2 describes the longitudinal data sets used in the analyses for CO_2 emissions and atmospheric concentrations, ambient temperatures, sea water composition, aerosol optical depths and socioeconomic variables. The analytical framework is outlined in Section 3 and certain methodological issues are tackled. The inter-relationships between climatic variables, and the data sets employed are summarized in the Figure 1. The empirical models are outlined in Section 4 and the maximum likelihood estimation methods are described. The stochastic properties of variables such as CO_2 emissions and ambient temperatures are discussed for motivating the modelling strategies. The results for country level CO_2 emissions and ambient temperatures are presented in Section 5. Section 6 presents the empirical models for minimum and maximum annual and quarterly temperatures compiled at weather stations in the Northern and Southern hemispheres. Section 7 presents the results for dissolved oxygen levels using data from oceanographic stations and for averages in $5° \times 5°$ grids constructed for longitudinal analyses. The results from various components of the analyses are summarized in Section 8 from an evidence-based climate mitigation policy perspective, and the need for further research on the melting of polar ice is underscored.

2 | THE DATA AND DESCRIPTIVE STATISTICS

Several longitudinal data sets were processed for the analyses and are noted in the Figure 1. First, annual data on CO_2 (and methane and nitrous oxide) emissions for countries were available for 1970–2016 from the Emissions Database for Global Atmospheric Research (2019). However, for harmonizing the longitudinal analyses, the data for 2000–2016 were analysed. Moreover, 3- and 5-yearly averages were constructed owing to the cyclical patterns in ambient temperatures. Second, socioeconomic variables such as gross domestic product (GDP) per capita, total population and industry value added as a percentage of GDP were taken from the World Development Indicators (World Bank,

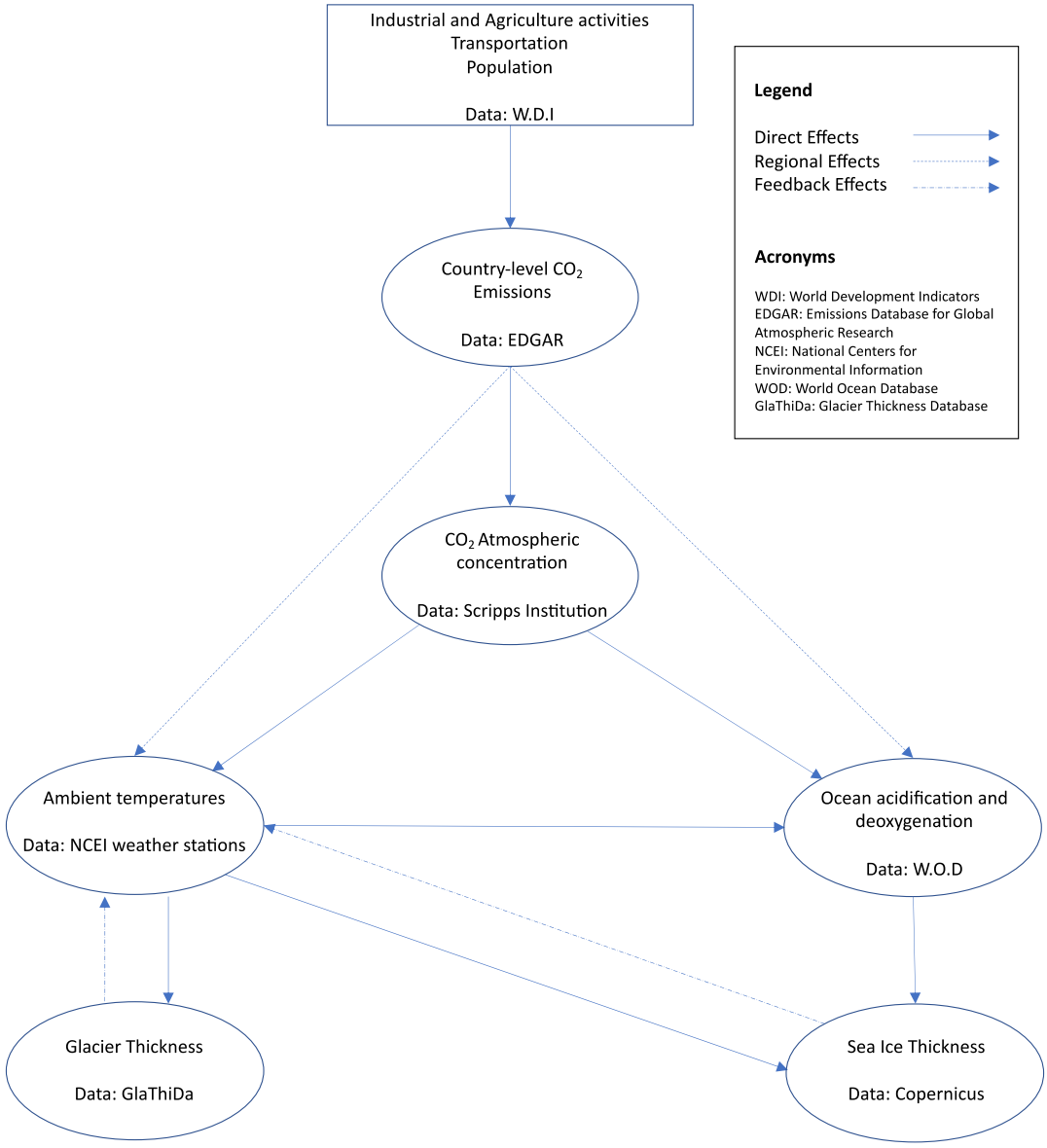

FIGURE 1 A graphical representation of the empirical inter-relationships between anthropogenic activities, CO_2 emissions and atmospheric concentrations, ambient temperatures and ocean deoxygenation [Colour figure can be viewed at wileyonlinelibrary.com]

2018) and merged with CO_2 emissions data using International Organization for Standardization-3 codes. The sample means of a subset of variables are reported in Table 1 and are discussed below.

Third, monthly data on minimum and maximum ambient temperatures were compiled at over 50,000 weather stations during 1960–2018 by the National Centers for Environmental Information (2018). However, many weather stations were dropped and new stations were added during this period. For the estimation of dynamic random effects models, it is necessary to analyse 'balanced panels' with complete time observations on stations; dropping and adding of weather stations can complicate the analyses of the trends underlying ambient temperatures. Thus, an annual data set was constructed for

TABLE 1 Sample means and standard deviations of 3-yearly averages of CO_2 emissions, ambient temperatures and socioeconomic variables for 163 countries during 2000-2016[a]

	2001		2004		2007		2010		2013		2016	
Year:	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Total CO_2 emissions, kton	142816	573095	158663	652278	175349	753824	185491	841984	197473	930340	199507*	938475
GDP per capita, USD	6552	10121	8843	13785	11959	17873	12803	18425	145273	20351	13366*	18684
Total population, 100,000	366.64	1336	380.65	1380	394.88	1422	409.40	1462	424.26	1501	439.48*	1539
Industry value added, % GDP	26.23	10.73	26.58	11.23	27.48	12.79	26.48	12.33	26.36	12.37	25.04	10.26
Minimum temperature, $^{\circ}C$[b]	14.71	7.99	14.75	8.17	14.84	8.00	14.86	8.21	14.87	8.09	15.12*	8.12
Maximum temperature, $^{\circ}C$[b]	24.56	8.21	24.46	8.34	24.60	8.10	24.63	8.39	24.70	8.25	24.98*	8.27
Aerosol optical depth[c], n	0.214	0.11	0.218	0.12	0.221	0.12	0.221	0.12	0.208	0.11	0.207*	0.12

[a]Longitudinal data on emissions and socioeconomic variables for 163 countries were taken from EDGAR and World Development Indicators databases.

[b]Minimum and maximum temperatures were computed using data from 8577 weather stations (Table 2).

[c]Aerosol optical depths were mapped with the data for countries.

*Changes from 2001 to 2016 were significant ($p<0.05$) using paired t-tests.

