DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

A statistical model of the global carbon budget

Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman

CREATES Research Paper 2020-18

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

A STATISTICAL MODEL OF THE GLOBAL CARBON BUDGET
(UPDATE TO GLOBAL CARBON BUDGET 2020)

MIKKEL BENNEDSEN, ERIC HILLEBRAND, AND SIEM JAN KOOPMAN

ABSTRACT. We propose a dynamic statistical model of the Global Carbon Budget as represented in the annual data set made available by the Global Carbon Project, covering the sample period 1959–2019. The model connects four main objects of interest: atmospheric carbon dioxide (CO_2) concentrations, anthropogenic CO_2 emissions, the absorption of CO_2 by the terrestrial biosphere (land sink), and by the ocean and marine biosphere (ocean sink). The model captures the global carbon budget equation, which states that emissions not absorbed by either land or ocean sinks must remain in the atmosphere and constitute a flow to the stock of atmospheric concentrations. Emissions depend on global economic activity as measured by World Gross Domestic Product while sink activities depend on the level of atmospheric concentrations and the Southern Oscillation Index. We use the model to determine the time series dynamics of atmospheric concentrations, to assess parameter uncertainty, to compute key variables such as the airborne fraction and sink rate, to forecast the Global Carbon Budget components from forecasts of World Gross Domestic Product and Southern Oscillation, and to conduct scenario analysis based on different possible future paths of global economic activity.

JEL codes: C32, C49, C51, C52, C53

Some Keywords: Global Carbon Budget, World Gross Domestic Product, CO_2 emissions, CO_2 concentrations, El Niño Southern Oscillation, airborne fraction, sink rate, climate economics, climate econometrics.

ACKNOWLEDGMENTS: We are grateful for comments from Jurgen Doornik, David Hendry, Søren Johansen, Andrew Martinez, and Felix Pretis, as well as from participants at the fourth conference on Econometric Models of Climate Change in Milan in 2019 and the Climate Econometrics online seminar 2020. EH posed this modeling problem as Case A of the Econometric Game in 2019 at the University of Amsterdam, and we are indebted to the organizers and participants of the Game. Special thanks go to Peter Boswijk and Joep Keuzenkamp. MB and EH acknowledge support from the Independent Research Fund Denmark.

Date: May 10, 2021.

Contents

1. Introduction

In this paper, we propose a dynamic statistical model for the time series collected in the annual Global Carbon Budget (GCB), collated and maintained by the Global Carbon Project (Friedlingstein et al., 2020). The model connects atmospheric CO_2 concentrations, anthropogenic emissions, and uptake by the terrestrial biosphere (land sink) and by the ocean and the marine biosphere (ocean sink). It includes the global carbon budget equation as a cornerstone. The model further specifies both sinks as linear in atmospheric concentrations, and it specifies emissions as a random walk with drift, either as a constant or as determined by economic activity. The dynamics of concentrations are determined by the global carbon budget equation. Since concentrations determine sinks activity in turn, the model captures synchronicity in the determination of key variables.

The model allows for the data-driven study of the global carbon cycle employing a relatively small model. By employing the GCB data set, it facilitates using observational data as well as the output of several large-scale general circulation models (GCM). Parameter uncertainty can be evaluated by way of statistical standard errors, in contrast to GCM or small-scale emulators. Parameter estimates are based on the statistical analysis of past data. In order to capture changes in the fuel mix and the dependence of emissions on economic activity in projections that extend far into the future, we consider scenarios for time-varying coefficients in the emissions equation. The model is linear and therefore incurs approximation errors, for example in the dependence of the sinks on concentrations.

We analyze the time series properties of the components of the GCB. The salient property of all series is that they are upward trending. From a number of unit roots tests and from theoretical relations, we extract a list of requirements that a statistical model of the data has to satisfy. The model we propose satisfies these requirements, and from its system equations we derive the time series dynamics of atmospheric CO_2 concentrations. This result shows that atmospheric concentrations follow single unit root dynamics that are, however, numerically quite close to a second unit root, and that they approach a second unit root as atmospheric concentrations increase. This finding might appear surprising at first sight, since the stock of atmospheric concentrations contains a cumulated sum of the unit root emissions process; however, as we show below, the result obtains because of the internal dynamics of the carbon cycle system.

We propose and compare two different specifications of the statistical model. Model I satisfies the list of requirements and involves the global time series variables of the GCB. Model II further includes a global variable of economic activity (World Gross Domestic Product, GDP) as a driver of emissions and the Southern Oscillation Index (SOI) variable as a proxy for the El-Niño/Southern

Oscillation (ENSO) cycle in the sinks dynamics. In addition, we include a number of dummy variables for specific unusual events in the relation of World GDP and emissions. We present the estimation results that include parameter estimation uncertainty measures.

In simulations of our model, we show the small-sample properties of the statistical estimators and their robustness to the model assumptions. We present and discuss the budget imbalance, the airborne fraction, and the sink rate implied by the models. We conduct a validation experiment where we compute the time series of the GCB using only data on World GDP and SOI for the last ten years of the sample, and we show that the model can produce confidence intervals from these two inputs that cover most observations of the GCB components. We nowcast and forecast emissions, sinks activity, and atmospheric concentrations from forecasts of World GDP and SOI for the period 2020–2022. Finally, we present an analysis of different scenarios for the path of the GCB components until the year 2050. The model proposed in this paper is well-suited for such scenarios analyses, since it integrates the carbon cycle (through the GCB components) with economic activity (through World GDP growth) and thus allows for a coherent way to construct projections of the dynamics of the global carbon cycle, conditional on an economic scenario. By considering different future scenarios for World GDP and for the dependence of emissions on GDP, we find that abatement levels needed to limit the amount of future global warming are very sensitive to the path of future economic growth. These findings highlight the need for decoupling emissions from economic growth (Haberl et al., 2020) and can thus have implications for commitments made by the global community, such as those arising from the Paris Agreement of 2015.

The remainder of the paper is organized as follows. Section 2 introduces the model, discusses central assumptions, and derives a solution for atmospheric concentrations of CO_2 from the system model equations. Section 3 presents Monte Carlo simulation results for the model. Section 4 describes the data set that we use and discusses time series properties of the data series. Section 5 presents estimation results, discusses residual diagnostics, and introduces an extension of the basic model to include World GDP and SOI, including estimation results and residual diagnostics for this extended model. Section 5.5 discusses limitations of the model as well as the implied budget imbalance, the airborne fraction, and the sink rate. Section 6 presents a validation analysis where the model is estimated on the first 51 years of data and validated on the last 10 years, a forecast exercise for three years out of sample, and scenario analyses until the year 2050. Section 7 concludes.

2. A statistical dynamic model for the global carbon budget

The statistical global carbon budget model proposed in this study is designed for four variables: atmospheric CO_2 concentrations (C_t^*), anthropogenic emissions (E_t^*), CO_2 uptake by the terrestrial biosphere (land sink, $S\_LND_t^*$) and CO_2 uptake by the ocean and marine biosphere (ocean sink, $S\_OCN_t^*$). The flow series E_t^*, $S\_LND_t^*$, and $S\_OCN_t^*$ are measured in gigatons of carbon (GtC) per year; the stock series C_t^* is measured in GtC. The foundation of our statistical model is the global carbon budget equation

$$G\_ATM_{t+1}^* := C_{t+1}^* - C_t^* = E_{t+1}^* - S\_LND_{t+1}^* - S\_OCN_{t+1}^*,$$

where G_ATM^* represents the change in atmospheric concentrations in GtC per year. The budget equation expresses that emissions not absorbed by either the land or ocean sink constitute a flow to the stock of atmospheric concentrations. The budget equation further implies a dynamic process for concentrations as given by

$$C_{t+1}^* = C_t^* + E_{t+1}^* - S\_LND_{t+1}^* - S\_OCN_{t+1}^*. \tag{1}$$

The dynamic equations for the emission and sink variables are given next. The dynamic evolution of emissions E^* is assumed to follow a random walk process with drift as given by

$$E_{t+1}^* = E_t^* + d + X_t^E, \tag{2}$$

where d is the constant drift or growth term and X_t^E is a stationary random innovation. The sinks are assumed to be linear in atmospheric concentrations

$$S\_LND_{t+1}^* = c_1 + \frac{\beta_1}{C_{1750}} C_{t+1}^*, \tag{3}$$

$$S\_OCN_{t+1}^* = c_2 + \frac{\beta_2}{C_{1750}} C_{t+1}^*, \tag{4}$$

where c_i, $i = 1, 2$, are constant intercepts, and the slopes $\beta_i / C_{1750} > 0$ represent the fractions of concentrations that are absorbed annually by the sinks. We discuss these linearity assumptions and the scaling of the parameters by pre-industrial concentrations C_{1750} =593.43GtC or 279ppm in the subsections below. See also the discussion of limitations in Section 5.5. We emphasize that X_t^E is the only source of randomness driving the state variables (1)–(4) which implies that the randomness of the GCB is only due to X_t^E.

The observed time series for these four variables are denoted with the same variable indicator but without the asterisk. Annual observations of the four variables at a global level are provided by the Global Carbon Project (Friedlingstein et al., 2020) for the period 1959 through 2019. Atmospheric concentrations are instrumental measurements. Emissions are computed from the use of fossil energy carriers as reported by countries' authorities. To compute our variable "anthropogenic emissions", we take fossil fuel emissions plus land use change emissions minus cement carbonation sink. These observations are typically subject to measurement errors and other irregularities due to data collection. The observations of the land and ocean sinks, on the other hand, are averages over the output of several GCM / Earth system models selected by the Global Carbon Project. We treat these observations statistically as data in our model. For the sink processes, the model should be understood as an approximation to the more complex climate models, and it only captures parts of the more detailed interrelations that are captured in the climatologically and mathematically more involved climate models, see the discussions below.

The deviations of the observed variables (without asterisk) from the unobserved model variables (with asterisk) are therefore a mix of measurement errors (in particular, for concentrations and emissions) and approximation errors (in particular, for land and ocean sink variables). The statistical treatment of GCM model output has recently been discussed in various contributions to the literature; see, for example, Castruccio et al. (2013), Holden et al. (2015), Castruccio and Guinness (2017) and Guinness and Hammerling (2018).

Given the dynamic specifications of the key four model variables, we complete the statistical model for the observed variables with the so-called measurement equations as given by

$$C_t = C_t^* + X_{1,t}, \tag{5}$$

$$S\_LND_t = S\_LND_t^* + X_{2,t}, \tag{6}$$

$$S\_OCN_t = S\_OCN_t^* + X_{3,t}, \tag{7}$$

$$E_t = E_t^* + X_{4,t}, \tag{8}$$

where $X_{1,t}$ and $X_{4,t}$, associated with C_t and E_t, can be mainly regarded as measurement errors and can be treated in the usual way as disturbances, and where $X_{2,t}$ and $X_{3,t}$, associated with S_LND_t and S_OCN_t, mainly represent processes that are not captured by our statistical model.

Under normality assumptions for the innovations in the $X_{i,t}$, $i = 1, \ldots, 4$, and X_t^E processes, the model variable and measurement equations constitute a multivariate linear Gaussian state space

model; see the discussion in Appendix A. The dynamics of the four unobserved model variables (with asterisks) are placed in the state vector and are driven by the single random innovation term X_t^E. We define autoregressive processes for the dynamic processes $X_{i,t}$, $i = 1, \ldots, 4$, and X_t^E, and place them in the state vector as well. We further allow for correlation between the innovations that are driving the processes $X_{i,t}$ and $X_{j,t}$, denoted as r_{ij}, for selected pairs (i, j). The state space framework facilitates the estimation of the parameters by the method of maximum likelihood, where the likelihood is evaluated by the Kalman filter, and its maximization relies on a numerical opimization method. The predicted, filtered, and smoothed estimates for the entries of the state vector are obtained by the Kalman filter and related methods; see Durbin and Koopman (2012) for a treatment. In Section 3, we verify the accuracy of our proposed estimation approach in a Monte Carlo study, both under correct and incorrect model specifications.

In the remaining part of this section, we discuss the linearity assumptions in (3) and (4), and we derive the dynamics of atmospheric concentrations C^* as the solution of the system of equations given by (1), (2), (3), and (4). These subsections refer to the *-processes throughout, but we suppress the superscript for notational ease.

2.1. **The land sink as a linear function of atmospheric concentrations.** Bacastow and Keeling (1973) suggest that the relationship between S_LND_t and atmospheric CO_2 concentrations C_t follows

$$S\_LND_t = \beta \log(C_t/C_{1750}), \tag{9}$$

where β is a constant and $C_{1750} = 593.43$ GtC is the pre-industrial atmospheric concentration. This formula captures the fertilization effect, whereby increased CO_2 concentrations lead to increased net primary production of plants. A second-order Taylor expansion of this formula around C_{1750} yields

$$S\_LND_t \approx \frac{\beta}{C_{1750}}(C_t - C_{1750}) - \frac{\beta}{2C_{1750}^2}(C_t - C_{1750})^2. \tag{10}$$

If $C_t - C_{1750}$ is small relative to C_{1750}, the second-order term is small, and a linear relation between S_LND_t and C_t is a good approximation. If C_t is large relative to C_{1750}, the second-order term becomes important. We can rewrite the Taylor expansion as

$$S\_LND_t \approx c_L + k_L(C_t)C_t, \tag{11}$$

where

$$c_L = -\frac{3}{2}\beta, \qquad k_L(C_t) = \frac{2\beta}{C_{1750}} - \frac{\beta}{2C_{1750}^2}C_t,$$

and $k_L(C_t)$ is decreasing in C_t. At the beginning of the sample, the atmospheric concentration is $C_{1959} = 670.83$GtC. The time series of concentrations is then given from the data by

$$C_t = C_{1959} + \sum_{\tau=1959}^{t} G\_ATM_\tau,$$

ending in $C_{2019} = 874.83$GtC. Therefore, in the beginning of the sample, $k_L(C_{1959}) = 0.0024\beta$, and at the end, $k_L(C_{2019}) = 0.0021\beta$. The assumption of a land sink that grows linearly in concentrations, $k_L(C_t) = k_L$, is thus not unreasonable on the sample. The linearity assumption was suggested already in Bacastow and Keeling (1973, p. 94) and it is discussed and applied, for example, in the context of the airborne fraction and sink rate of CO_2, in Raupach et al. (2014), Raupach (2013), Gloor et al. (2010), Canadell et al. (2007b), and Bennedsen et al. (2019b).