TABLE 2 Sample means and standard deviations of 5-yearly averages for annual and quarterly temperatures assessed from 8577 weather stations and CO_2 measurements from 10 stations in the Northern and Southern hemispheres during 1985-2018[a]

Year:	1987		1992		1997		2002		2007		2012		2017	
Variables:	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**.D**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Northern hemisphere														
Annual min temp, °C	4.183	7.70	4.545	7.67	4.646	7.56	4.846	7.49	4.867	7.42	4.876	7.45	5.306 + *	7.63
Annual max temp, °C	16.108	7.91	16.240	7.80	16.336	7.83	16.611	7.69	16.612	7.68	16.467	7.80	16.788+ *	7.93
Min temp quarter 1, °C	−4.392	10.26	−3.470	10.13	−3.461	10.12	−3.454	10.02	−3.568	10.10	−3.903	10.25	−2.917+ *	9.98
Min temp quarter 2, °C	8.459	6.63	8.469	6.65	8.278	6.80	8.701	6.73	8.684	6.60	8.861	6.73	9.141+ *	6.71
Min temp quarter 3, °C	13.106	6.03	13.258	5.82	13.630	5.80	13.757	5.68	13.856	5.70	14.115	5.64	14.335+ *	5.80
Min temp quarter 4, °C	−0.136	8.83	0.093	8.87	0.446	8.62	0.565	8.68	0.686	8.44	0.684	8.46	1.371+ *	8.67
Max temp quarter 1, °C	6.838	10.52	7.553	10.28	7.498	10.46	7.661	10.33	7.517	10.48	7.079	10.61	8.074+ *	10.31
Max temp quarter 2, °C	21.476	6.96	21.021	6.80	20.710	6.96	21.411	6.86	21.391	6.81	21.347	7.02	21.696+ *	6.84
Max temp quarter 3, °C	25.798	5.97	25.876	5.79	26.283	5.80	26.509	5.71	26.515	5.68	26.676	5.74	26.760+ *	5.69
Max temp quarter 4, °C	10.756	9.27	10.785	9.34	11.179	9.30	11.094	9.21	11.302	9.09	11.281	9.15	11.816+	9.28
Annual atmos. CO_2, PPM	349.94	-	357.09	-	364.95	-	373.95	-	384.21	-	394.66	-	405.63	-
Southern hemisphere														
Annual min temp, °C	13.852	7.07	13.783	7.13	13.926	7.17	13.932	7.15	13.955	7.24	14.047	7.25	14.493 +	7.29
Annual max temp, °C	24.942	7.13	25.055	7.23	25.129	7.16	25.341	7.13	25.532	7.10	25.554	7.15	25.970+	7.17
Min temp quarter 1, °C	17.385	6.15	17.433	6.16	17.590	6.15	17.632	6.18	17.721	6.22	17.840	6.16	18.134+	6.22
Min temp quarter 2, °C	12.135	7.90	12.091	8.05	12.013	8.05	12.106	8.04	11.867	8.18	12.130	8.20	12.358+	8.30
Min temp quarter 3, °C	9.943	8.27	9.830	8.31	10.246	8.39	9.946	8.36	10.059	8.48	10.122	8.53	10.343+	8.64
Min temp quarter 4, °C	15.213	6.81	15.203	6.82	15.293	6.86	15.372	6.91	15.580	6.94	15.580	6.90	15.858+	6.84
Max temp quarter 1, °C	28.718	6.54	28.656	6.54	28.692	6.43	28.866	6.43	29.084	6.46	29.024	6.46	29.497+	6.51
Max temp quarter 2, °C	22.631	7.76	22.795	7.88	22.745	7.79	22.953	7.74	23.032	7.723	23.073	7.74	23.292+	7.78
Max temp quarter 3, °C	21.221	8.35	21.287	8.39	21.555	8.48	21.662	8.35	21.860	8.37	21.974	8.39	21.942+	8.50
Max temp quarter 4, °C	26.730	7.23	26.840	7.29	26.815	7.18	27.144	7.20	27.443	7.19	27.498	7.19	27.783+	7.08
Annual atmos. CO_2, PPM	347.47	-	354.53	-	362.14	-	371.16	-	380.95	-	390.93	-	401.88	-

[a]There were 7886 and 691 weather stations, respectively, in the Northern and Southern hemispheres with complete data.

+Paired *t*-tests were significant ($p<0.05$) for changes in temperatures between 1987 and 2017.

*Independent *t*-tests were significant for the changes in temperatures in Northern versus Southern hemispheres.

1985–2018, and 3 and 5-yearly averages were analysed using data from 7886 to 691 weather stations, respectively, in the Northern and Southern hemispheres. The results for 5-yearly averages are reported in Table 2 for economizing on space. Fourth, aerosol optical depths were computed by Li et al., (2016) from remote sensing and land-based stations. The aerosol optical depth variable was merged with the weather stations data using geodesic coordinates; averages of aerosol depths were also computed at the country level. Due to space constraints, sample means of aerosol optical depths are reported for countries in Table 1 but are omitted from Table 2.

Fifth, the data on atmospheric CO_2 concentrations from six measuring stations in the Northern hemisphere (Point Barrow, Mauna Loa, La Jolla, Cape Kumukahi, Christmas Island, Alert Nunavut) and four stations in the Southern hemisphere (South Pole, American Samoa, Baring Head, Kermadec islands) were obtained from the Scripps Institution of Oceanography (2019). While monthly data collection at Mauna Loa started in 1957 (Keeling, 1960), all 10 stations were operational from 1985, and annual and quarterly averages for hemispheres were constructed for 1985–2018; sample means of 5-yearly annual averages are reported in Table 2. Lastly, data on sea water composition were available for 2000–2018 for over 380,000 oceanographic stations from the National Centers for Environmental Information (2019). Complete data pooled over time on dissolved oxygen levels, salinity, pH and water temperatures were available for approximately 284,000 stations. However, most stations had only a single observation for a given year so that annual averages were constructed for approximately 45,000 stations. Lastly, for investigating the dynamics of dissolved oxygen levels, $5° \times 5°$ grids were constructed from the station data. The sample means of variables are not reported in Tables 1 and 2 to save space and the findings are discussed below.

It would be helpful to briefly discuss the sample means and *t*-tests presented in Tables 1 and 2 for motivating the longitudinal analyses. First, at the country level in Table 1, there was a strong upward trend in mean CO_2 emissions for 3-yearly averages that increased from 142816 kton in 2001 to 199507 kton in 2016, that is, a 40% increase. By contrast, there were small but significant increases ($p < 0.05$) of 0.41° and 0.42° centigrade in 3-yearly averages of minimum and maximum temperatures respectively; annual temperature series exhibited cyclical patterns. Aerosol optical depths did not exhibit a significant trend and there was a slight decline in 2016. Lastly, there were significant increases in 3-yearly averages of GDP per capita and total population; industry value added as a percentage of GDP did not exhibit a significant trend.

In Table 2, the results for the Northern and Southern hemispheres showed significant increases between 1987 and 2017 in the 5-yearly averages of annual minimum and maximum ambient temperatures as well in quarterly temperatures. Application of *t*-tests showed that the changes in temperatures between 1987 and 2017 were significantly greater in the Northern hemisphere except for maximum temperatures in the fourth quarter. Also, there were significant increases between 1987 and 2017 in mean atmospheric CO_2 concentrations in both hemispheres. At every time point, CO_2 concentrations were higher in the Northern hemisphere and increased by approximately 1.5 ppm per annum. Overall, the sample means in Tables 1 and 2 suggest that it would be useful to conduct longitudinal analyses for CO_2 emissions and ambient temperatures at the country level, and model the effects of atmospheric CO_2 concentrations on annual and quarterly ambient temperatures compiled from weather stations in the hemispheres.

3 | THE ANALYTICAL AND METHODOLOGICAL FRAMEWORKS FOR CLIMATIC DATA

The availability of longitudinal remote sensing data and in situ measurements from weather and oceanographic stations provide valuable opportunities for statistical analyses for the formulation of

climate mitigation policies. As discussed in the ensuing subsections, it is important to formulate suitable empirical models and employ appropriate estimation procedures for drawing robust inferences. The present methodological approach is in the spirit of Fisher (1935) emphasizing the need for 'making theories elaborate' (e.g. Keogh & Cox, 2014), and the longitudinal analyses seem consistent with a 'teleological' approach to the 'entire historical process' underlying variables (von Neumann, 1961). In addition, the results of Mann and Wald (1943) and Koopmans et al., (1950) regarding possible misspecification of the actual distribution function of the errors were utilized, and 'quasi' maximum likelihood methods were employed for drawing robust inferences (Bhargava, 1987).

3.1 | On the use of temperature 'anomalies'

Although it is useful for illustrative purposes to subtract sample mean of temperatures estimated for a previous time interval for creating temperature 'anomalies', this procedure complicates testing of statistical hypotheses especially in finite samples. For analysing the stochastic properties of a temperature series (y_t) from a weather station with unknown mean μ and variance σ^2, we can postulate the simple model:

$$y_t = c + u_t \quad (t = 1, \ldots, T) \tag{1}$$

where c is a scalar and u's are random variables that follow a first-order autoregression:

$$u_t = \rho\, u_{t-1} + e_t \quad (t = 1, \ldots T) \tag{2}$$

Here, e_t are randomly distributed variables with mean zero and variance σ_e^2. Note that the problem of testing hypotheses regarding stochastic properties of u's, governed by the value of ρ, remains invariant with respect to groups of transformations of the type:

$$y^* = \gamma_0 y + h \tag{3}$$

where γ_0 and h are non-zero scalars (Lehmann, 1985). Under the second group of transformations, the T x1 vector of deviations of y's from their sample mean ($y_t - \bar{y}$) and the (T-1) × 1 vector of first differences of y's ($y_t - y_{t-1}$) are maximal invariants, that is, there is no information loss in utilizing them for statistical inferences. By contrast, subtracting the mean temperature ($\overline{\psi}$) computed for an earlier time interval does not account for variation in $\overline{\psi}$ in the confidence intervals and tests of hypotheses. In fact, if the linear regression model is postulated in Equation (1), the use of first differences (D y) does not entail any information loss due to the matrix identity (Sargan & Bhargava, 1983):

$$\Phi_0 = D' \left(DD'\right)^{-1} D \tag{4}$$

where Φ_0 is a T × T symmetric idempotent matrix converting variables into deviations from means:

$$\Phi_0 = I_T - (1/T)\, qq', \text{ q is a T} \times \text{1vector of ones}$$

and D is the (T-1) × T difference operator with elements $d_{ij} = -1$, if i = j; $d_{ij} = 1$, if j = I + 1; $d_{ij} = 0$, otherwise.

Lastly, when longitudinal data on N stations are available, Equation (1) can be replaced by:

$$y_{it} = c_i + u_{it}\, (i = 1, \ldots, N; t = 1, \ldots, T) \tag{5}$$

and the deviations of $y_{it's}$ from their respective station time means and station first differences are maximal invariants under a more general group of transformations. Some finite sample statistics for testing null hypotheses $\rho = 0$ (or $\rho = 1$ in the non-stationary case) versus the alternative hypothesis $0 < \rho < 1$ (Bhargava et al., 1982) for ambient temperatures in countries are presented in Section 5.

3.2 | Country and station level analyses for effects of CO_2 emissions and concentrations on temperatures

The diffusion of CO_2 from emission sources to the atmosphere is a complex process and its dynamics depends on geodesic variables such as latitude, longitude and elevation, as well as on air–sea fluxes, and wind patterns and speed (Kump et al., 2013). Moreover, oceans serve as major carbon sinks although there are uncertainties in the timing and magnitudes of absorption (Lombardi et al., 2006). Note that CO_2 emissions are calculated for countries by the International Energy Agency (2019) using reported activity levels in sectors such as construction, manufacturing, power industries and travel. Thus, it is not feasible to include such activities directly into empirical models for CO_2 emissions using country level data. Instead, one can include broad socioeconomic indicators such as GDP per capita and population levels in countries. Moreover, it is critical to utilize the data on atmospheric CO_2 concentrations that are ultimate indicators of CO_2 emissions. We proceed below by estimating empirical models using country level CO_2 emission data and model the effects of economic and demographic variables. Similar models were estimated for the average minimum and maximum temperatures from weather stations in the countries.

Next, since countries are often heterogeneous, effects of atmospheric CO_2 concentrations were modelled on annual and quarterly minimum and maximum ambient temperatures compiled at weather stations. Because the data were available for six and four stations, respectively, in the Northern and Southern hemispheres, average CO_2 concentrations were computed at the hemispheric levels. The average CO_2 concentrations can be introduced as 'macro' variables in empirical models, that is, at a given point in time, these variables assume the same values for the weather stations in the hemisphere. The trends underlying T annual (or quarterly) observations on temperatures can be modelled in dynamic random effects models by including a set of (T-2) time indicator variables or via spline functions (Section 4.2). Lastly, annual (or quarterly) atmospheric CO_2 concentrations can be included in the model in place of one or more time indicator variables and statistical tests for model adequacy can be applied (Sargan, 1980; Wald, 1947).

3.3 | Ocean deoxygenation and the use of 5° × 5° grids

The construction of gridded data for land and oceans often involves interpolations applied to the underlying sparsely distributed variables (Cressie, 1993). Moreover, simplifying assumptions on autocorrelation functions are sometimes necessary for analytical tractability (Wiener, 1949). While alternative techniques can be employed for construction of grids for oceans, it is important that the interpolations do not alter the dynamic properties of critical variables such as dissolved oxygen levels in oceans. Moreover, unlike weather stations, oceanographic stations are temporary and attempt to cover the vast oceans. For example, data from over 380,000 oceanographic stations were available during 2000–2018 from the World Ocean Database (National Centers for Environmental Information, 2019). Thus, it seems reasonable to construct averages for 5° × 5° grids without invoking interpolations that would be necessary for finer grids. While the data on sea water composition were available for 781

5° × 5° grids, due to the small number of repeated observations, two 10-yearly and three 7-yearly averages were constructed for 308 and 105 grids respectively.

4 | EMPIRICAL MODELS AND THEIR ESTIMATION

It is important to analyse the stochastic properties of dependent variables such as CO_2 emissions and ambient temperatures for developing comprehensive empirical models. For example, while annual data on CO_2 emissions by countries show a high degree of persistence, annual temperatures exhibit cyclical patterns and the estimated serial correlation coefficients were small (Section 5). Thus, it would be useful to employ 3- or 5-yearly averages using comprehensive empirical models. In the next subsections, empirical models for CO_2 emissions at country level, ambient temperatures from weather stations and dissolved oxygen levels in the oceans are outlined and the estimation methods are discussed.

4.1 | Country levels analyses for CO2 emissions and temperatures

The model for CO_2 emissions (Table 3) by ith country in time period t is given by:

$$\begin{aligned}\ln\left(\mathrm{CO_2\ emissions}\right)_{it} = {} & a_0 + a_1 \ln\left(\mathrm{GDP\ per\ capita}\right)_{it} + a_2 \ln\left(\mathrm{Total\ population}\right)_{it} \\ & + a_3 \ln\left(\mathrm{Industrial\ value\ added}\right)_{it} + a_4\left[\ln\left(\mathrm{GDP\ per\ capita}\right)\right]^2_{it} + a_5(\mathrm{Time\ period\ 3})_t \\ & + a_6(\mathrm{Time\ period\ 4})_t + a_7(\mathrm{Time\ period\ 5})_t + a_8(\mathrm{Time\ period\ 6})_t \\ & + a_9 \ln\left(\mathrm{CO_2\ emissions}\right)_{it-1} + u_{it}\ (i = 1, 2, \ldots, N; t = 2, 3, 4, 5, 6) \end{aligned} \quad (6)$$

Here, ln represents natural logarithms and this transformation was helpful for reducing heteroscedasticity. Also, 3-yearly averages were utilized for 163 countries in six time periods during 2000–2016. The dynamic model in (6) contained previous level of CO_2 emissions as an explanatory variable thereby enabling a distinction between short and long run effects of explanatory variables. Note that the coefficient of the lagged dependent variable (a_9) can be greater or equal to one so that such models can tackle non-stationarity of the dependent variable. Also, coefficients of explanatory variables in logarithms are the short run 'elasticities' (percentage change in dependent variable resulting from 1% change in the explanatory variable). In view of data availability at six time points, at most four time indicator variables can be included in (6); initial observations on CO_2 emissions were modelled using a 'reduced form' that had its own constant term (Section 4.4). Initially, a more general formulation for the model in (6) was estimated and it included squared of the total population and industrial value added variables. However, due to the modest sample sizes, these variables and some of the coefficients of time indicator variables were statistically insignificant and were dropped from the models in Table 3. Lastly, simple models for CO_2 emissions that included only time indicator variables are useful for investigating possible misspecification of the actual distribution function of the errors; the fourth-order moments of the fitted residuals in the six time periods are presented in Table 3.