An alternative expression for the fertilization effect was put forward in Gifford (1993):

$$S\_LND_t = NPP_t - NPP_{1750} = \frac{a(C_t - C_b)}{1 + b(C_t - C_b)} - \frac{a(C_{1750} - C_b)}{1 + b(C_{1750} - C_b)},$$

where NPP is net primary production and C_b is the atmospheric concentration where net primary production is zero ($C_b = 65.9$GtC or 31ppm in Gifford (1993)). The parameter a plays a comparable role to the parameter β in the Bacastow-Keeling formula. A Taylor expansion of the formula to second order in the point $C_t = C_{1750}$ yields:

$$\begin{aligned} S\_LND_t &\approx \frac{a}{(1 + b(C_{1750} - C_b))^2}(C_t - C_{1750}) - \frac{ab}{(1 + b(C_{1750} - C_b))^3}(C_t - C_{1750})^2 \\ &\approx k_G(C_t)(C_t - C_{1750}), \end{aligned}$$

and

$$k_G(C_t) = \frac{a}{(1 + b(C_{1750} - C_b))^2} - \frac{ab(C_t - C_{1750})}{(1 + b(C_{1750} - C_b))^3}$$

decreasing in C_t. Using a typical parameter configuration employed in Gifford (1993), $a = 0.461$ and $b = 0.005$, we get $k_G(C_{1959}) \approx 0.08$ and $k_G(C_{2019}) \approx 0.06$. Thus, the Gifford formula has more curvature for typical parameter values, and the linear approximation is less accurate than for the Bacastow-Keeling specification.

Both the Bacastow-Keeling and the Gifford formulae imply that a regression of S_LND on C_t/C_{1750} and a constant should yield estimated intercepts and regression coefficients that are equal in magnitude with opposite signs, of course up to small sample error. (For the Bacastow-Keeling formula, the first-order term decomposes into $-\beta$ plus $\beta\, C_t \,/\, C_{1750}$, analogously for the Gifford formula.) Table 1 shows that the estimated coefficients are numerically similar in magnitude. Indeed, the null

hypothesis that intercept plus slope equal zero cannot be rejected for S_LND. The conclusion from the hypothesis test is not sensitive to whether or not a heteroskedasticity adjustment is applied.

TABLE 1. Simple linear regressions of the sinks series on normalized atmospheric concentrations C_t/C_{1750}. Standard errors in parentheses. F is the F-statistic for a test that intercept plus slope equal zero. $C_{1750} = 593.43$GtC. The Durbin-Watson statistic indicates first-order serial correlation when it deviates from the benchmark value of 2.

Series	intercept	C_t/C_{1750}	Durbin-Watson	F
S_LND	-6.72 (1.12)	7.12 (0.88)	2.02	2.37 (p=0.13)
S_OCN	-4.86 (0.24)	5.18 (0.18)	0.66	36.62 (p=1e-7)

2.2. **The ocean sink as a linear function of atmospheric concentrations.** As in the case of the land sink, linearity of the ocean sink in atmospheric concentrations has been employed in the literature, see for example Raupach et al. (2014), Raupach (2013), Gloor et al. (2010), Canadell et al. (2007b), and Bennedsen et al. (2019b). For the CO_2 flux from atmosphere to ocean, Joos et al. (1996) and Joos et al. (2001) specify the relation

$$S\_OCN_t = k_{O,ppm}(pCO2_t^a - pCO2_t^s),$$

where $pCO2_t^a$ and $pCO2_t^s$ are the partial pressures of CO_2 in the atmosphere at sea level and in the ocean surface layer, respectively. The coefficient $k_{O,ppm}$ determines the gas exchange between the atmosphere and the ocean surface layer. The atmospheric partial pressure at sea level is simply a linear transformation of the atmospheric CO_2 concentration. The partial pressure of CO_2 in the surface layer of the ocean is specified as

$$pCO2_t^s = \left(pCO2_0^s + \delta_1\Phi(L)S\_OCN_t + \delta_2[\Phi(L)S\_OCN_t]^2 + \ldots + \delta_5[\Phi(L)S\_OCN_t]^5\right)\exp(\gamma(T_t - T_0)),$$

where $pCO2_0^s$ is initial partial pressure in the surface ocean (equal to pre-industrial atmospheric partial pressure assuming equilibrium of the pre-industrial era ocean surface layer with the atmosphere). Since dissolution of CO_2 in the surface ocean depends on temperature, the coefficients $\delta_i = \delta_i(T_0)$ depend on the initial temperature T_0, and the expression in parentheses is multiplied by an exponential evaluated in the temperature difference $T_t - T_0$ between time t and time 0, multiplied by a coefficient γ. The lag polynomial $\Phi(L)$ is a time-invariant linear filter that describes the dissolution of carbon from the atmosphere in the surface ocean over time:

$$\Phi(L)S\_OCN_t = \phi_1 S\_OCN_{t-1} + \phi_2 S\_OCN_{t-2} + \ldots + \phi_{t-1}S\_OCN_1.$$

We present the model for an annual sampling frequency. It is specified in Joos et al. (1996) and Joos et al. (2001) for higher resolutions as well, with different $\Phi(L)$ filters for subannual responses, but we abstract from this due to the nature of the data. This model for the ocean sink is employed, for example, in the widely used MAGICC model (Meinshausen et al., 2011). It is clearly a highly nonlinear description of the uptake and dissolution process and how it depends on the pressure differential, the temperature, and temporal dynamics. The key features of the specification for the purposes of our statistical analysis are:

a) After suitable unit conversions, the model can be written in the form

$$S\_OCN_t = k_{O,GtC}[C_t - C_{1750} - f(\Phi(L)S\_OCN_t)], \tag{12}$$

that is, with a leading linear term in current atmospheric CO_2 concentrations.

b) When we approximate equation (12) as

$$S\_OCN_t \approx c_O + k_O C_t, \tag{13}$$

lags of S_OCN_t, and hence lags of C_t, enter into the approximation error both linearly and non-linearly up to fifth powers through function f. This introduces serial correlation (linear terms) and memory in higher moments (non-linear terms). We show in Section 4 that on the data sample, this memory can be sufficiently captured with a linear first-order autoregressive filter, such that the residuals appear as Gaussian white noise.

c) The linear specification in concentrations (13) is similar to the linearized specification for the land sink (11). In particular, estimating a linear regression on C_t/C_{1750}, we expect an estimated intercept of the same magnitude as the regression coefficient, with opposite sign. Table 1 shows that the coefficients are numerically similar in magnitude, but the F test rejects the null hypothesis that their sum is zero. The conclusion does not depend on whether or not a heteroskedasticity adjustment is applied, which is the case here. The Durbin-Watson statistic in the table shows evidence for serial correlation.

2.3. **The dynamics of atmospheric concentrations.** We assume sinks linear in concentrations, $S\_LND_t = c_L + k_L C_t +$ stationary error, and $S\_OCN_t = c_O + k_O C_t +$ stationary error. We employ a random walk with drift for emissions $E_t = d + E_{t-1} + X_t^E = E_0 + dt + x_t$, where E_0 are initial emissions, d is the drift term, $x_t = \sum_{\tau=1}^{t} X_\tau^E$, and X_t^E is a stationary process. Using the GCB

equation

$$G\_ATM_t = C_t - C_{t-1} = E_t - S\_LND_t - S\_OCN_t$$
$$= c + dt + x_t - k_L C_t - k_O C_t + \text{stationary error},$$

where $c = E_0 - c_L - c_O$, and denoting $q = 1/(1 + k_L + k_O)$, we arrive at a first-order difference equation

$$(1 - qL)\, C_t = qc + qdt + qx_t + q\varepsilon_t,$$

where ε_t collects the stationary and serially correlated error terms from the two sinks. The deterministic trend and the I(1) process x_t dominate. The difference equation for C_t admits the following solution, where C_0 are initial concentrations,

$$(14)\quad C_t = q^t \left[C_0 - \frac{cq}{1-q} + \frac{dq^2}{(1-q)^2}\right] + \left[\frac{cq}{1-q} - \frac{dq^2}{(1-q)^2}\right] + \frac{dq}{1-q}t + \sum_{j=0}^{t-1} q^{j+1} x_{t-j} + \sum_{j=0}^{t-1} q^{j+1}\varepsilon_{t-j}$$
$$= o(1) + O(1) + O(t) + I(1) + I(0) = O(t) + I(1).$$

Note that the variance of an $I(1)$ process is $O(t)$. The estimated coefficients k_L and k_O are small, from Table 1 we obtain $\hat{k}_L = 7.12/C_{1750} \approx 0.012$ and $\hat{k}_O \approx 5.18/C_{1750} = 0.009$. Thus, the first-order polynomial

$$(15)\qquad 1 - qz \approx 1 - \frac{1}{1 + 0.012 + 0.009} z$$

has its root $z_0 = 1 + k_L + k_O$ close to, but above, unity. The lag polynomial $1 - qL$ is a stationary filter, but close to simple first differences $1 - L$. Therefore, the term $\sum q^{j+1} x_{t-j}$ is an I(1) process, but close to I(2). Note that the coefficient of the linear trend is $dq/(1-q) = d/(k_L + k_O)$. This means that the drift dt from emissions is divided by approximately 0.02. This is the reason why the trend appears dominant in the graph of C_t (see panel [A] of Figure 3 for a plot of the atmospheric concentration time series), even though it is linear. If the lag polynomial was equal to first differences,

$$(1 - L)C'_t = c + dt + x_t + \varepsilon_t,$$

the solution would be

$$C'_t = C_0 + \left[c + \frac{d}{2}\right] t + dt^2 + \sum_{\tau=1}^{t} x_\tau + \sum_{\tau=1}^{t} \varepsilon_\tau,$$

that is a sum of a linear trend, a quadratic trend, an I(2) process, and an I(1) process. In other words, the orders of integration of the terms of the solution (14) would increase by one.

Note that as C_t increases, and thus the distance $C_t - C_{1750}$ from pre-industrial concentrations, the increasingly negative second-order terms in the sink processes move the root of the lag polynomial (15) closer to unity (compare Equations (11) and (13)). As terrestrial plants saturate and the ocean acidifies, they take up less CO_2 than at concentrations close to pre-industrial levels (Raupach et al., 2014; Bennedsen et al., 2019b). Therefore, the higher the level of atmospheric concentrations, the closer they get to explosive I(2) dynamics.

3. Monte Carlo simulation study

To validate the maximum likelihood estimation procedure for the unknown parameters in our model, we conduct a Monte Carlo simulation study. We consider the basic model as set out in equations (1) to (8) in the previous section. Given this model specification and with a choice of parameter values, which we regard as the "true" model and the "true" parameter values, we simulate data for the four time series variables $\{C_t, S\_LND_t, S\_OCN_t, E_t\}$ from the basic model. We notice that before we can simulate the four "observed" time series, we first need to simulate paths for the "unobserved" variables C_t^* and E_t^* from their dynamic model equations in (1) and (2), respectively. Our Monte Carlo study consists of M simulations. For each simulation, we estimate the parameters in the basic model using the set of the four simulated time series (see Section 5 for further details on the estimation procedure). In this way, we obtain a collection of M estimates for the parameter vector. We assess the accuracy of the estimates by comparing the estimated parameter values with their corresponding "true" values. In Table 2 we report the sample bias and sample standard error for all parameters. Our Monte Carlo study is based on $M = 1,000$ simulations and on three different time series lengths $T = 30$, $T = 60$ and $T = 120$. In our empirical study, we have $T = 61$.

First, we study the performance of the maximum likelihood estimation method for correct model specification: we consider the same model for simulation as for estimation. The "true" parameter values in this Monte Carlo study are chosen so that they resemble the estimated parameter values in the empirical study of Section 5. Given that our model is linear and Gaussian, and given that the exact maximum likelihood estimation method has generally good finite sample properties, we expect a good performance overall. However, for the small sample cases (e.g. $T = 30$ and $T = 60$), it is possible that the variances for latent components such as $X_{i,t}$, for $i = 1, \ldots, 4$, can be subject to the "pile-up" problem which refers to the case that a high number of estimated variances are equal to the boundary value of zero, see the discussions in Shephard (1993) and Stock and Watson (1996). In particular, the "true" variance for $X_{4,t}$ equals the relatively small value of 0.001 and a

TABLE 2. **Results of Monte Carlo Study.** Sample mean bias and sample standard errors (in parentheses) are reported for the parameter estimates based on $M = 1,000$ simulations from the basic model specification (1) to (8) where $X_{2,t}$, $X_{4,t}$ and X_t^E are Gaussian noise variables with zero mean and variance $\sigma^2_{\eta 2}$, $\sigma^2_{\eta 4}$ and $\sigma^2_{\eta 5}$, respectively, where $X_{1,t}$ and $X_{3,t}$ are autoregressive processes of order 1 with zero mean, variance $\sigma^2_{\eta 1}$ and $\sigma^2_{\eta 3}$, and autoregressive coefficient ϕ_1 and ϕ_3, respectively, and where r_{12} and r_{13} are the correlations for the pairs $(X_{1,t}, X_{2,t})$ and $(X_{1,t}, X_{3,t})$, respectively. The parameter values are selected to be close to those obtained in the empirical study of Section 5. Results are presented for different model specifications which are described in the main text.