The $u_{it's}$ in (6) were error terms that can be decomposed in a simple random effects fashion as:

$$u_{i\,t} = \delta_i + v_{it}\ (i = 1, 2, \ldots, N;\ t = 2, 3, \ldots, T) \quad (7)$$

TABLE 3 Maximum likelihood estimates of dynamic random effects models for 3-yearly averages for annual minimum and maximum temperatures explained by time indicators and CO_2 emissions for countries during 2000–2016[a]

Dependent variable:	**ln (CO_2), kton[b]**		**ln (Min temp), Kelvin**		**ln (Max temp), Kelvin**		**ln (Max temp), Kelvin**		**ln (CO2), kton**	
Explanatory variables:	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Constant	2.247	0.134	0.935	0.015	5.393	0.014	0.064	0.002	−1.292	0.197
Time period 3, 0–1	0.018*	0.009	0.0002	0.0003	0.0005	0.0003	-		-	
Time period 4, 0–1	0.020	0.015	−0.00001	0.0002	0.0006*	0.0003	-		-	
Time period 5, 0–1	0.040*	0.013	−0.00001	0.0002	0.0008*	0.0003	-		-	
Time period 6, 0–1	0.041*	0.012	0.0008*	0.0002	0.002*	0.0004	-		-	
ln (CO_2), kton	-		-		-		−0.0001	0.0001	-	
ln (GDP per capita), USD	-		-		-		-		0.168*	0.042
Industrial value added, % GDP	-		-		-		-		0.065*	0.011
ln (Total population), 1000,000	-		-		-		-		0.063*	0.004
[ln (GDP per capita)]^2	-		-		-		-		−0.008*	0.002
Lagged dependent variable	0.763*	0.017	0.835*	0.003	0.053*	0.003	0.989*	0.0003	0.929	0.003
Between/within variance, n	-		-		-		0.0002*	0.0001	-	
Within variance	-		-		-		0.00001	-	-	
2 x Maximized log-likelihood fn.	4079.58		11280.93	0.038	10942.42		9525.63		4125.47	

[a]Slope coefficients and standard errors are reported from estimating dynamic random effects models using six 3-yearly averages for CO_2 emissions, and minimum and maximum temperatures for 163 countries in six time periods.

[b]The fourth-order moments of residuals from the model for ln (CO_2) in the six time periods were, respectively, 2.480, 2.643, 2.641, 2.664, 2.806 and 2.705.

*$p<0.05$.

where δ_i are country-specific random effects that were assumed to be distributed with zero mean and constant variance, and v_{it} were distributed with zero mean and constant variance. A more general formulation for u_{it}, assuming that they were drawings from a multivariate normal distribution with a $T \times T$ dispersion matrix Ω was first employed. The validity of the decomposition (7) was tested using likelihood ratio statistics (Section 4.4). For example, if the errors v_{it} were serially correlated and/or the random effects (δ_i) had a time-varying coefficient, then the likelihood ratio tests were likely to reject the decomposition (7) in favour of the general random effects model with unrestricted dispersion matrix Ω.

4.2 | Models for 5-yearly averages of minimum and maximum temperatures in the hemispheres

The empirical model for 5-yearly averages of annual minimum (and maximum) temperatures from 7886 weather stations in the Northern hemisphere in seven time periods (Specification 1, Table 4) is given by:

$$\begin{aligned}\ln(\text{Minimum temperature})_{it} = {} & b_0 + b_1 \ln(\text{Latitude})_i + b_2 \ln(\text{Longitude})_i + b_3 \ln(\text{Elevation})_I \\ & + b_4(\text{Time period 3})_t + b_5(\text{Time period 4})_t + b_6(\text{Time period 5})_t \\ & + b_7(\text{Time period 6})_t + b_8(\text{Time period 7})_t + b_9 \ln(\text{Minimum temperature})_{it-1} \\ & + u_{it} \quad (i = 1, 2, \ldots, N; t = 2, \ldots, 7).\end{aligned} \tag{8}$$

Note that the geodesic coordinates account for the region characteristics. A similar model was estimated using the data from 691 weather stations in Southern hemispheres (Table 5). Because the coefficient of the Time period 3 indicator variable was not significantly different from zero in the empirical models and for reducing collinearity between time indicators and macro variables, atmospheric CO_2 concentrations replaced the Time period 4 variable in Specification 2 of Tables 4 and 5. Furthermore, Table 6 reports the results for 5-yearly averages of maximum ambient temperatures for quarters in the Northern and Southern hemispheres; the results for minimum temperatures were similar and are omitted. Lastly, while cubic spline functions were employed for capturing trends in annual temperatures in (8) (Wongsai et al., 2017), likelihood ratio statistics indicated a preference for analysing 3- or 5-yearly averages.

4.3 | Cross-sectional and longitudinal models for dissolved oxygen levels

The cross-sectional model estimated using the data for dissolved oxygen levels from 284,379 oceanographic stations (Specification 1, Table 7) is given by:

$$\begin{aligned}\ln(\text{Dissolved oxygen})_i = {} & c_0 + c_1(\text{Latitude})_i + c_2(\text{Longitude})_i + c_3(\text{Year})_I \\ & + c_4 \ln(\text{Depth})_i + c_5 \ln(\text{Salinity})_i + c_6 \ln(\text{pH})_i + c_7 \ln(\text{Temperature})_i \\ & + u_i \, (i = 1, 2, \ldots, N)\end{aligned} \tag{9}$$

Note that inclusion of the year of the observations partly captures the trends in dissolved oxygen levels although this model cannot account for heterogeneity across stations. A similar model (Specification 2) was estimated using yearly averages for oceanographic stations for 44,979 stations.

TABLE 4 Maximum likelihood estimates of dynamic random effects models for 5-yearly averages for annual minimum and maximum temperatures in the Northern hemisphere explained by geodesic, time indicators and atmospheric CO_2 variables during 1985–2018[a]

	ln (Minimum temperatures), Kelvin				ln (Maximum temperatures), Kelvin			
Dependent variable:	**Specification 1[c]**		**Specification 2**		**Specification 1[d]**		**Specification 2**	
Explanatory variables:	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Constant	2.884	0.002	2.897	0.002	2.832	0.001	2.695	0.002
ln (Latitude), 0–90	−0.020*	0.0002	−0.021*	0.0002	−0.023*	0.0002	−0.023*	0.0001
ln (Longitude), 0–180	−0.001*	0.0001	−0.001*	0.0002	−0.003*	0.0002	−0.003*	0.0001
ln (Elevation), m	−0.003*	0.0001	−0.003*	0.0001	−0.001*	0.0001	−0.001*	0.0001
Time period 4, 0–1	0.0004*	0.0001	-		0.001*	0.0001	-	
Time period 5, 0–1	0.0001*	0.0001	−0.0003*	0.0001	0.0003*	0.0001	−0.001*	0.0001
Time period 6, 0–1	0.0001*	0.0001	−0.0005*	0.0001	−0.0002*	0.0001	−0.002*	0.0001
Time period 7, 0–1	0.002*	0.0001	0.001*	0.0001	0.001*	0.0001	−0.001*	0.0001
ln (Atmospheric CO_2), PPM[b]	-		0.007*	0.0008	-		0.019*	0.001
Lagged dependent variable	0.504*	0.0003	0.494*	0.001	0.519*	0.001	0.522*	0.0004
Between/within variance, n	4.327*	0.055	4.498*	0.118	3.859*	0.058	3.788*	0.062
Within variance	0.00002		0.00002		0.00002		0.00002	
2 x Maximized log-likelihood fn.	578567.11		578535.48		570780.07		570751.62	

[a]Slope coefficients and standard errors are reported for seven 5-yearly averages of temperatures from 7886 weather stations in the Northern hemisphere.

[b]Atmospheric CO_2 concentrations were averaged over six stations in Northern hemisphere.

[c]The fourth-order moments of residuals from the model for ln (Minimum temperatures) in the seven time periods were, respectively, 4.185, 4.147, 4.304, 4.076, 4.629, 4.122 and 11.763.

[d]The fourth-order moments of residuals from the model for ln (Maximum temperatures) in the seven time periods were, respectively, 3.248, 3.993, 3.217, 3.646, 3.248, 3.152 and 7.073.

*$p<0.05$.

TABLE 5 Maximum likelihood estimates of dynamic random effects models for 5-yearly averages for annual minimum and maximum temperatures in the Southern hemisphere explained by geodesic, time indicators and atmospheric CO_2 variables during 1985–2018[a]

	ln (Minimum temperatures), Kelvin				ln (Maximum temperatures), Kelvin			
Dependent variable:	**Specification 1**		**Specification 2**		**Specification 1**		**Specification 2**	
Explanatory variables:	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Constant	3.807	0.0004	2.422	0.006	3.412	0.001	2.862	0.013
ln (Latitude), 0–90	0.056*	0.0004	0.045*	0.001	0.057*	0.001	0.050*	0.001
ln (Longitude), 0–180	−0.001*	0.0002	−0.0004	0.0003	0.001*	0.0002	0.001	0.0004
ln (Elevation), m	−0.002*	0.0001	−0.002*	0.0001	0.0001	0.0001	0.0001	0.0001
Time period 4, 0–1	0.0001	0.0001	-		0.001*	0.0001	-	
Time period 5, 0–1	0.0002*	0.0001	−0.0002	0.0002	0.001*	0.0001	−0.0001	0.0001
Time period 6, 0–1	0.0005*	0.0001	−0.0002	0.0002	0.001*	0.0001	−0.001*	0.0002
Time period 7, 0–1	0.002*	0.0001	0.001*	0.0002	0.002*	0.0001	0.0003	0.0002
ln (Atmospheric CO_2), PPM[b]	-		0.008*	0.002	-		0.016*	0.001
Lagged dependent variable	0.416*	0.0004	0.533*	0.001	0.359*	0.001	0.444*	0.001
Between/within variance, n	3.533*	0.154	2.148*	0.095	5.553*	0.213	4.067*	0.234
Within variance	0.00001		0.00001		0.00001		0.00001	
2 x Maximized log-likelihood fn.	55911.09		55889.37		53972.10		53974.47	

[a]Slope coefficients and standard errors are reported for seven 5-yearly averages of temperatures from 691 weather stations in the Southern hemisphere.

[b]Atmospheric CO_2 concentrations were averaged over the four stations in Southern hemisphere.

*$p<0.05$.

TABLE 6 Maximum likelihood estimates of dynamic random effects models for 5-yearly averages for quarterly maximum temperatures in the Northern and Southern hemispheres explained by geodesic, time indicators and atmospheric CO_2 variables during 1985–2018[a]

	ln (Maximum temperatures), Kelvin							
Dependent variable:	**Quarter 1**		**Quarter 2**		**Quarter 3**		**Quarter 4**	
Explanatory variables:	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Northern hemisphere								
Constant	4.945	0.0006	4.558	0.0002	3.543	0.001	3.687	0.005
ln (Latitude), 0–90	−0.056*	0.001	−0.031*	0.00003	−0.016*	0.0002	−0.037*	0.0002
ln (Longitude), 0–180	−0.009*	0.0003	−0.001*	0.0002	−0.003*	0.0001	−0.005*	0.0001
ln (Elevation), m	−0.003*	0.0003	−0.0007*	0.00001	−0.0001*	0.00001	−0.003*	0.0001
ln (Atmospheric CO_2), PPM[b]	0.004*	0.0002	0.019*	0.00003	0.012*	0.0001	0.016*	0.0003
Lagged dependent variable	0.165*	0.0003	0.203*	0.0001	0.380*	0.0003	0.361*	0.001
Between/within variance, n	31.287*	0.135	18.358*	0.102	14.701*	0.106	12.961*	0.097
Within variance, n	0.00001	-	0.00001	-	0.00001	-	0.00001	-
Southern hemisphere								
Constant	4.136	0.0001	4.423	0.0004	2.676	0.006	3.692	0.003
ln (Latitude), 0–90	0.055*	0.00002	0.087*	0.0001	0.057*	0.0004	0.060*	0.001
ln (Longitude), 0–180	0.003*	0.00003	−0.00007	0.00005	−0.0009*	0.0003	0.002*	0.0002
ln (Elevation), m	0.001*	0.00002	−0.0009*	0.0002	−0.0004*	0.0001	0.001*	0.0002
ln (Atmospheric CO_2), PPM[b]	0.018*	0.0001	0.012*	0.0001	0.008*	0.0006	0.020*	0.0004
Lagged dependent variable	0.213*	0.00002	0.148*	0.0001	0.481*	0.001	0.286*	0.0007
Between/within variance, n	18.876*	0.926	10.272*	0.621	5.831*	0.035	13.642*	0.207
Within variance, n	0.00001	-	0.00001	-	0.00001		0.00001	

[a]Slope coefficients and standard errors are reported for seven 5-yearly averages of temperatures from 7886 weather stations in the Northern hemisphere and 691 stations in the Southern hemisphere.

[b]Atmospheric CO_2 concentrations were averaged over six stations in the Northern hemisphere and four stations in the Southern hemisphere.

*$p<0.05$.

TABLE 7 Ordinary least squares and efficient estimates for static random effects models for dissolved oxygen levels from oceanographic stations and for constructed 5° × 5° degree grids during 2000–2018 explained by geodesic, temperature and acidification variables[a]

Dependent variable:	ln (Dissolved oxygen), µmol/kg									
Model:	Specification 1 (Stations: Monthly pooled)[b]		Specification 2 (Stations: Annual pooled)[b]		Specification 3 (Grids: Two time periods)[c]		Specification 4 (Grids: Three time periods)[d]		Specification 5 (Grids: Three time periods)[d]	
Explanatory variables:	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Constant	−6.012	0.373	7.806	0.359	−10.150	1.092	0.461	1.028	−17.434	4.112
Latitude	−0.002*	0.00003	0.0004*	0.0001	-		-		-	
Longitude	−0.0003*	0.00001	−0.0001*	0.00002	-		-		-	
Year	−0.004*	0.0002	−0.003*	0.0002	-		-		-	
ln (Depth), m	−0.051*	0.0006	−0.071*	0.001	−0.023*	0.001	−0.064*	0.009	−0.064*	0.008
ln (Salinity), n	−0.090*	0.001	−0.477*	0.001	−0.182*	0.039	−0.108*	0.033	−0.097*	0.030
ln (pH), n	10.192*	0.029	2.272*	0.027	8.233*	0.532	2.864*	0.514	11.614*	2.010
ln (Temperature +2), °C	−0.312*	0.001	−0.268*	0.002	−0.318*	0.022	−0.105	0.032	6.778*	1.696
ln (pH) x ln(Temperature+2)	-		-		-		-		−3.372*	0.829
R^2	0.48		0.60		-		-		-	
Number of stations or grids	284379		44979		308		105		105	

[a]Slope coefficients and standard errors are reported using data from oceanographic stations during 2000–2018.

[b]Cross-sectional models in Specifications 1 and 2 were estimated by ordinary least squares using, respectively, monthly and annual data.

[c]Longitudinal Specification 3 was estimated by a stepwise procedure for static random effects models using two 10-yearly averages for 5° × 5° grids.

[d]Specifications 4 and 5 were estimated by a stepwise procedure using three 7-yearly averages for the grids.

*$p<0.05$.

In addition, random effects models were estimated for 5° × 5° grids using two 10-yearly (Specification 3) and three 7-yearly (Specification 4) averages for investigating the dynamics of dissolved oxygen levels, while controlling for unobserved heterogeneity. Lastly, Specification 5 accounted for interactions between water temperatures and acidification using three 7-yearly averages, that is, the effect of temperatures for dissolved oxygen levels was postulated to depend on pH levels, and that of pH levels for dissolved oxygen on temperatures; the interactive effects were computed at their respective sample means (Section 7).

4.4 | Statistical and econometric methods

The dynamic random effects models for variables such as CO_2 emissions and ambient temperatures were estimated by maximum likelihood methods (Bhargava & Sargan, 1983); static random effects models for dissolved oxygen levels were estimated using efficient stepwise random effects models (Bhargava, 1991). The distribution theory assumed that the number of countries or stations (N) was large but the number of time periods (T) was fixed. The estimation techniques treated previous observations on dependent variables as 'endogenous', that is, correlated with the errors u_{it}. For exposition purposes, dynamic models can be written in a simultaneous equation framework by defining a 'reduced form' equation for initial observations and a 'triangular' system of (T-1) 'structural' equations for remaining time periods:

$$y_{i1} = \sum_{j=1}^{m} z_{ij}\zeta_j + \sum_{j=1}^{n}\sum_{k=1}^{T} \upsilon_{jk} x_{ijk} + u_{i1} \quad (i = 1, \ldots, N) \tag{10}$$

and

$$\underset{(T-1)\times T}{B}\ \underset{T\times N}{Y'} + \underset{(T-1)\times m}{C_1}\ \underset{m\times N}{Z'} + \underset{(T-1)\times nT}{C_2}\ \underset{nT\times N}{X'} = \underset{(T-1)\times N}{U'} \tag{11}$$

Here, Y, Z and X are, matrices containing, respectively, observations on the dependent, time invariant and time-varying explanatory variables; dimensions of the matrices are written below the respective symbols. B is a (T-1) x T lower triangular matrix of coefficients such that

$$B_{ii} = \alpha,\ B_{i,\,i+1} = -1,\ B_{ij} = 0 \text{ otherwise} \quad (i = 1, \ldots, T-1;\ j = 1, \ldots, T) \tag{12}$$

where α is the coefficient of lagged dependent variable. The matrices C_1 and C_2 contain coefficients of time invariant and time-varying variables respectively; the matrix U contains the error terms.