true par.		correct spec.	set $\beta_{1,2}=0$	set $\phi_{1,3}=0$	set $r_{12,13}=0$	br G_{ATM}	br Var(η_t)
				$T = 30$			
c_1	-7.22	.2672 (6.17)	3.664 (136.)	.0356 (.965)	-.1726 (8.54)	0.137 (6.92)	.2612 (6.34)
c_2	-4.93	.1153 (4.10)	2.887 (107.)	-.0192 (.440)	-.1531 (6.32)	0.062 (5.23)	.1339 (5.47)
β_1	7.0	-.0268 (1.00)	−−	-.0374 (.938)	-.0134 (.990)	-0.026 (.921)	-.0557 (1.43)
β_2	5.5	.0208 (.456)	−−	.0187 (.423)	.0236 (.449)	0.019 (.415)	.0327 (.671)
d	0.14	-.0009 (.033)	-.0009 (.033)	-.0009 (.033)	-.0009 (.033)	0.061 (.033)	-.0010 (.033)
ϕ_1	0.8	-.0813 (.167)	-.0261 (.150)	−−	-.0795 (.159)	-0.081 (.168)	-.0979 (.177)
ϕ_3	0.7	-.1124 (.184)	-.0445 (.168)	−−	-.1120 (.178)	-0.112 (.184)	-.1331 (.196)
$\sigma^2_{\eta 1}$	0.90	-.0469 (.265)	-.0035 (.271)	-.0208 (.249)	-.0641 (.229)	-0.052 (.266)	.7847 (.487)
$\sigma^2_{\eta 2}$	0.70	-.0158 (.189)	-.0062 (.187)	-.0159 (.188)	-.0238 (.165)	-0.016 (.188)	.6964 (.404)
$\sigma^2_{\eta 3}$	0.01	-.0007 (.003)	-.0003 (.003)	-.0003 (.003)	-.0008 (.003)	-0.001 (.003)	.0093 (.006)
$\sigma^2_{\eta 4}$	0.001	.0018 (.004)	.0019 (.004)	.0018 (.004)	.0018 (.004)	0.006 (.011)	.0027 (.005)
$\sigma^2_{\eta 5}$	0.03	-.0042 (.011)	-.0042 (.011)	-.0042 (.011)	-.0042 (.011)	0.096 (.030)	-.0036 (.012)
r_{12}	-0.65	.0061 (.117)	.0043 (.113)	.0140 (.114)	−−	0.004 (.118)	.0163 (.127)
r_{13}	-0.15	-.0096 (.162)	-.0086 (.157)	-.0007 (.150)	−−	-0.011 (.162)	-.0085 (.185)
				$T = 60$			
c_1	-7.22	-.0285 (.446)	-0099 (.216)	-.0181 (.335)	-.0043 (.758)	-.0214 (.377)	-.0338 (.836)
c_2	-4.93	.0005 (.317)	-0093 (.167)	.0029 (.208)	.0179 (.549)	.0041 (.256)	.0035 (.620)
β_1	7.0	.0194 (.287)	−−	.0157 (.271)	.0186 (.286)	.0179 (.264)	.0300 (.410)
β_2	5.5	-.0048 (.178)	−−	-.0028 (.167)	-.0047 (.177)	-.0041 (.164)	-.0065 (.258)
d	0.14	-.0006 (.023)	-0006 (.023)	-.0006 (.023)	-.0006 (.023)	.0293 (.023)	-.0006 (.023)
ϕ_1	0.8	-.0359 (.091)	-0111 (.085)	−−	-.0356 (.089)	-.0361 (.091)	-.0424 (.099)
ϕ_3	0.7	-.0474 (.112)	-0171 (.103)	−−	-.0457 (.111)	-.0474 (.112)	-.0594 (.121)
$\sigma^2_{\eta 1}$	0.90	-.0276 (.168)	-0104 (.168)	-.0152 (.167)	-.0387 (.145)	-.0296 (.168)	.9031 (.385)
$\sigma^2_{\eta 2}$	0.70	-.0076 (.127)	-0049 (.127)	-.0069 (.128)	-.0150 (.114)	-.0076 (.127)	.7351 (.295)
$\sigma^2_{\eta 3}$	0.01	-.0003 (.002)	-0002 (.002)	-.0002 (.002)	-.0003 (.002)	-.0003 (.002)	.0100 (.004)
$\sigma^2_{\eta 4}$	0.001	.0014 (.003)	0015 (.003)	.0014 (.003)	.0014 (.003)	.0028 (.005)	.0024 (.004)
$\sigma^2_{\eta 5}$	0.03	-.0027 (.008)	-0028 (.008)	-.0028 (.008)	-.0027 (.008)	.0464 (.015)	-.0023 (.009)
r_{12}	-0.65	-.0026 (.077)	-0010 (.077)	.0017 (.077)	−−	-.0034 (.077)	.0010 (.088)
r_{13}	-0.15	-.0008 (.103)	-0004 (.102)	.0008 (.099)	−−	-.0012 (.103)	-.0032 (.113)
				$T = 120$			
c_1	-7.22	-.0034 (.123)	-.0008 (.030)	-.0045 (.126)	-.0036 (.129)	-.0029 (.123)	-.0013 (.166)
c_2	-4.93	-.0013 (.107)	.0005 (.027)	-.0003 (.102)	-.0018 (.106)	-.0014 (.102)	-.0049 (.135)
β_1	7.0	.0017 (.079)	−−	.0021 (.076)	.0017 (.078)	.0014 (.074)	-.0001 (.113)
β_2	5.5	.0009 (.066)	−−	.0005 (.062)	.0013 (.065)	.0010 (.062)	.0036 (.094)
d	0.14	-.0003 (.016)	-.0003 (.016)	-.0003 (.016)	-.0003 (.016)	.0144 (.016)	-.0003 (.016)
ϕ_1	0.8	-.0173 (.050)	-.0075 (.048)	−−	-.0169 (.049)	-.0176 (.050)	-.0208 (.056)
ϕ_3	0.7	-.0242 (.073)	-.0113 (.068)	−−	-.0244 (.072)	-.0243 (.073)	-.0309 (.082)
$\sigma^2_{\eta 1}$	0.90	-.0090 (.116)	-.0024 (.117)	-.0033 (.115)	-.0131 (.103)	-.0097 (.116)	.9506 (.270)
$\sigma^2_{\eta 2}$	0.70	.0041 (.089)	.0052 (.088)	.0042 (.088)	.0011 (.080)	.0041 (.089)	.7603 (.209)
$\sigma^2_{\eta 3}$	0.01	-.0002 (.001)	-.0001 (.001)	-.0001 (.001)	-.0002 (.001)	-.0002 (.001)	.0104 (.003)
$\sigma^2_{\eta 4}$	0.001	.0007 (.002)	.0008 (.002)	.0007 (.002)	.0007 (.002)	.0012 (.003)	.0016 (.002)
$\sigma^2_{\eta 5}$	0.03	-.0015 (.006)	-.0016 (.006)	-.0015 (.006)	-.0015 (.006)	.0227 (.008)	-.0011 (.006)
r_{12}	-0.65	-.0008 (.054)	.0002 (.054)	.0008 (.053)	−−	-.0010 (.053)	-.0008 (.059)
r_{13}	-0.15	.0005 (.069)	.0004 (.069)	.0013 (.069)	−−	.0004 (.069)	-.0013 (.080)

strong pile-up is evidenced by the relatively large bias and large standard error, especially for the two smaller sample sizes. All other model parameters show good performances with small biases and small standard errors. The bias is measured as the sample mean over the set of M parameter estimates, minus the "true" parameter value. The standard error is obtained from the corresponding sample variance.

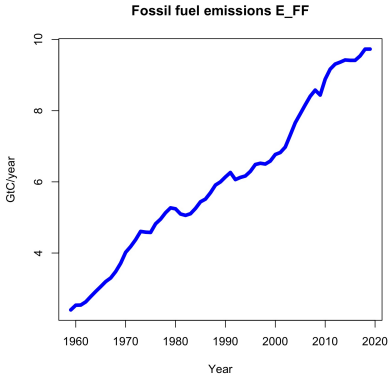
Next, we extend the Monte Carlo study and assess the estimation accuracy to cases where the "true" model for the data generation process (DGP) and the model for estimation are different in specific ways. We start by considering cases where we set particular values for the β, ϕ and r_{ij} parameters in the DGP. However, during the subsequent estimation process, we restrict a particular set of parameters to zero, as if they are not present in the model. The remaining parameters are estimated as usual. We expect that the estimation of the other parameters will be affected in this erroneous setting: the estimates of the remaining parameters are subject to the incomplete model specification. The results in Table 2 confirm that model misspecification affects the estimation accuracy, especially when sample sizes are smaller. However, the model is sufficiently flexible to provide an overall good fit at the cost of increased inaccuracies in the estimation of related parameters.

Finally, we also present results for cases where the DGP model is distorted. We consider the case of a structural break in the growth term in C_t^*, located in the middle of the time series and specied as a one-off change (dummy) in G_ATM_t of 10 units. In another case we distort the DGP model in the middle of the time series with a change in the variance of $X_{i,t}$ for $i = 1, \ldots, 4$; in the second half of the time series, all variances are multiplied by $\sqrt{10}$. For these additional two misspecification cases in the DGP, the parameters are estimated using the original model specification (ignoring the breaks). These estimation results are also presented in Table 2. The reported biases for these cases show that the break in variance leads to more severe distortions in estimation accuracy for all measurement variances and to some extent also for the β coefficients. The break in the growth term has most impact on the estimation accuracy for the growth term coefficient d itself which is to be expected. Other parameter estimates are not heavily affected by the severe breaks in the DGP. This last point provides some evidence of our flexible and robust modeling framework.

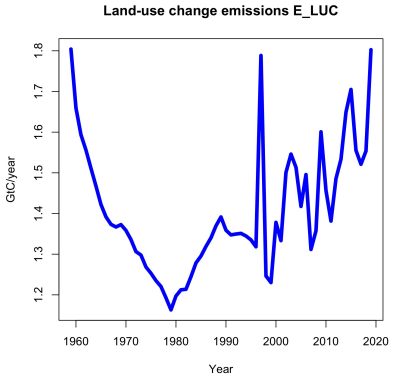
4. The data set and its time series properties

Figure 1 displays the time series data set related to the GCB that we employ in our study, both in levels and in first differences. All series are annual, observed from 1959 to 2019, and are taken from the global file of Friedlingstein et al. (2020), available at `https://www.icos-cp.eu/`

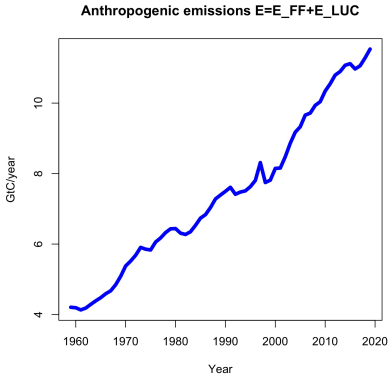
FIGURE 1. GCB annual time series, from 1959 to 2019. E_FF: fossil fuel emissions, E_LUC: land-use change emissions, E: anthropogenic emissions (E=E_FF+E_LUC), G_ATM: growth of atmospheric concentrations, S_LND: land sink, S_OCN: ocean sink.