The errors affecting the model in Equations (10) and (11) were assumed to be independent drawings from a multivariate normal distribution with a T × T symmetric positive definite dispersion matrix Ω. The validity of the decomposition for u_{it} as in Equation (7) was tested using likelihood ratio statistics that were distributed for large N as Chi-squared variables with $[\{T(T+1)/2\} - 2]$ degrees of freedom. Note that the general formulation assuming multivariate normal distribution does not require the coefficient of the lagged dependent variable to be less than unity, that is, dependent variables can follow non-stationary processes. This is also the case where the random effects decomposition in (7) is employed and the variance of the initial observations (u_{i1}) and covariance between u_{i1} and remaining u_{it} (t = 2, 3, …, T) are unrestricted parameters. The validity of stationarity assumptions can be tested

using a likelihood ratio statistic that is distributed, for large N, as a Chi-squared variable with two degrees of freedom.

The 'concentrated likelihood functions' of the model in Equations (10) and (11) are obtained by analytically maximizing the log-likelihood function with respect to the 'nuisance' parameters ζ_j and $\upsilon_{j\,k}$ in the reduced form (10). Note that these formulations are more general than the dynamic models considered by Anderson and Hsiao (1981) that focused on time invariant explanatory variables. For dynamic models with time-varying explanatory variables (Anderson & Hsiao, 1982), it is essential to model the 'systematic' part of the initial observations (y_{i1}) using explanatory variables in the model; realizations of time-varying variables in different time periods are included in the reduced form (10) with unrestricted coefficients. Moreover, it may be useful in certain applications to treat y_{ir} ($r > 1$) as the initial observations for adequately capturing their dependence on previous realizations of time-varying variables (Bhargava, 1987). While such formulations will reduce the numbers of time observations in the estimation, adequacy of the model for initial observations enhances the computation of maximum likelihood estimates.

Furthermore, the assumption of normality of errors is not critical for the 'quasi' maximum likelihood estimates computed under the multivariate normality assumption; the estimated model parameters are consistent and asymptotically normally distributed provided that $(4 + \varepsilon)$th order moments ($\varepsilon > 0$) of the actual distribution function of the errors exist (Koopmans et al., 1950; Mann & Wald, 1943). Moreover, when the dispersion matrix Ω is unrestricted, the asymptotic standard errors of parameters obtained by maximizing the quasi likelihood function are unaffected by the actual distribution function (Bhargava, 1987). However, when the simple random effects decomposition (7) for errors is postulated, asymptotic standard errors of the estimated parameters of Ω depend on the fourth-order moments of actual distribution function of errors. The role of multivariate normality assumption for CO_2 emissions by countries and for the minimum and maximum ambient temperatures from weather stations in the Northern hemisphere is investigated, respectively, in Tables 3 and 4 by reporting the fourth-order moments of the fitted residuals.

Lastly, a numerical optimization routine (E04 JBF) (Numerical Algorithm Group, 1990) was used in a FORTRAN program for computing the parameters using concentrated likelihood functions for the cases where the dispersion matrix Ω was unrestricted, and where it corresponded to the random effects decomposition (7) with different assumptions on the initial observations. Asymptotic standard errors of the parameters were computed by numerically approximating the second derivatives of maximized log-likelihood functions. Note that while one can estimate dynamic models with 'fixed' effects (i.e. dummy variables) using data on large number of units (N) with only a few time observations, the increase in number of dummy variables with sample size leads to the problem of 'incidental parameters' (Neyman & Scott, 1948) vitiating certain properties of maximum likelihood estimators. From an empirical modelling standpoint, use of random effects models obviates the need for estimating the coefficients of large numbers of dummy variables thereby enhancing the efficiency of estimates.

5 | RESULTS FROM MODELS FOR CO_2 EMISSIONS AND AMBIENT TEMPERATURES USING COUNTRY-LEVEL DATA

Table 3 presents the maximum likelihood estimates from model for 3-yearly averages of CO_2 emissions, and minimum and maximum ambient temperatures for 163 countries for the period 2000–2016. Columns 2–4 report the results for CO_2 emissions and minimum and maximum temperatures from simple dynamic random effects models that included only a set of four time indicator variables. In

Column 5, effects of CO_2 emissions on maximum temperatures were investigated. Column 6 presents the results for CO_2 emissions that were explained by GDP per capita, total population and industrial value added expressed as a percentage of the GDP. Note that temperatures were converted into Kelvin by adding 273.15 and were then transformed into natural logarithms.

First, in the simple dynamic model for CO_2 emissions in Column 2, estimated coefficients of the indicators for Time periods 3, 5 and 6 were positive and significant ($p < 0.05$) indicating an upward trend in CO_2 emissions at the country level. The estimated coefficient of the lagged dependent variable was 0.76 and was significant. The likelihood ratio test rejected the null hypothesis that the errors affecting the model could be decomposed in a simple random effects manner, as in Equation (7). The fourth-order moments estimated from the residuals in Table 3 were in the interval [2.48, 2.80] and were not far from the value 3 for the normal distribution.

Second, in the models for minimum temperatures, the estimated coefficients of time indicators were only significant for Time period 6. The estimated coefficient of the lagged dependent variable was 0.84 and was significant. In contrast, in the model for maximum temperatures, the estimated coefficients of the indicator variables for Time periods 4, 5 and 6 were positive and significant but the estimated coefficient of the lagged dependent variable was small although significant. The empirical models for ambient temperatures for countries covered heterogeneous geographical regions that can affect the robustness of estimated parameters. Moreover, when annual data on minimum and maximum temperatures were employed, the coefficients of time indicator variables were generally not significant. In fact, Durbin–Watson type test statistics for longitudinal data with country fixed effects and time indicators (Bhargava et al., 1982) for minimum and maximum temperatures were close to 1.49 thereby rejecting the null hypothesis that $\rho = 0$ (or $\rho = 1$) in Equation (2). The estimated serial correlation coefficients for the errors were in the neighbourhood of 0.30.

Third, the model in Column 5 for maximum ambient temperatures included the CO_2 emissions by countries. However, the estimated coefficient was not statistically significant and the results were sensitive to model specification such as to the inclusion of time indicator variables. Moreover, aerosol optical depths were not significant predictors of maximum temperatures presumably because their effects were likely to be temporary (Samset, 2019) and 3-yearly averages were employed in the estimation.

Lastly, the model for CO_2 emissions at the country level showed significant effects of GDP per capita, total population and industrial value added. Nonlinearities with respect to GDP per capita were apparent and the CO_2 emissions increased with GDP per capita although at a declining rate. The estimated coefficient of the lagged dependent variable was 0.93 that was large and indicated persisting effects of GDP per capita, total population and industrial value added for CO_2 emissions. Note that the population density variable replaced total population in an alternative version of the model but its coefficient was not significantly different from zero. Overall, the estimated coefficients in Table 3 were sensitive to the specification of models presumably due to the heterogeneity across countries and the modest number of observations available; such factors are also likely to hamper the application of non-parametric methods.

6 | RESULTS FOR MINIMUM AND MAXIMUM ANNUAL AND QUARTERLY TEMPERATURES IN THE HEMISPHERES

The results from the models for 5-yearly averages of annual minimum and maximum ambient temperatures in the Northern and Southern hemispheres are presented in Tables 4 and 5 respectively. The results for 5-yearly averages of quarterly maximum temperatures are in Table 6; results for minimum quarterly temperatures were similar and are suppressed. Note that Tables 4 and 5 report the results for

Specifications 1 and 2 that included, respectively, four time indicator variables and where the atmospheric CO_2 concentrations replaced the indicator variable for Time period 4.

First, the estimated coefficients of time indicator variables in the models for annual minimum and maximum temperatures in Northern hemisphere (Specification 1) were positive and statistically significant for periods 4, 5, 6 and 7 corresponding to the years 2002, 2007, 2012 and 2017 respectively. This was also the case for time indicators for periods 5, 6 and 7 in the Southern hemisphere in Table 5; such increases were also evident in the analyses of 3-yearly averages of minimum and maximum temperatures. Note that the upward trends in ambient temperatures in the hemispheres were obscured by cyclical patterns when the data at annual intervals were employed.

Second, the estimated coefficients of atmospheric CO_2 concentrations in Specification 2 in the models for minimum and maximum temperatures were significant in the Northern and Southern hemispheres in Tables 4 and 5 respectively. Coefficients of the time indicator variables were often insignificant in Specification 2 reiterating the importance of CO_2 concentrations for ambient temperatures. While the estimated elasticities were in the interval [0.01, 0.02], the long run elasticities were twice as large in magnitude (see Section 8). Moreover, persistence of CO_2 in the atmosphere for decades suggests that reducing emissions will have lasting benefits for temperatures. Third, the estimated coefficients of explanatory variables were similar for random effects models with the unrestricted dispersion matrix Ω and where the simple random effects decomposition as in (7) was assumed for minimum and maximum temperatures in the hemispheres. While the likelihood ratio tests indicated a preference for the model with unrestricted Ω matrix, the between and within station variations in temperatures are reported in Tables 4 and 5, partly for comparisons with the results using quarterly data in Table 6.

Fourth, the estimated coefficients of lagged dependent variables in Specification 2 for minimum and maximum temperatures were generally similar for the two hemispheres (Tables 4 and 5). However, for quarterly temperatures in Table 6, coefficients of the lagged dependent variables were lower presumably due to disaggregation of the data for seasons. Moreover, the ratios of between to within variances were twice as large for quarterly temperatures in Table 6 than for annual temperatures reported in Tables 4 and 5. Fifth, the fourth-order moments for minimum and maximum temperatures in the Northern hemisphere in Table 4 were in the intervals [4.12, 11.76] and [3.15, 7.07] respectively. While the departures from normality seemed small in most time periods, there were increases in the last time period corresponding to the year 2017 reflecting more extreme minimum and maximum temperatures. Similar results were obtained for the Southern hemisphere but are not reported in Table 5.

Lastly, the models in Tables 4–6 were estimated with aerosol optical depths included as explanatory variables for the sub-period 2000–2018 for which the data were available. However, this variable was often estimated with small and positive coefficients that were not statistically significant. As noted previously, aerosols can have short-term effects for reducing temperatures (Magnus et al., 2011) and there are conceptual issues in aggregating data for different types of aerosols (Section 8). By contrast, atmospheric CO_2 concentrations were invariably significant in the models for annual and quarterly minimum and maximum ambient temperatures in the hemispheres.

7 | RESULTS FROM CROSS-SECTIONAL AND LONGITUDINAL MODELS FOR DISSOLVED OXYGEN LEVELS IN OCEANS

Table 7 reports the results from cross-sectional models for dissolved oxygen levels using the data from oceanographic stations in Specifications 1 and 2; longitudinal analyses for $5° \times 5°$ grids are presented in Specifications 3, 4 and 5. First, in Specifications 1 and 2, the year variable was estimated with a

significant negative coefficient indicating a decline in dissolved oxygen levels. Moreover, depth of measurements, and water salinity and temperatures were negatively and significantly associated with dissolved oxygen levels. By contrast, pH levels were estimated with a large positive coefficient in Specification 1 although the estimate was lower in Specification 2 for annual averages. Thus, higher temperatures and ocean acidification were significantly associated with lower dissolved oxygen levels.

Next, Specification 3 estimated the model using data for 308 5° × 5° grids in two time periods and the results were qualitatively similar to those for Specifications 1 and 2; there were some differences in the magnitudes of coefficients of depths, salinity, pH and temperatures. However, Specifications 1 and 2 did not control for heterogeneity across stations; this was tackled in Specification 3 via the variance covariance matrix of the errors affecting dissolved oxygen levels. Note that the results from dynamic version of the models containing previous dissolved oxygen levels were similar to those for Specification 3. However, due to small number of time observations, it was preferable to focus on static models with serially correlated and/or heteroscedastic errors in Specifications 3, 4 and 5.

Furthermore, Specification 4 and 5 were estimated using the data on 105 5° × 5° grids at three time points; the results for Specifications 3 and 4 were similar. Note that Specification 5 included an interactive term between pH levels and temperatures and its coefficient was significant. The partial derivatives of dissolved oxygen levels with respect to pH and temperatures, respectively, were:

$$11.614 - 3.372 \ln\ (\text{Temperature} + 2.0) \tag{13}$$

and

$$6.778 - 3.372 \ln (\text{pH}) \tag{14}$$

Thus, the points of inflexion with respect to natural logarithms of pH and temperatures were 3.44 and 2.01 respectively. These computations translated into pH levels of 7.46 and temperatures of 29.31° centigrade at the respective sample means. Thus, the results from Specification 5 implied that the effects of higher temperatures for dissolved oxygen levels were negative if pH levels were greater than 7.46 that was generally the case; minimum pH level for the grids was 7.19. Furthermore, the effects of higher pH levels for dissolved oxygen were positive if temperatures were below 29.31° centigrade that was invariably the case. Thus, pH levels above 7.46 were likely to dampen the adverse effects of higher temperatures for dissolved oxygen levels. While these computations are suggestive of the underlying nonlinearities and interactions, they underscore the need for estimating threshold effects using data on larger numbers of grids and for longer time spans.

8 | CONCLUSION AND DISCUSSION

The formulation of evidence-based climate mitigation policies is an intricate task partly due to data limitations and methodological aspects. This paper tackled several issues and presented elaborate analyses of longitudinal data for understanding the dynamic inter-relationships between CO_2 emissions and atmospheric concentrations, ambient temperatures and ocean deoxygenation. First, results at the country level in Table 3 showed the importance of economic activity and population levels for CO_2 emissions. Thus, commitments by countries for reducing greenhouse gas emissions are critical (United Nations, 2019) and should be monitored due to the negative externalities imposed by emitters on the global environment.

Second, analyses of the data from over 8,500 weather stations on minimum and maximum annual and quarterly ambient temperatures in the hemispheres provided useful insights. Ambient temperatures

increased at faster rates between 1985 and 2018 in the Northern hemisphere, and the longitudinal analyses invariably showed that atmospheric CO_2 concentrations were positively and significantly associated with 5-yearly averages of annual and quarterly temperatures (Tables 4–6). For illustrative purposes, CO_2 concentrations increased between 1987 and 2017 (Table 2) from 349.94 to 405.63 ppm constituting a 15.91% increase. From Table 4, the long run elasticity of maximum temperatures in Kelvin with respect to CO_2 concentrations in the Northern hemisphere was approximately 0.04. Thus, the empirical model using data from weather stations and controlling for latitudes, longitudes and altitudes predicted an increase in maximum temperatures of 0.64% from the mean in 1987, that is, of 1.85 Kelvin.

Third, for comparisons with the implications of Arrhenius (1896), elasticities of global average annual maximum temperatures with respect to CO_2 concentrations recorded in Mauna Loa were estimated using time series data for 1960–2018 and the results were sensitive to model specification. For example, in a simple model with a linear time trend, the elasticity of maximum temperatures in Kelvin with respect to CO_2 concentrations was 0.056. However, coefficients of lagged maximum temperatures in dynamic versions of the models were generally insignificant. While computations using average temperatures are indicative of the role of atmospheric CO_2 concentration for global warming (Sherwood et al., 2020), the empirical results in Tables 4–6 using disaggregated longitudinal data invariably showed significant effects of CO_2 concentrations on ambient temperatures. The longitudinal analyses were advancements over the use of annual temperature 'anomalies' that introduce a degree of arbitrariness in statistical analyses and ignore the seasonal patterns; it would be useful to tabulate recent quarterly means for the hemispheres in climate databases. Analyses of *in situ* and remote sensing data are likely to provide deeper insights for climate turning points than projections from simple models (Bhargava, 2018).

Fourth, aerosol optical depths were not significantly associated with ambient temperatures presumably because aerosols include diverse pollutants such as black carbon, sulphates and nitrates whose long-term effects may be ambiguous (Samset, 2019). Fifth, the lags between CO_2 emissions and absorption by oceans depend on factors such as sea–air fluxes, geodesic coordinates and wind patterns; ultimate effects of CO_2 emissions are reflected in atmospheric concentrations and in ocean acidification. It was seen in Table 7 that higher water temperatures and lower pH levels were negatively and significantly associated with dissolved oxygen levels. The empirical results from $5° \times 5°$ grids provide evidence of interactive effects of temperatures and pH for dissolved oxygen levels, while controlling for heterogeneity across grids. The problems facing marine life seem urgent since their metabolic activity is increased in lower oxygen and higher temperatures environments thereby creating a vicious circle (Breitburg et al., 2018).