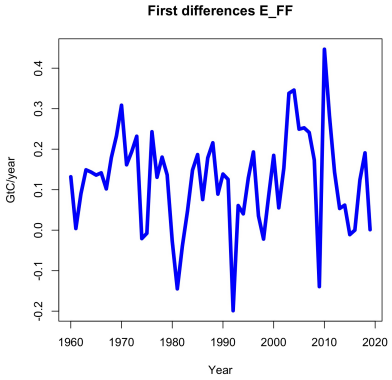
(A) E_FF



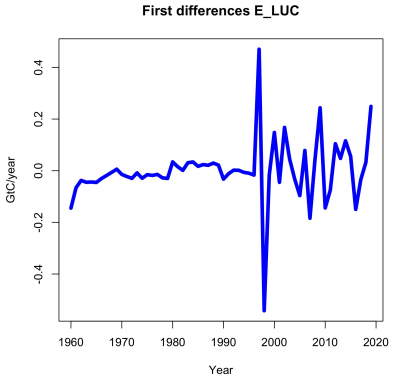
(B) E_LUC



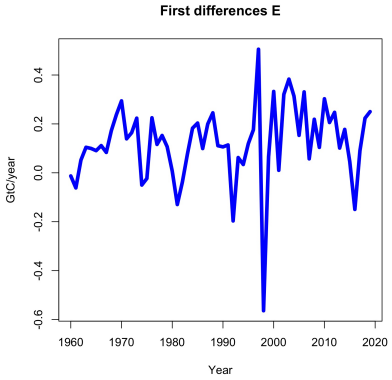
(C) E



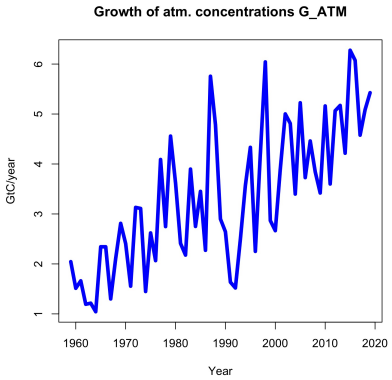
(D) First differences E_FF



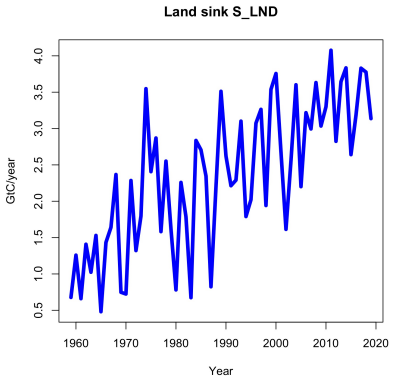
(E) First differences E_LUC



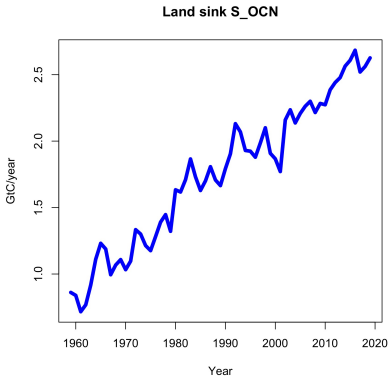
(F) First differences E



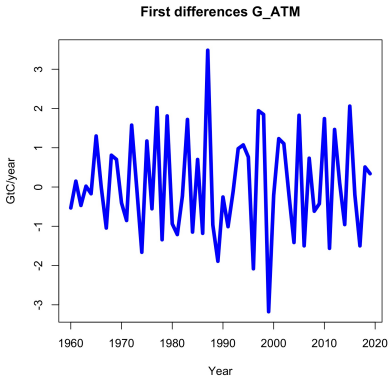
(G) G_ATM



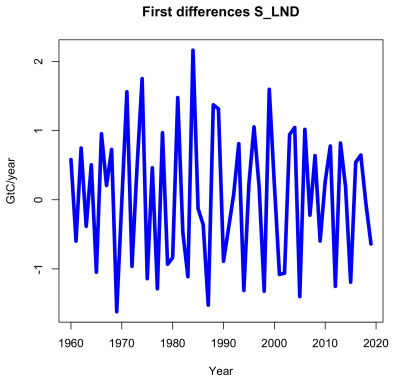
(H) S_LND



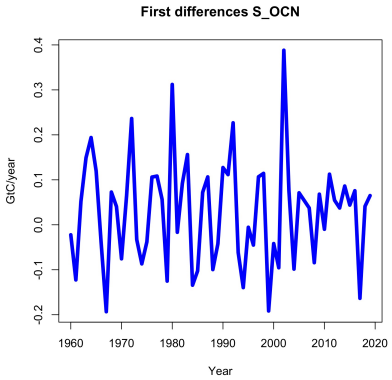
(I) S_OCN



(J) First differences G_ATM



(K) First differences S_LND



(L) First differences S_OCN

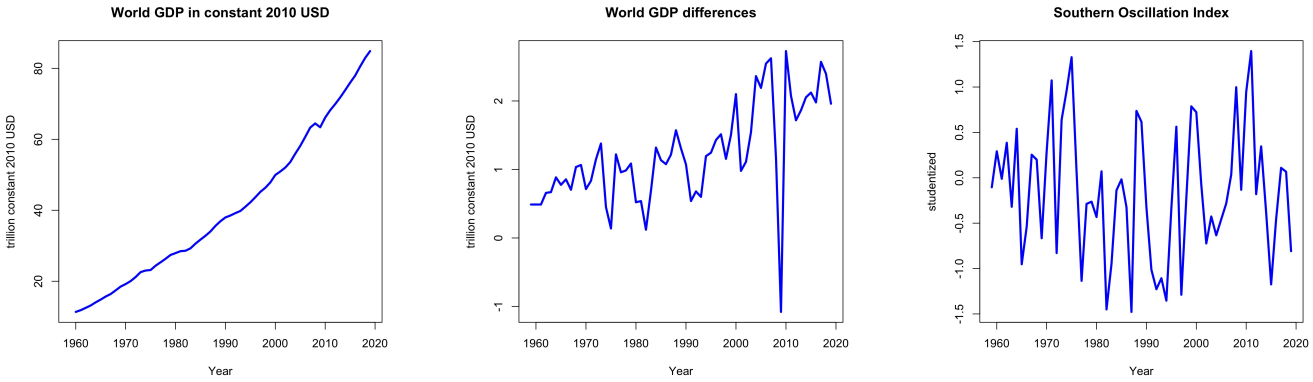
science-and-impact/global-carbon-budget/2020. All series are measured in gigatons of carbon (GtC) per year. Panel [A] in Figure 1 presents fossil fuel emissions E_FF, which are calculated including cement carbonation from the global file. Panel [B] shows land-use change emissions E_LUC, which are the average of three series prepared by Houghton and Nassikas (2017), Hansis et al. (2015), and Gasser et al. (2020). Panel [C] shows the sum of E_FF and E_LUC, labeled anthropogenic emissions E. Panels [D] through [F] show the first differences of the series above. Panel [G] shows growth in atmospheric concentrations; these are the first differences of the so-called Keeling curve of atmospheric concentrations (Dlugokencky and Tans, 2020). Panel [H] shows the land sink, which is the mean of the output of 17 different models (Haverd et al., 2018; Melton et al., 2020; Yuan et al., 2014; Lawrence et al., 2019; Tian et al., 2015; Meiyappan et al., 2015; Delire et al., 2020; Mauritsen et al., 2019; Sellar et al., 2019; Smith et al., 2014; Poulter et al., 2011; Lienert and Joos, 2018; Zaehle and Friend, 2010; Vuichard et al., 2019; Walker et al., 2017; Kato et al., 2013; Yue and Unger, 2015). Panel [I] shows the ocean sink, which is the mean of the output of 9 different models (Buitenhuis et al., 2013; Schwinger et al., 2016; Paulsen et al., 2017; Berthet et al., 2019; Law et al., 2017; Hauck et al., 2020; Liao et al., 2020; Doney et al., 2009; Aumont et al., 2015) and evaluated using Landschützer et al. (2016); Rödenbeck et al. (2014); Denvil-Sommer et al. (2019); Gregor et al. (2019); Watson et al. (2020). Panels [J] through [L] show the first differences of the series above.

Figure 2 displays the time series that are used to extend the model in Section 5. Panel [A] shows World GDP in constant 2010 USD obtained from World Bank (2021b). The year-to-year differences of World GDP in Panel [B] are employed as drift in the random walk model for emissions. The Southern Oscillation Index (Panel [C]) is used as an explanatory variable for the sinks processes and obtained from Climatic Research Unit (2021); Ropelewski and Jones (1987). The Southern Oscillation Index is the studentized measure of differences in atmospheric pressure at sea level between Tahiti and Darwin, Australia. Positive (negative) values correspond to La Niña (El Niño) phases.

The most conspicuous feature of the time series variables E_FF, E, G_ATM, S_LND, and S_OCN in Figure 1 is the upward trend, whereas the variable E_LUC is not trending. We conduct augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) tests for unit roots to determine the order of integration of the series and whether stationarity can be achieved by either de-trending or by differencing. We keep the notation of the original sources for each test.

The test results are presented in Table 3: all tests do not reject the unit root hypothesis for the E_FF time series, including the ADF test that allows for a drift term. In case of the variable E_LUC,

FIGURE 2. Annual time series, from 1959 to 2019, of World Gross Domestic Product (GDP, constant 2010 USD), in levels and in first differences, and the Southern Oscillation Index (SOI)



(A) World GDP (B) World GDP differences (C) South. Oscillation Idx

the tests are not unanimous: the ADF and PP tests point towards stationarity while the KPSS test rejects the stationarity hypothesis. Anthropogenic emissions E inherit the agreement of the tests on E_FF and clearly exhibit a unit root, with drift term according to ADF. All tests agree that G_ATM, S_LND, and S_OCN are trend-stationary and that SOI as well as the budget imbalance BIM=E-G_ATM-S_LND-S_OCN are stationary. All tests agree that atmospheric concentrations C contain a unit root, and G_ATM, which are first differences of C, are unanimously identified as trend-stationary. One might be tempted to conclude from these results that C has a dominant quadratic trend, but this would ignore the important system aspect of the series that we discussed in Section 2.3. First differences of World GDP are found to be trend-stationary by all tests.

It may be counter-intuitive that G_ATM, S_LND, and S_OCN are trend-stationary with upward trends since, according to the global carbon budget, the main driver of the system are emissions E, which follow a random walk with drift according to the tests. The drift term in E is then a linear trend, but the deviations from this linear trend follow a random walk, not a stationary process. The key to understanding this finding is that even though emissions E are cumulated in concentrations, the sinks are subtracted, which are linear in concentrations. This leads to the rich dynamics in equation (14) from which it follows that C_t is $O(t)+I(1)$, and thus $\Delta C_t = G\_ATM_t$ is indeed stationary. The deterministic trend $dt/(k_L+k_O)$ is, however, so dominant in C_t due to the near second unit root that on a relatively small sample of 61 observations, it looks nearly indistinguishable from a quadratic trend, and there still seems a linear trend in first differences.

TABLE 3. **Unit root tests.** ***, **, * denote significance at 0.01, 0.05, 0.10 level according to tabulated unit root distributions, (***) 0.01 significance according to standard normal.

Augmented Dickey-Fuller test: The ADF test regression is given by $\Delta y_t = \alpha + \beta t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_t$, where we select k by Bayes Information Criterion (BIC). We test the hypothesis $H_0 : \pi = 0$ using the t-statistic $\hat{\tau}_\tau$ that follows a Dickey-Fuller distribution (Fuller, 1996, p. 562, 642). Φ_3 is the statistic in Table VI in Dickey and Fuller (1981). $\hat{\tau}_\beta$ is the standard t-statistic on the trend coefficient. $\hat{\tau}_\mu$ is a statistic from Table 10.A.2 in (Fuller, 1996, p. 642). Φ_1 is the statistic in Table IV in Dickey and Fuller (1981).

Phillips-Perron test: The PP test regression is given by $y_t = \mu + \beta(t - \frac{1}{2}T) + \alpha y_{t-1} + u_t$, where T is the sample size. Phillips and Perron (1988) define the statistic $Z(t_{\hat{\alpha}})$ that follows a Dickey-Fuller distribution and is robust to serial correlation in the error term u_t. This obviates the selection of the tuning parameter k in the ADF tests. We test the hypothesis $H_0 : \alpha = 1$ using $Z(t_{\hat{\alpha}})$. If the test rejects, the series is either stationary or trend-stationary, depending on the test of $H_0 : \beta = 0$ using the $Z(t_{\hat{\beta}})$ statistic proposed in Phillips and Perron (1988). The $Z(t_{\hat{\beta}})$ statistic follows the distribution for $\hat{\tau}_{\beta\tau}$ in Table III in Dickey and Fuller (1981).

KPSS test: The KPSS test considers the data-generating model $y_t = \xi t + r_t + \varepsilon_t$, $r_t = r_{t-1} + u_t$, $u_t \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$, and ε_t stationary. The test is for $H_0 : \sigma_u^2 = 0$, that is, unlike the ADF and Phillips-Perron tests, here the null hypothesis is trend-stationarity. The Lagrange multiplier test statistic is proposed in Kwiatkowski et al. (1992). The asymptotic distribution for $\xi = 0$ is a first-level Cramér-von Mises distribution, the asymptotic distribution for $\xi \neq 0$ is a second-level Cramér-von Mises distribution. For LM(trend), the residuals e_j are from a regression on a constant and a trend, for LM(constant), on a constant only.

	Augmented Dickey-Fuller test						**Phillips-Perron test**			**KPSS test**		
Series	$\hat{\tau}_\tau$	Φ_3	$\hat{\tau}_\beta$	$\hat{\tau}_\mu$	Φ_1	Decision	$Z(t_{\hat{\alpha}})$	$Z(t_{\hat{\beta}})$	Decision	LM(trend)	LM(constant)	Decision
E_FF	$-1.927_{K=1}$	1.861		$-0.402_{K=1}$	6.802***	I(1) w/ drift	-1.784		I(1)	0.190**		I(1)
E_LUC	$-2.804^{(***)}_{K=1}$	5.541*		$-1.102_{K=1}$	0.694	stationary	-3.74**	2.273	stationary	0.310***		I(1)
E	$-1.536_{K=1}$	1.424		$0.418_{K=1}$	10.267***	I(1) w/ drift	-1.892		I(1)	0.258***		I(1)
G_ATM	$-5.678^{***}_{K=1}$		4.357***			trend-stationary	-7.024***	5.245***	trend-stationary	0.061	1.259***	trend-stationary
C	$0.266_{K=1}$	10.324***				I(1)	-0.355		I(1)	0.379***		I(1)
S_LND	$-6.956^{***}_{K=1}$		5.527***			trend-stationary	-8.055***	5.667***	trend-stationary	0.046	1.397***	trend-stationary
S_OCN	$-4.580^{***}_{K=1}$		4.507***			trend-stationary	-4.223***	4.094***	trend-stationary	0.102	1.576***	trend-stationary
BIM	$-4.458^{***}_{K=1}$		0.078			stationary	-5.360***	-0.808	stationary	0.125*	0.221	stationary
SOI	$-5.024^{***}_{K=1}$		0.064			stationary	-5.787***	-0.381	stationary	0.101	0.112	stationary
DGDP	$-5.775^{***}_{K=1}$		4.075**			trend-stationary	-6.213***	3.766***	trend-stationary	0.113	1.147***	trend-stationary

5. Parameter estimation by the maximum likelihood method

The analyses of the dependence of the sink processes on atmospheric concentrations in Sections 2.1 and 2.2, together with the unit root analyses of the constituents of the global carbon budget in Section 4, imply that a statistical model of the GCB should have the following properties:

a) Emissions E_t follow a random walk process with a drift.

b) Land and ocean sinks depend linearly on atmospheric concentrations C_t. Since concentrations are monotonically increasing in time, are smooth, and are only slightly convex on the sample (see Figure 3 [A]), they constitute the "trend" as detected in the unit root tests of Section 4.

c) The stationary residual processes X_2, X_3 of the linear relations of the sinks with concentrations should allow for serial correlation, which is in particular expected to be present in the ocean sink, see Table 1.

d) Atmospheric concentrations C_t follow a deterministic trend plus I(1) dynamics through the global carbon budget equation.

The model specified in equations (1) to (8), and summarized in the appendix in equation (25), satisfies these properties. In this section, we estimate the model on the Global Carbon Project data set. We discuss residual diagnostics and parameter estimates. We then extend the model to include El-Niño/Southern Oscillation dynamics in the sinks, World GDP changes in emissions, and a number of dummy variables for outliers in the equation for emissions. We show that extending the model along these three directions improves the goodness-of-fit while the residual diagnostics demonstrate that the model describes the data well.

5.1. **Details of parameter estimation.** Model (1) to (8) is presented in a linear state space form in equation (25) in the appendix. The parameters in the model are estimated by maximum likelihood using state space methods (Durbin and Koopman, 2012). The state vector contains the dynamic (stationary and non-stationary) features of the model and the linear regression effects. The stationary elements of the state vector ($X_{i,t}$ for $i = 1, \ldots, 4$) are initialized based on corresponding unconditional moments while the non-stationary elements and fixed coefficients are subject to diffuse initial conditions. The state equation disturbance vector is given by $\eta_t = (\eta_{1,t}, \ldots, \eta_{5,t})'$ where $\eta_{i,t}$ corresponds to the disturbance in the autoregressive processes $X_{i,t}$, for $i = 1, \ldots, 4$ and where $\eta_{5,t}$ is the disturbance in the autoregressive process X_t^E. We assume that the disturbances are mutually independent with the exception of having a non-zero covariance for $\eta_{1,t}$ (C) on the one hand and for $\eta_{2,t}$ (S_LND) and $\eta_{3,t}$ (S_OCN) on the other hand. Hence, the corresponding correlations, as indicated

by r_{12} and r_{13}, are non-zero. We also have considered other correlations between disturbances being treated as non-zero. However, such more elaborate versions of the model have resulted in estimated correlations that are close to zero and insignificant.

The 13×1 parameter vector is given by

$$\psi = (\beta_1, \beta_2, \phi_1, \phi_3, \phi_4, \phi_E, \sigma^2_{\eta_1}, \sigma^2_{\eta_2}, \sigma^2_{\eta_3}, \sigma^2_{\eta_4}, \sigma^2_{\eta_5}, r_{12}, r_{13})'.$$

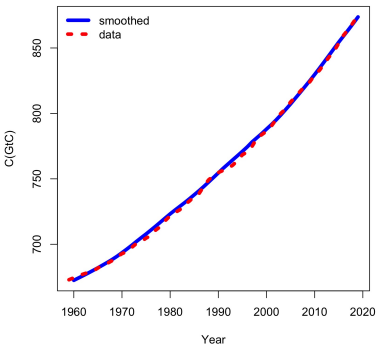
while the fixed coefficients c_1, c_2 and d are placed in the state vector, see Appendix A. The estimation of ψ is based on maximizing the log-likelihood function that is evaluated by an augmented Kalman filter in order to account for the diffuse conditions in the state vector (Durbin and Koopman, 2012, p. 173).

Alternatively, we can consider the standard Kalman filter that is initialized as indicated but the diffuse conditions are replaced by a mean zero and variance K, with K a large value, say $K = 10^7$. This so-called *big-K* method yields very similar results. The coefficients c_1, c_2, and d are added to the state vector, with transition equations that enforce constancy (a random walk process without innovations). In this way, we concentrate out these unknown fixed coefficients from the log-likelihood function. Their estimates and standard errors are returned by the Kalman filter as filtered state values and the square root of the filtered state variances, respectively.

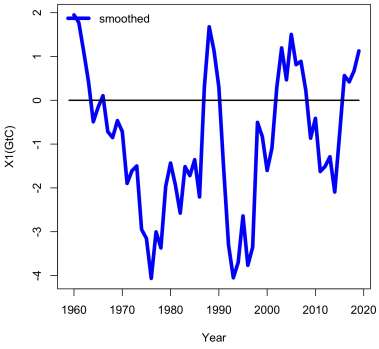
5.2. **Model I: Results and residual diagnostics.** Table 4 displays the estimated parameter values and their standard errors. Figure 3 presents the smoothed states, together with the time series data of C, G_ATM, E, S_LND and S_OCN. The intercepts c_1, c_2 and slopes β_1, β_2 coefficients for the sinks equations are estimated with the same order of magnitude (with opposite signs). These results can be compared with those in Table 1 obtained from simple regressions. Figure 3 shows in panels [D] and [F] that the sinks processes are upward trending with slight positive curvature since they are scaled concentrations (panel [A]). The sinks data exhibit stationary variations around these trends. This is the reason why the unit roots tests in Section 4 identified these time series as trend-stationary.

The serial correlation in the error process $X_{1,t}$ (in C) and $X_{3,t}$ (in S_OCN) is reflected in the estimated autoregressive parameters $\hat{\phi}_1 = 0.86(0.08)$ and $\hat{\phi}_3 = 0.67(0.10)$, respectively. These estimates imply substantial serial correlation, but they are far from unity, and thus these are clearly stationary processes. The smoothed values of the processes are shown in panels [B] and [G] of Figure 3, where the memory in the series can be seen. The process $X_{2,t}$, on the other hand, is indistinguishable from white noise, and the diagnostics on S_LND show that this is a sufficient description of the data (Table 5). A version of the model with an additional parameter ϕ_2 shows a small and insignificant

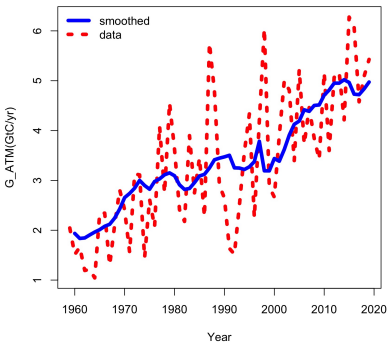
FIGURE 3. Smoothed values for state vector elements of model (1) to (8)



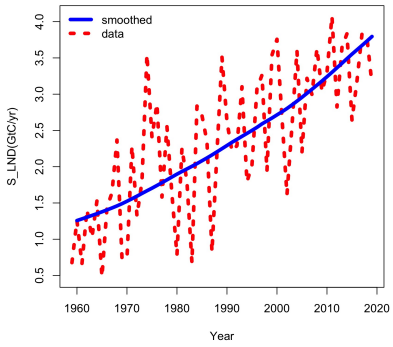
(A) C_t



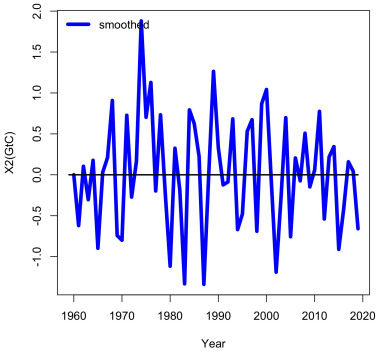
(B) $X_{1,t}$



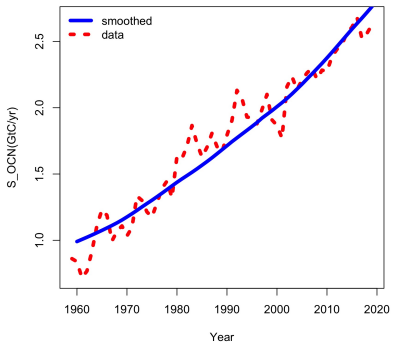
(C) G_ATM_t



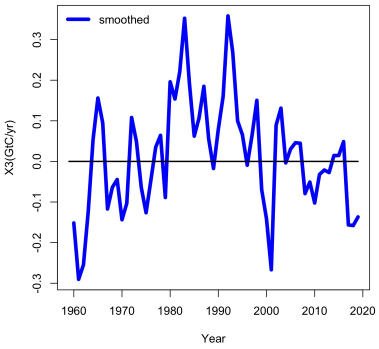
(D) S_LND_t



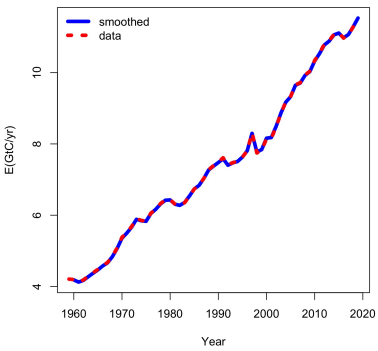
(E) $X_{2,t}$



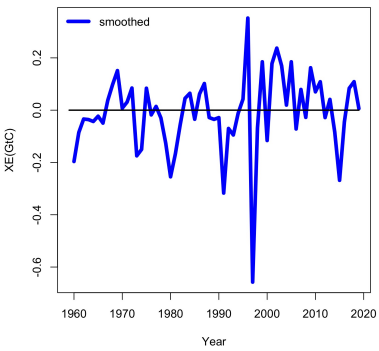
(F) S_OCN_t



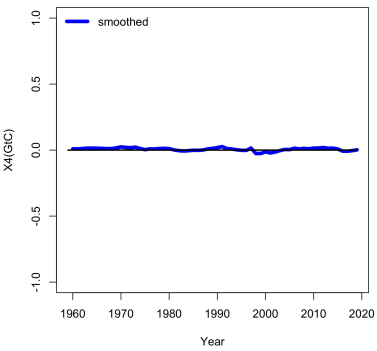
(G) $X_{3,t}$



(H) E_t



(I) X_t^E



(J) $X_{4,t}$

TABLE 4. **Parameter estimates** for model (1) to (8). We indicate standard error by "std err" and filtered estimate for an element of the state vector by "filt.".

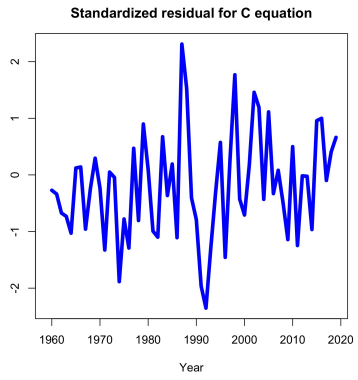
Coefficients			Variances		
	estimate	std err		estimate	std err
c_1 (filt.)	-7.22	0.05	$\sigma^2_{\eta_1}$	0.93	0.26
c_2 (filt.)	-4.93	0.04	$\sigma^2_{\eta_2}$	0.46	0.08
β_1	7.48	0.97	$\sigma^2_{\eta_3}$	0.012	0.002
β_2	5.22	0.41	$\sigma^2_{\eta_4}$	0.001	0.027
d (filt.)	0.14	0.02	$\sigma^2_{\eta_5}$	0.024	0.032
ϕ_1	0.86	0.08	r_{12}	-0.68	0.07
ϕ_3	0.67	0.10	r_{13}	-0.20	0.11
ϕ_4	0.76	0.20			
ϕ_E	0.06	0.26			

estimate for this coefficient. The process $X_{4,t}$ in the measurement equation of E is estimated to be essentially zero, as shown in Figure 3 (panel [J]). It can be concluded that the random walk with drift and stationary error X^E for E* in equation (2) is a sufficient description of the data, and X_4 can be omitted. The average growth in emissions is $\hat{d} = 0.14$ GtC/year. There is a strong correlation between the innovations in the X_1 residual process in C and in the X_2 residual process in S_LND. This may reflect that within the climate models the generation of S_LND is calibrated to agree with the concentration measurements (Friedlingstein et al., 2020).

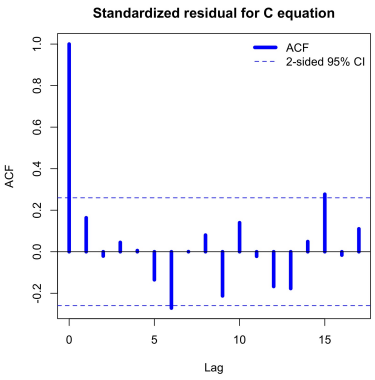
Table 5 presents the diagnostic statistics for the standardized prediction residuals for the model where the parameters are replaced by their corresponding maximum likelihood estimates. The residuals for the four dependent variables are presented in Figure 4, together with their correlograms. The Ljung-Box statistics indicate that there is no first-order serial correlation in the residuals of any of the four series. However, the Durbin-Watson test statistic for the residuals of the C equation shows some distance from the benchmark value of two, indicating some serial correlation. On the other hand, the correlograms in Figure 4 for both C and E do not show any strong evidence of residual serial correlation. The Jarque-Bera test for the null of normally distributed residuals is strongly rejected for the E equation, caused by both strong skewness and high leptokurtosis. There is no serial correlation left in the residual from E, on the other hand, as the diagnostic statistics show. The diagnostic statistics for the residuals of the S_LND and S_OCN equations show a good fit in Table 5, but the correlogram in Figure 4 shows some marginally significant spikes at higher lags for the S_LND equation.

5.3. **Model II: Extension with World GDP and SOI variables.** Although the residual diagnostics for the sink equations in the Model I do not provide much evidence of model misspecifications

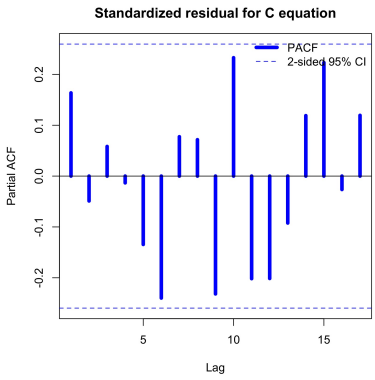
FIGURE 4. Standardized prediction residuals for model (1) to (8), ACF autocorrelation function, PACF partial autocorrelation function. Note that the ACF shows the value for lag zero (equal to one by definition), while the PACF does not.



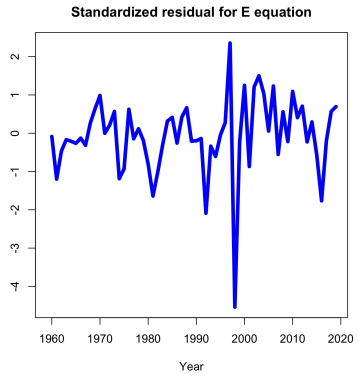
(A) Std. residuals C



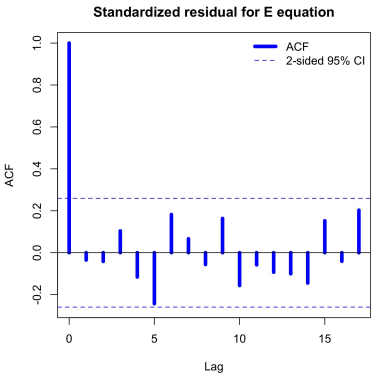
(B) ACF



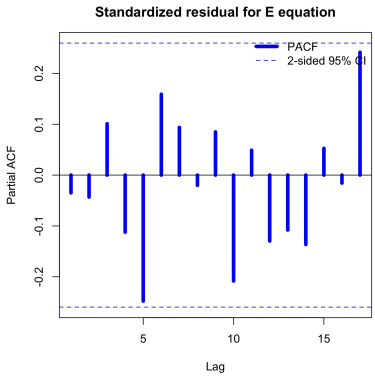
(C) PACF



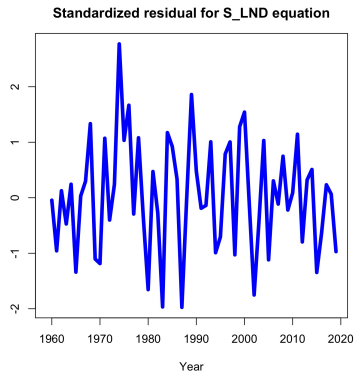
(D) Std. residuals E



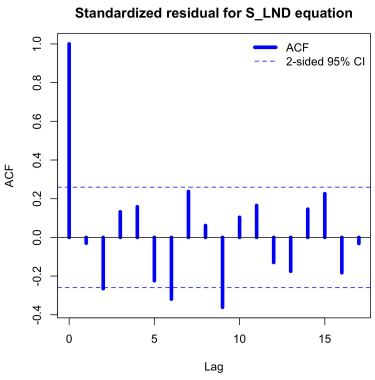
(E) ACF



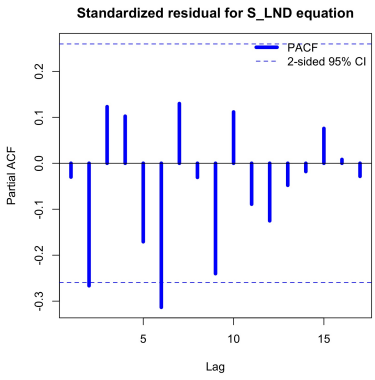
(F) PACF



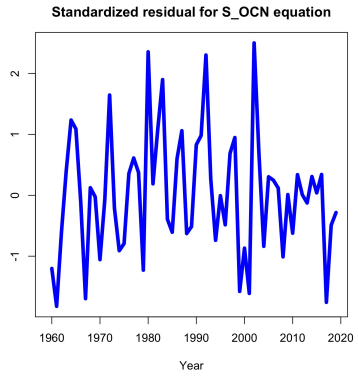
(G) Std. residuals S_LND



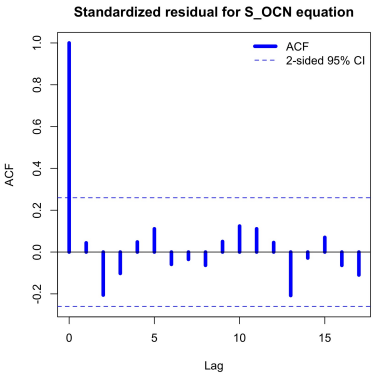
(H) ACF



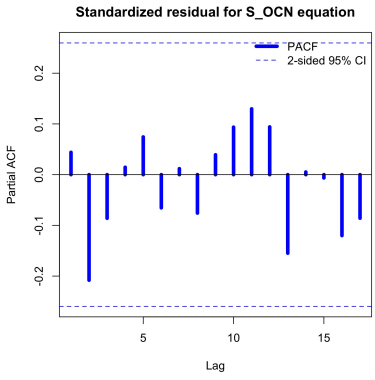
(I) PACF



(J) Std. residuals S_OCN



(K) ACF



(L) PACF

TABLE 5. **Residual diagnostics** for standardized prediction residuals from model (1) to (8). We present sample mean, standard deviation "std dev", skewness "skew", kurtosis "kurt", Jarque-Bera test for normality "JB", Ljung-Box test statistic for first-order autocorrelation "LB(1)", Durbin-Watson statistic "DW", and significance at 0.01-level "***".