Lastly, while it was noted in the Figure 1 that higher temperatures and ocean acidification are likely to reduce sea ice and glacier thicknesses, such issues and feedback effects (Heimann & Reichstein, 2008) need to be investigated in future research using thickness data (Copernicus Climate Change Service, 2019; GlaThiDa Consortium, 2019). The analyses will present methodological challenges such as the need for incorporating regional effects of pollution for glacier thickness, and difficulties in assessing ocean acidification when they are covered by ice. The evidence in this paper underscored the need for adopting a long-term policy perspective because greenhouse gases remain in the atmosphere for several decades. Greater emphasis should be placed on CO_2 absorption by the oceans and on dumping of waste that can raise water temperatures and accelerate sea ice melting. In addition, analyses of data on emissions from agriculture (Food & Agriculture Organization, 2019; World Data Centre for Greenhouse Gases, 2021) will be useful for the formulation of climate and agriculture policies. Eliminating fossil fuel subsidies and devoting greater resources for renewable energy generation will be helpful for reducing greenhouse gas emissions. From an evidence-based climate mitigation perspective, policies at the country and global levels need to strike a balance between economic

activity and population growth for minimizing environmental degradation and for enhancing long-term sustainability.

ACKNOWLEDGEMENTS

While retaining the responsibility for the analyses, the author thanks Juan Echenique for his valuable help in extracting *in situ* and remote sensing data, and Dr. Jing Li for sharing the data on aerosol optical depths. The revisions of the paper have substantively benefited from the helpful comments from two reviewers, an Associate Editor and the Editor.

ORCID

Alok Bhargava https://orcid.org/0000-0002-0399-0949

REFERENCES

Anderson, T.W. & Hsiao, C. (1981) Estimation of dynamic models with error components. *Journal of the American Statistical Association*. 76, 598–606.

Anderson, T.W. & Hsiao, C. (1982) Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1), 47–82.

Arrhenius, S. (1896) On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine*, 41, 237–276.

Bhargava, A. (1987) Wald tests and systems of stochastic equations. *International Economic Review*, 28, 789–808.

Bhargava, A. (1991) Identification and panel data models with endogenous regressors. *Review of Economic Studies*, 58, 129–140.

Bhargava, A. (2018) Climate variability, rice production and groundwater depletion in India. *Environmental Research Letters*, 13(3), 034022. https://doi.org/10.1088/1748-9326/aade9

Bhargava, A., Franzini, L. & Narendranathan, W. (1982) Serial correlation and the fixed effects model. *Review of Economic Studies*, 49(4), 533–549.

Bhargava, A. & Sargan, J.D. (1983) Estimating dynamic random effects models from panel data covering short time periods. *Econometrica*, 51(6), 1635–1660.

Breitburg, D., Levin, L.A., Oschlies, A., Grégoire, M., Chavez, F.P., Conley, D.J. et al. (2018) Declining oxygen in the global ocean and coastal waters. *Science*, 359(46), 1–11.

Copernicus Climate Change Service. (2019) Sea ice area. https://climate.copernicus.eu/sea-ice-cover-october-2019

Cressie, N. (1993) *Statistics for spatial data*. New York: John Wiley & Sons.

Emissions Database for Global Atmospheric Research. (2019) Emissions data and maps. https://edgar.jrc.ec.europa.eu/overview.php?v=booklet2018

Fisher, R.A. (1935) *The design of experiments*. Edinburgh: Oliver and Boyd.

Food and Agriculture Organization. (2019) FAOSTAT database. http://www.fao.org/faostat/en/#data/RF

GlaThiDa Consortium (2019) *Glacier thickness database 3.0.1*. Zurich, Switzerland: World Glacier Monitoring Service. https://doi.org/10.5904/wgms-glathida-2019-03

Heimann, M. & Reichstein, M. (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, 451(7176), 289–292.

Intergovernmental Panel on Climate Change (2014) Climate Change 2014: Mitigation of climate change. http://www.ipcc.ch/report/ar5/wg3/

International Energy Agency. (2019). Data and statistics. https://www.iea.org/data-and-statistics?country=WORLD%26fuel=Energy%2520supply%26indicator=Coal%2520production%2520by%2520type"

Keeling, C. (1960) The concentration and isotopic abundances of carbon dioxide in the atmosphere. *Tellus*, 12(2), 200–203.

Keeling, R., Kortzinger, A. & Gruber, N. (2010) Ocean deoxygenation in a warming world. *Annual Review of Marine Science*, 2(1), 199–229.

Keogh, R. & Cox, D. (2014) *Case-control studies*. Cambridge: Cambridge University Press.

Koopmans, T., Rubin, H. & Leipnik, R. (1950) Measuring the equation systems of dynamic economics. In: Koopmans, T.C. (Ed.) *Statistical inference in dynamic economic models*. New York: John Wiley and Sons, pp. 54–237.

Kump, L., Kasting, J. & Crane, R. (2013) *The earth system*. London: Pearson Education Inc.
Lehman, E. (1985) *Testing statistical hypotheses*, 2nd edition. New York: John Wiley & Sons.
Li, J., Li, X., Carlson, B., Kahn, R., Lacis, A., Dubovik, O. & et al. (2016) Reducing multisensory satellite monthly mean aerosol optical depth uncertainty: 1. Objective assessment of current AERONET locations. *Journal of Geophysical Research Atmospheres*, 121, 13609–13627.
Linacre, E. (1992) *Climate data and resources*. London: Routledge.
Lombardi, S., Altunina, L. & Beaubien, S. (2006) *Advances in the geological storage of carbon dioxide*. Netherlands: Springer.
Magnus, J., Melenberg, B. & Muris, C. (2011) Global warming and local dimming: the statistical evidence. *Journal of the American Statistical Association*. 106(494), 452–468.
Mann, H. & Wald, A. (1943) On the statistical treatment of linear stochastic equations. *Econometrica*, 11, 173–220.
National Centers for Environmental Information (2018) Land-based station data. https://www.ncdc.noaa.gov/data-access/land-based-station-data
National Centers for Environmental Information (2019). World Ocean Database. https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.nodc:0164586#__sid=js0
Neyman, J. & Scott, E. (1948) Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1–32.
Numerical Algorithm Group. (1990) *Numerical algorithm group library*. Mark 13. Oxford: Oxford University.
Pigou, A. (1932) *The economics of welfare*, 4th edition. London: Macmillan.
Robeson, S. (1997) Spherical methods for spatial interpolation: Review and evaluation. *Cartography and Geographical Information Systems*, 24(1), 3–20.
Samset, B. (2019) Aerosols and climate. Oxford Research Encyclopedia, Climate Science. http://oxfordre.com/climatescience
Sargan, J.D. (1980) Some tests of dynamic specification for a single equation. *Econometrica*, 48, 879–898.
Sargan, J.D. & Bhargava, A. (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica*, 51, 153–174.
Scripps Institution of Oceanography. (2019) Scripps CO2 Program Data. https://library.ucsd.edu/dc/collection/bb3381541w
Sherwood, S.C., Webb, M.J., Annan, J.D., Armour, K.C., Forster, P.M., Hargreaves, J.C. et al. (2020) An assessment of earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4), 1–92. https://doi.org/10.1029/2019RG000678
Stern, N. (2006) *The economics of climate change*. Cambridge: Cambridge University Press.
Torge, W. (2001) *Geodesy*, 3rd edition. New York: Walter de Gruyter.
United Nations (2019) United Nations Climate Change Conference, 25. https://unfccc.int/cop25
von Neumann, J. (1961) Methods in the physical sciences. In: Taub, A. (Ed.) *John von Neumann: Collected works*. VI, Oxford: Oxford University Press.
Wald, A. (1947) *Sequential analysis*. New York: Dover Publications.
Wang, C.Y., Wang, S. & Carroll, R. (1997) Estimation in choice-based sampling with measurement error and bootstrap analysis. *J. Econometrics*, 77(1), 65–86.
Wiener, N. (1949) *Extrapolation, interpolation, and smoothing of stationary time series*. New York: John Wiley & Sons.
Wongsai, N., Wongsai, S. & Huete, A. (2017) Annual seasonality extraction using cubic spline function and decadal trend in temporal daytime MODIS LST data. *Remote Sensing*, 9, 1254. https://doi.org/10.3390/rs9121254.
World Bank (2018) World Development Indicators. https://data.worldbank.org/products/wdi
World Data Centre for Greenhouse Gases (2021) Data Archive. Available from https://gaw.kishou.go.jp/

How to cite this article: Bhargava A. (2022) Econometric modelling of carbon dioxide emissions and concentrations, ambient temperatures and ocean deoxygenation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185, 178–201. https://doi.org/10.1111/rssa.12732