Residual	mean	std dev	skew	kurt	JB	LB(1)	DW
C	-0.183	0.964	0.205	2.765	0.568	1.607	1.653
E	-0.033	1.009	-1.448	8.506	98.38***	0.073	2.060
S_LND	0.061	1.022	0.047	2.615	0.399	0.054	2.037
S_OCN	0.093	0.978	0.377	2.988	1.447	0.116	1.908

and likewise show a reasonable goodness-of-fit, the proposed model that reduces them to scaled concentrations seems simple. Since the sink time series are not measurements but calibrated model output, the variations around the trend given by concentrations are meaningful. A statistical model such as (1) to (8) is not designed to capture the rich dynamics of the dynamic global vegetation models or Earth system models behind the S_LND series. One major determinant of the dynamics beyond concentrations is the El Niño/Southern Oscillation (ENSO) cycle, however, and there are observational data available for this cycle (Climatic Research Unit, 2021; Ropelewski and Jones, 1987). We include a measure of ENSO in the sinks processes, and we opt for the Southern Oscillation Index (SOI). We have also estimated versions that include Niño 3.4 and Oceanic Niño Indices, and the results were very similar, likely because our model is specified for an annual sampling frequency, where the differences between the indices do not matter as much.

In Bennedsen et al. (2021), we have shown that U.S. CO_2 emissions can be modeled effectively by industrial production indices, leading to accurate forecasts. Friedlingstein et al. (2020) and earlier vintages of the GCB (Friedlingstein et al., 2019; Le Quéré et al., 2018, 2017) model and forecast emissions by measures of gross domestic product (GDP), following Raupach et al. (2007). The energy economics literature has discussed the relation of energy consumption and macroeconomic activity at length (e.g., Stern, 1993, 2000; Oh and Lee, 2004; Lee, 2005; Zhang and Cheng, 2009; Ozturk, 2010). Motivated by this, we replace the constant drift d in equation (2) for E by a coefficient times differences in the World Bank's World GDP series in trillion (10^{12}) constant 2010 Dollars (Series ID NY.GDP.MKTP.KD). Simply adding a coefficient times changes in World GDP to the constant drift d results in an insignificant estimated constant of $\hat{d} = 0.007$ with a standard error of 0.03 and a significant estimated coefficient on differences of World GDP of 0.11 with a standard error of 0.02. The mean difference of World GDP on the sample is 1.22 trillion dollars per year, thus $1.22 * 0.11 \approx 0.13$ corresponds to the magnitude of $\hat{d}$ in Table 4. The data thus clearly prefer differences of World GDP as the time-varying drift in the random walk for E.

We include three dummy variables in the state and measurement equations and one in the variance of the innovation $\varepsilon_{5,t}$ of the stationary error X_t^E for emissions E_t^*. The selection of the dummies was conducted using a variety of methods and criteria. The point of departure was a purely data-driven search with AutoMetrics (Doornik, 2009; Pretis et al., 2018). The set of selected dummies was modified by a search with the following criteria: (i) The number of dummies should be minimal. (ii) The dummies should have an identifying event or narrative. (iii) The dummies should be statistically significant at least at the 10% level. (iv) The numerical maximum likelihood should be reasonably close to the highest value we saw in all of our experiments. (v) The residual diagnostics as well as the filtered and smoothed states and disturbances should be reasonable.

The final set of dummy variables contains: (1) 1991 in the state equation for G_ATM captures the Pinatubo minimum (Bousquet et al., 2000; Angert et al., 2004). (2) 1991 in the state equation for E*: The collapse of the Soviet Union, the 1990 oil price shock, and the first Gulf War in 1991 are associated with a decrease in emissions. The relation between oil price crises, energy consumption, and macroeconomic activity has been discussed at length in the econometrics and energy economics literature, see, e.g., Hamilton (1983); Perron (1989); Hamilton (1996, 2003); Barsky and Kilian (2004); Kilian (2008, 2009); Stern and Kander (2012). (3) 1997 in the measurement equation for E: There is a strong spike in levels of E in 1997 due to burning of South East Asian peatlands (Houghton and Nassikas, 2017). (4) 1996 in variance: Panels [D], [E], and [F] of Figure 1 show that first differences of E inherit an increase in variance from first differences of E_LUC in 1996. This change in variance in E_LUC is displayed in all three source time series (Houghton and Nassikas, 2017; Hansis et al., 2015; Gasser et al., 2020).

Given these considerations, we propose the following specification for Model II

$$C^*_{t+1} = C^*_t + G\_ATM^*_{t+1}, \qquad G\_ATM^*_{t+1} = E^*_{t+1} - S\_LND^*_{t+1} - S\_OCN^*_{t+1} + \beta_7 I1991, \tag{16}$$

$$E^*_{t+1} = E^*_t + \beta_5 \Delta GDP_{2010,t+1} + \beta_8 I1991 + X^E_t, \tag{17}$$

$$S\_LND^*_{t+1} = c_1 + \frac{\beta_1}{C_0} C^*_{t+1} + \beta_3 SOI_{t+1}, \qquad S\_OCN^*_{t+1} = c_2 + \frac{\beta_2}{C_0} C^*_{t+1} + \beta_4 SOI_{t+1}. \tag{18}$$

The measurement equations (5) to (7) remain unchanged while (8) is amended as $E_t = E^*_t + \beta_6 I1997$ so that the $X_{4,t}$ process is removed from the equation for E based on the results in the Model I. The remaining disturbance processes $X_{i,t}$, for $= 1, 2, 3$, and X^E_t remain while the latter is subject to a break in its error variance, that is $X^E_{t+1} = \phi_E X^E_t + \eta_{4,t}$ with $\eta_{4,t} \sim N(0, \sigma^2_{\eta_4} \times (s^2_E \, I_{t \geq 1996}))$, which captures the change in variance in the E residuals from 1996 onwards.

The additional new coefficients β_j, for $j = 3, \ldots, 9$, can all be concentrated out of the log-likelihood functions by including them as constants in the state vector of the state space representation of the model. The scale coefficient s_E is added to the parameter vector ψ while $\sigma^2_{\eta_5}$ and ϕ_4 are removed from ψ. The estimation of this 12×1 parameter vector is done by maximum likelihood as described above.

TABLE 6. **Residual diagnostics** for standardized prediction residuals from Model II, see Table 5 for details.

Residual	mean	std dev	skew	kurt	JB	LB(1)	DW
C	-0.140	0.979	0.167	3.272	0.471	0.350	1.840
E	0.004	0.983	-0.269	3.095	0.759	0.062	1.928
S_LND	0.164	0.995	0.285	2.530	1.387	0.223	2.111
S_OCN	0.120	0.951	0.440	3.394	2.360	0.648	2.191

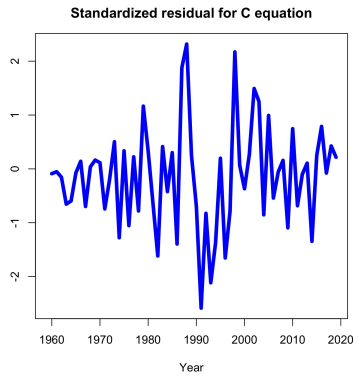
5.4. **Model II: Results and residual diagnostics.** In Table 6 and Figure 5 we present the residual diagnostics from Model II. There is no evidence of non-normality remaining in the standardized prediction residuals, and no evidence of serial correlation. The correlograms in Figure 5 show no evidence of serial correlation, except from two marginally significant spikes in the PACF of the C equation at higher lags. All autocorrelation coefficients for the sinks processes, in particular, are far below the significance bands. We conclude that Model II provides a good statistical description of the data.

Table 7 reports the parameter estimates of Model II and Figure 6 shows the corresponding smoothed states, together with the data series for comparison, where available. In Table 7, the basic patterns we have described above for Table 4 remain. The coefficients pertaining to SOI in the sink processes are highly significant. They are of opposite sign: La Niña phases (positive SOI) correspond to higher land uptake whereas they correspond to lower ocean uptake. This is to be expected, see, for example, Feely et al. (1999) and Haverd et al. (2018). The increase in variance of $\eta_4 = \eta^E$ in 1996 is highly significant and more than quintuples the pre-1996 variance. The dummy variables are all of the expected sign. Figure 6 shows that the inclusion of SOI in the sinks makes them much more dynamic, tracing the data better. G_ATM inherits the enhanced dynamics and also presents a better fit.

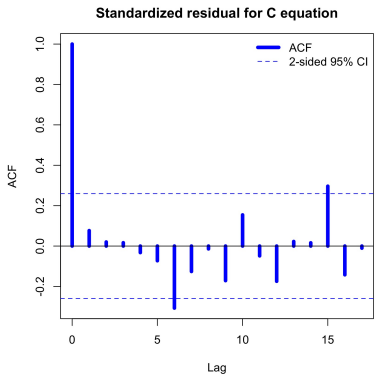
5.5. **Discussion of empirical results.**

5.5.1. *Limitations.* Section 2 discussed common models for the sink processes and how they can be approximated by specifications that are linear in concentrations (Bacastow and Keeling, 1973; Gifford,

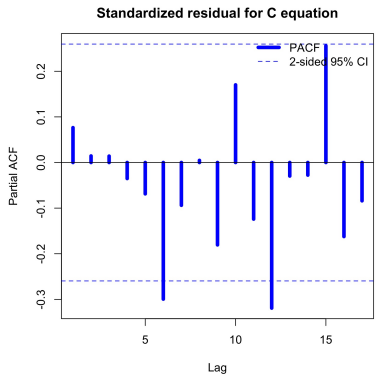
FIGURE 5. Standardized prediction residuals for Model II, ACF autocorrelation function, PACF partial autocorrelation function. Note that the ACF shows the value for lag zero (equal to one by definition), while the PACF does not.



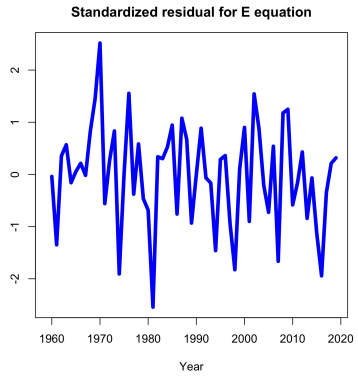
(A) Std. residuals C



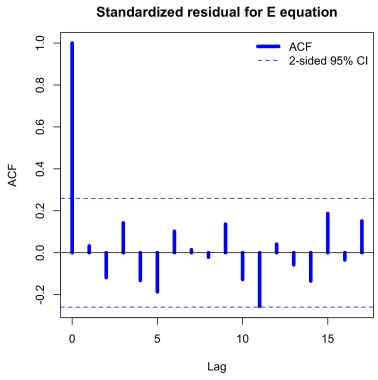
(B) ACF



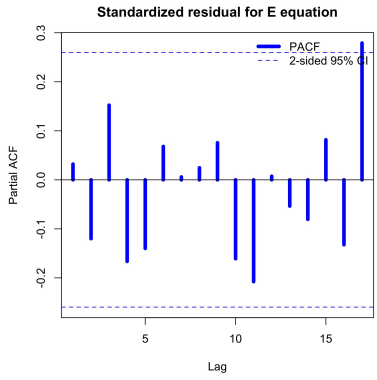
(C) PACF



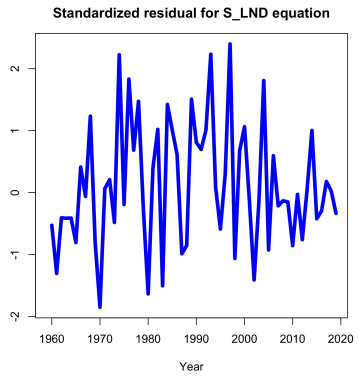
(D) Std. residuals E



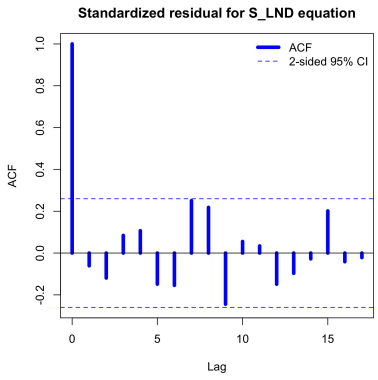
(E) ACF



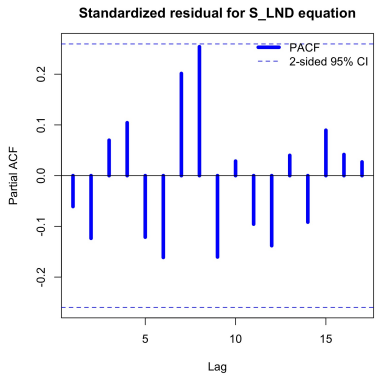
(F) PACF



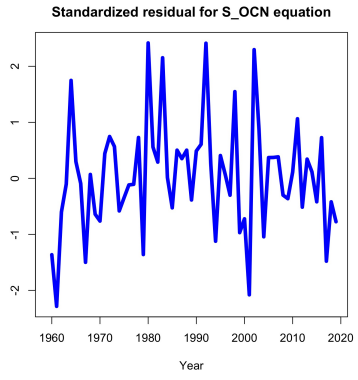
(G) Std. residuals S_LND



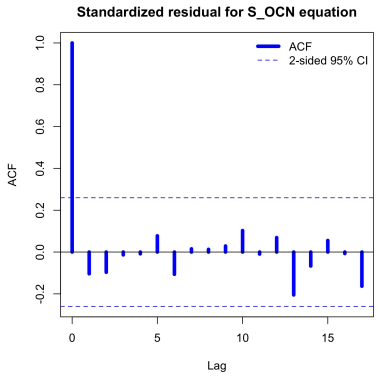
(H) ACF



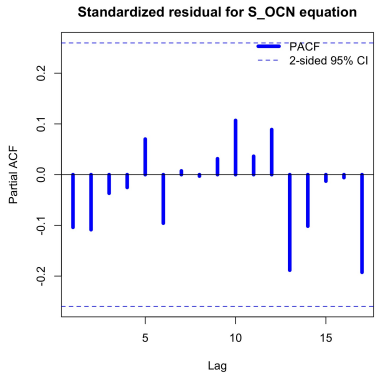
(I) PACF



(J) Std. residuals S_OCN



(K) ACF



(L) PACF

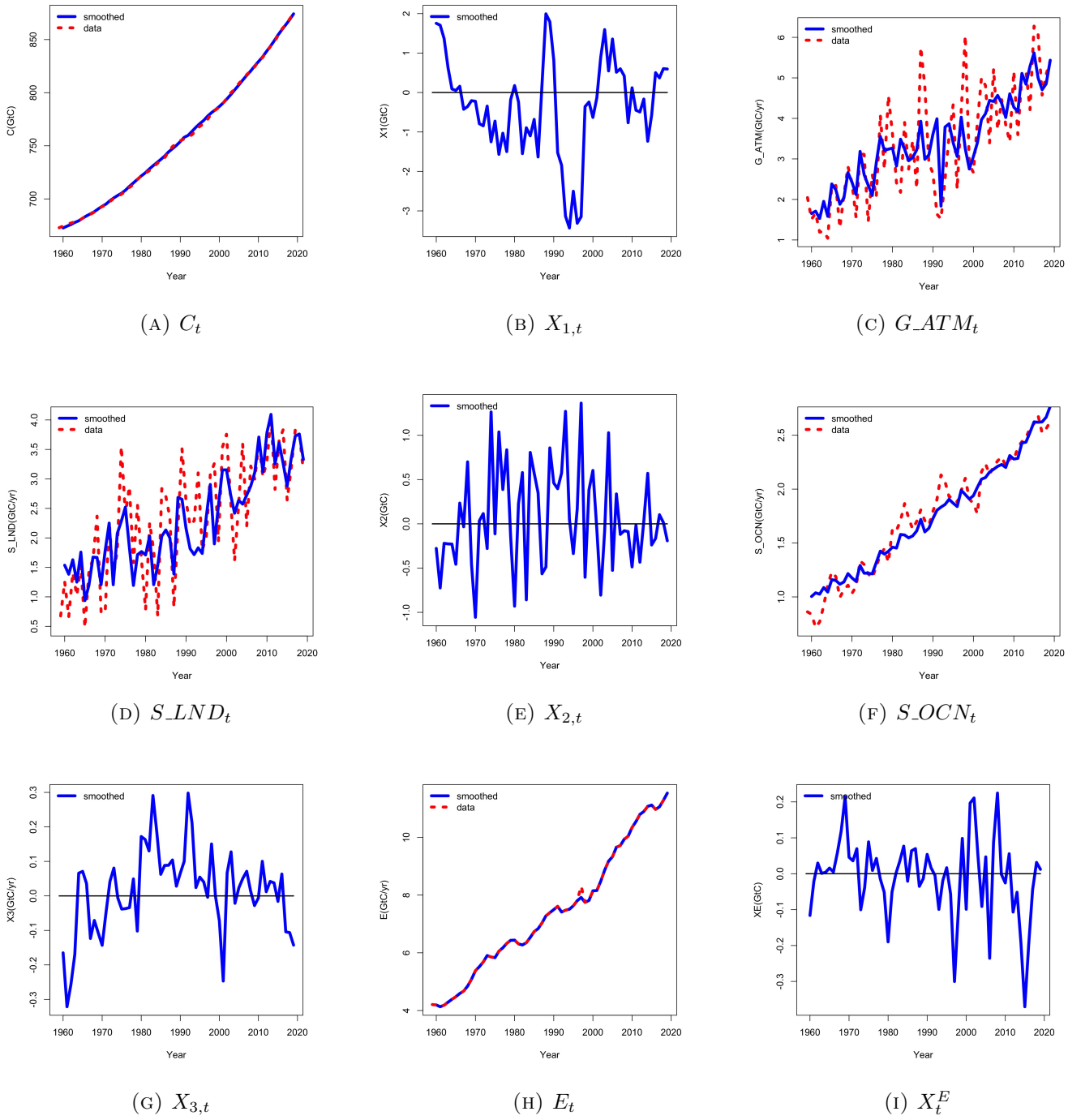
TABLE 7. **Parameter estimates** for Model II, see Table 4.

Coefficients			Variances		
	estimate	std err		estimate	std err
c_1 (filt.)	-6.70	0.04	$\sigma^2_{\eta_1}$	0.66	0.13
c_2 (filt.)	-4.63	0.03	$\sigma^2_{\eta_2}$	0.32	0.06
β_1	7.12	0.49	$\sigma^2_{\eta_3}$	0.01	0.002
β_2	4.99	0.29	$\sigma^2_{\eta_4}$	0.004	0.001
β_3 (filt.)	0.57	0.09	r_{12}	-0.60	0.09
β_4 (filt.)	-0.06	0.02	r_{13}	-0.01	0.12
β_5 (filt.)	0.11	0.01	s_E	2.28	0.51
β_6 (filt.)	0.39	0.06			
β_7 (filt.)	-2.04	0.66			
β_8 (filt.)	-0.29	0.06			
ϕ_1	0.77	0.08			
ϕ_3	0.62	0.12			
ϕ_E	0.39	0.12			

1993). This means, on the other hand, that there is approximation error in the sink processes in Equations (3) and (4). The residual diagnostics are encouraging in the sense that there is little evidence of serial correlation remaining, but this may change as new data with higher atmospheric concentrations are observed, which will increase the influence of second-order terms in the Taylor approximations in equations (10) and (13). Also, dependence in higher moments cannot be detected with these measures of linear dependence.

Since model parameters are estimated on the historical sample, they necessarily reflect the technological conditions that prevailed during the sample period. Forecasts or scenario analyses conducted with the model write these technological conditions forth through the use of the estimated parameters. This holds in particular for the fuel mix that is behind emissions E (e.g., Friedlingstein et al., 2020). Hence, Model II cannot shed any light on the effects of changes in the energy mix to, say, higher shares of renewable sources. Note that while this is a limitation of Model II, it is not a limitation of the state space model class or the general modeling strategy followed in this paper. One could specify a model that instead of emissions uses consumption quantities of fossil and renewable energy carriers as exogenous drivers. Then, observations, forecasts, and scenarios of changes in the consumption quantities of different energy carriers could be employed as input to the model instead of emissions. We have decided not to pursue this avenue since we wanted to engage the time series of the GCB, including emissions. Emissions are in essence linear combinations of the consumed energy carrier quantities, weighted by the physical coefficients that specify how much CO_2 is released by consumption of one unit of the respective carriers (Marland and Rotty, 1984). Note that because of this definition of emissions, it is not feasible to specify fuel mix variables as explanatory variables

FIGURE 6. **Smoothed estimates** for state vector elements from Model II.



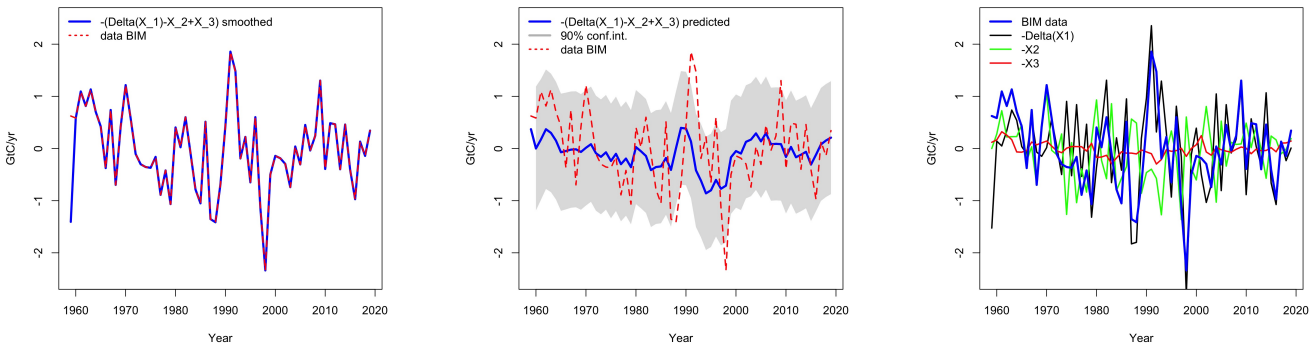
(A) C_t (B) $X_{1,t}$ (C) G_ATM_t

(D) S_LND_t (E) $X_{2,t}$ (F) S_OCN_t

(G) $X_{3,t}$ (H) E_t (I) X_t^E

on the right-hand side of the E equation, since this will lead to definitional overlap on the left-hand and right-hand side variables. The construction of emissions data is, on the other hand, independent of macroeconomic national accounting and the construction of World GDP data, so that there is no definitional overlap in our specification of the E equation. To conduct scenario analysis in Section 6, we shall make use of the versatility of the state space model framework and specify a time-varying slope coefficient for the emissions equation in order to capture changes in the fuel mix that are projected to decouple emissions from economic growth (Haberl et al., 2020).

FIGURE 7. Budget imbalance BIM, smoothed values, one-year ahead predictions, and components



(A) Smoothed $-(\Delta X_1 + X_2 + X_3) + \beta_6 I1997 - \beta_7 I1991$

(B) One-year ahead predictions

(C) Components $-\Delta X_1$, $-X_2$, $-X_3$

5.5.2. *Budget Imbalance.* The model facilitates the study of quantities of interest that have been discussed in the literature. For example, Friedlingstein et al. (2020) define and discuss the budget imbalance BIM, which is the residual of the GCB equation. The budget imbalance is an important time series in its own right. For instance, it can be used to assess whether the different GCB data sources are internally consistent (Friedlingstein et al., 2020). Moreover, it has recently been suggested that the future budget imbalance may be used to detect potential misreportings of global CO_2 emissions (Peters et al., 2017). Hence, accurate modelling and prediction of the budget imbalance is an important endeavour.

Model II implies $BIM^* = E^* - G\_ATM^* - S\_LND^* - S\_OCN^* = 0$ by way of the equation for G_ATM^*. The corresponding measure of the imbalance in Model II is

$$\begin{aligned} BIM_t &= E_t - \Delta C_t - S\_LND_t - S\_OCN_t, \\ &= E_t^* - G\_ATM_t^* - S\_LND_t^* - S\_OCN_t^* + \beta_6 I1997 - (X_{1,t} - X_{1,t-1}) - X_{2,t} - X_{3,t}, \\ &= \beta_6 I1997 - \beta_7 I1991 - \Delta X_{1,t} - X_{2,t} - X_{3,t}. \end{aligned}$$

Figure 7 [A] shows the smoothed values of this process together with the data. There is perfect overlap: Since the state processes E*-G_ATM*-S_LND*-S_OCN* yield zero by definition of the equation for G_ATM*, all data variation in BIM must be absorbed by the X-processes, which are freely competing for the data variation in the estimation. Note that $X_{4,t}$ in equation (8) for E was found to be insignificant, indicating that there is no stationary variation in emissions that contributes to

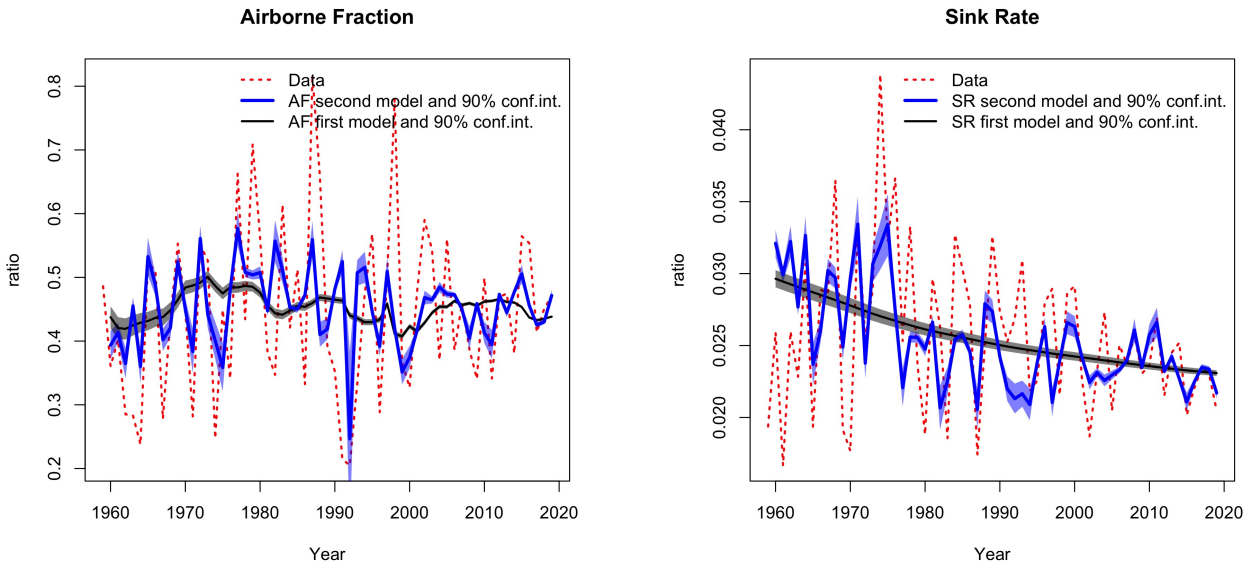
BIM. From the estimation of the X-processes we can conclude that BIM is stationary, since it is a linear combination of processes that were all tested to be stationary.

TABLE 8. Empirical covariance matrix of the smoothed values of the components of BIM

	ΔX_1	X_2	X_3
ΔX_1	0.767	-0.258	0.001
X_2	$\star$	0.313	0.010
X_3	$\star$	$\star$	0.015
Variance decomposition	0.85	0.11	0.04

Figure 7 [B] shows the one-year ahead prediction for BIM obtained from the one-year ahead predictions of the X-processes, together with 90% pointwise confidence intervals. Figure 7 [C] shows the individual components $-\Delta X_1$, $-X_2$, and $-X_3$. The panel indicates visually that most of the variation originates in $-\Delta X_1$ and $-X_2$. Table 8 reports the empirical variances and covariances of the smoothed component processes $-\Delta X_1$, $-X_2$, and $-X_3$ of BIM. The variance decomposition confirms the visual inspection of the components: $-\Delta X_1$ (G_ATM) contributes 85% and X_2 (S_LND) 11% to the total variation of BIM. The contribution of X_3 (S_OCN) is smallest at 4%.

FIGURE 8. **Airborne fraction and sink rate**, their smoothed estimates from both Model II and Model I, with the data for comparison



(A) Airborne fraction

(B) Sink rate

5.5.3. *Airborne fraction and sink rate.* Figure 8 shows the airborne fraction in panel [A] and the sink rate in panel [B]. The airborne fraction

$$AF = \frac{G\_ATM}{E}$$

is the part of emissions that remains in the atmosphere. On the sample, this fraction is on average 0.44, but is shows substantial variation in the data. The sink rate

$$SR = \frac{S\_LND + S\_OCN}{C}$$

is a measure of the capacity of the sinks to absorb atmospheric CO_2. Whether or not the airborne fraction is increasing and/or the sink rate is decreasing has been subject of debate (Canadell et al., 2007a; Raupach et al., 2008; Knorr, 2009; Le Quéré et al., 2009; Gloor et al., 2010; Raupach et al., 2014; Rayner et al., 2015; Ballantyne et al., 2015; Bennedsen et al., 2019b).

Figure 8 shows, however, that by calculating AF^* and SR^*, the versions of the ratios with *-processes in the numerators and denominators, we obtain new estimates of these variables that focus on a few key relations of the sinks, and thus they show much less variation. Raupach et al. (2014) and Bennedsen et al. (2019b) show that there is no evidence of an increase in the airborne fraction but that there is evidence of a decline in the sink rate. In Figure 8 this can be seen with the naked eye, in particular for the results from model (1) to (8) that reduces the sinks to linear regressions on concentrations. The figure also shows confidence intervals for both models obtained from simulation smoothing (Durbin and Koopman, 2002). Note that these confidence intervals should not be interpreted such that the data should fall within the boundaries with a certain coverage probability. Rather, their correct interpretation is that if we were given a large number of trajectories sampled from the models, and if we were to extract the smoothed state variables and compute the airborne fraction and the sink rate from these trajectories, the confidence bands will cover these imputed variables 90% of the time.

6. Validation, forecasting and future scenarios

The exogenous drivers of Model II are World GDP changes and SOI. Based on given trajectories of these two exogenous variables, all other model variables, being emissions E, land sink S_LND, ocean sink S_OCN, growth of atmospheric concentrations G_ATM, and atmospheric concentrations C, can be determined when ignoring the Gaussian random processes. This facilitates three lines of inquiry: (1) validation, (2) forecasting, and (3) scenario analysis. In validation exercises, the model is estimated on an estimation subsample, and conditional on the data for the exogenous elements, the model is employed to produce implied values for the endogenous variables on a validation subsample that can then be compared to the observations. In forecasting exercises, forecasts of the exogenous drivers are obtained, either by specifying forecast models for them or by getting them from external

sources, and they are plugged into the system for generating forecasts of the endogenous variables. In scenario analyses, usually for long-term projections, artificial trajectories for the exogenous drivers are assumed that are converted, via the model, to corresponding artificial trajectories for the endogenous variables. We discuss these three approaches in the following subsections.

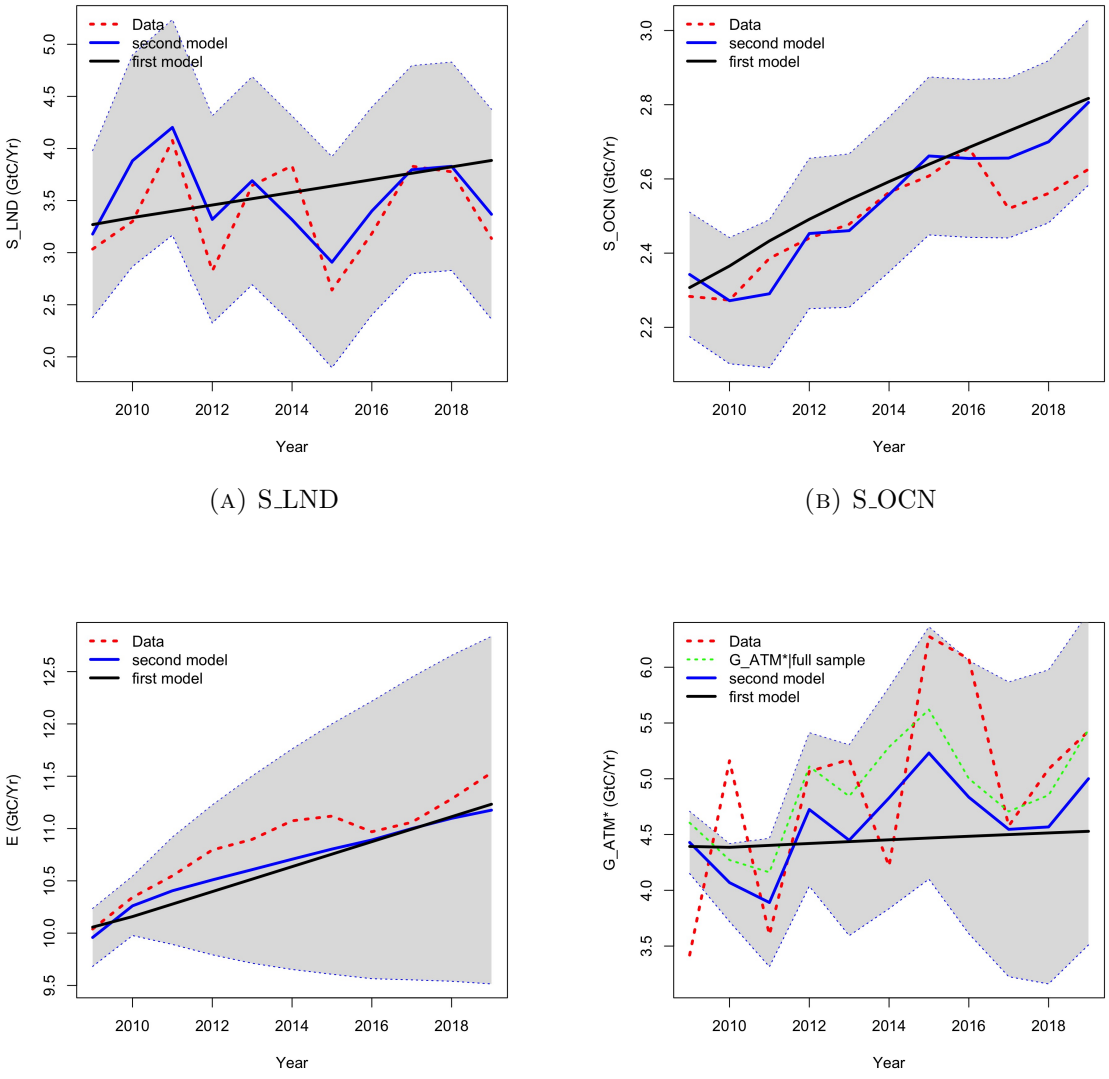
The presented results are all obtained from the Kalman smoother, which is applied to the state space specifications of Models I and II. The smoothed estimates of the state vector are obtained together with the corresponding variances, and the unknown parameters are replaced by their maximum likelihood estimates; the details are provided in (Durbin and Koopman, 2012, Chapter 4). The projections for out-of-sample periods are computed by treating the observations for these periods as missing. It is well established that the Kalman filter and related methods can treat missing values without further complications. In all the three lines of inquiry, the results are obtained by the application of these methods.

6.1. **Validation.** We conduct a validation exercise on the last 10 years of the sample, that is, the validation subsample extends from 2010 to 2019. The parameters are estimated for the period 1959–2009. Using the estimated parameters and the values of $\Delta GDP_{2010,t}$ and SOI_t for the validation sample 2010–2019, the implied values of the system are computed for these 10 years. This results in implied values both for the measurement and for the state variables.

Figure 9 shows the results for the four variables S_LND, S_OCN, E, and G_ATM*. Note that the first three variables are in the measurement equation while G_ATM* is a state variable. We have chosen to present G_ATM* instead of C because there are more interesting dynamics in changes of C, and the differences between the models are easier to distinguish than in levels of C.

The validation is performed on both Model I (black) as well as Model II (blue); the pointwise 90% confidence intervals are based on Model II. In panels [A] to [C], the benchmarks all lie within the point-wise 90% confidence bands. For G_ATM*, the 2010 data point falls outside the confidence bands. Note, however, that confidence bands for state processes should not actually be compared against data points but against the estimated state process given the whole sample. This is plotted in the dashed green curve, which does fall within the confidence bands. Black and blue lines show that Model II captures more of the dynamics in the benchmark (and in the data) than Model I, equations (1) to (8). For E, there is little difference in the fitted values from either model, indicating that the enriched dynamics of Model II are due to the dependence on SOI. For the sink states, there is a marked difference between the models (panels [A] and [B]), with the extended specification capturing the data substantially better.

FIGURE 9. **Validation exercise on subsample 2010–2019** for Models I and II.



(A) S_LND (B) S_OCN

(C) E (D) G_ATM*

In summary, the validation exercise shows that Model II provides a good statistical description of the GCB data. On a validation sample of 10 years, it can generate implied values and confidence bands from parameters estimated on the first 50 years and from data on ΔGDP and on SOI such that the data and the smoothed states obtained from the full sample lie reasonably within confidence bounds.

6.2. **Forecasting.** Given the data sample period from 1959 to 2019, we carry out a forecast/nowcast analysis based on Models I and II, for the periods 2020 to 2022. This necessitates forecasting World GDP changes and the Southern Oscillation Index (SOI).

For World GDP growth, we resort to the forecasts regularly provided and updated by the International Monetary Fund (IMF) and the World Bank (WB). Table 9 shows hindcasts for 2020, reported

in IMF (July 2019) and World Bank (June 2019), and nowcasts and forecasts for 2021 and 2022, respectively, reported in IMF (June 2020) and World Bank (June 2020).

TABLE 9. Forecasts of World GDP growth provided by International Monetary Fund and World Bank (in percent). All hindcasts, nowcasts, and forecasts are from IMF (Jan 2021) and World Bank (Jan 2021a). Data are from World Bank (2021b).

	2020	2021	2022
IMF	-3.5%	5.5%	4.2%
World Bank	-4.3%	4.0%	3.8%

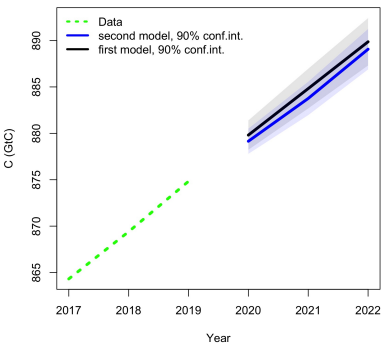
We forecast the monthly Southern Oscillation Index (SOI) obtained from Climatic Research Unit (2021) and Ropelewski and Jones (1987) using the structural time series model of Harvey (1989). This unobserved components time series model consists of a level component, a monthly seasonal component, a second-order stochastic cycle (with the cycle-period being estimated as close to 4 years), and a first-order autoregressive component; similar specifications are adopted in Durbin and Koopman (2012, Section 3.2) and Azevedo et al. (2006). Further details of the model and the estimation results are presented in Appendix B. This model is used for the generation of the forecasts for SOI.

TABLE 10. **Forecasts for** C**,** G_ATM**,** E**,** S_LND**, and** S_OCN from Models I and II, given forecasts of World GDP growth and SOI. Forecast standard errors in parentheses.

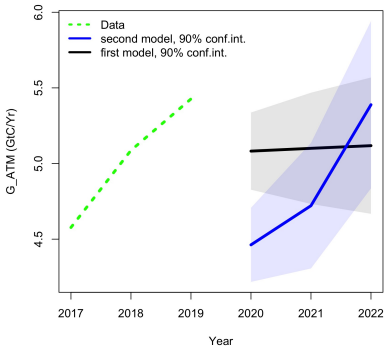
	2020			2021			2022		
	Model II		Model I	Model II		Model I	Model II		Model I
	IMF	WB		IMF	WB		IMF	WB	
C (GtC)	879.15 (0.829)	879.07 (0.829)	879.80 (0.964)	883.76 (1.097)	883.47 (1.097)	884.82 (1.303)	889.07 (1.329)	888.53 (1.329)	889.86 (1.556)
G_ATM (GtC/Yr)	4.462 (0.149)	4.388 (0.149)	5.082 (0.155)	4.721 (0.253)	4.510 (0.253)	5.100 (0.224)	5.389 (0.337)	5.136 (0.337)	5.118 (0.275)
E (GtC/Yr)	11.212 (0.152)	11.136 (0.152)	11.663 (0.159)	11.721 (0.261)	11.505 (0.261)	11.786 (0.231)	12.129 (0.351)	11.865 (0.351)	11.908 (0.286)
S_LND (GtC/Yr)	4.015 (0.686)	4.014 (0.569)	3.793 (0.675)	4.243 (0.569)	4.240 (0.569)	3.854 (0.675)	3.895 (0.569)	3.889 (0.569)	3.915 (0.675)
S_OCN (GtC/Yr)	2.647 (0.097)	2.647 (0.097)	2.707 (0.108)	2.701 (0.115)	2.698 (0.115)	2.776 (0.130)	2.811 (0.121)	2.807 (0.121)	2.838 (0.140)

Table 10 and Figure 10 show the forecasts together with estimates of the forecast uncertainty (forecast standard errors in Table 10 and 90% point-wise confidence bands in Figure 10). Figure 10 only shows the forecast from the IMF forecasts of World GDP. On the three points of the forecast sample, the forecasts from the two models are distinct in a number of ways. In forecasting atmospheric concentrations C, Model II generates lower forecasts due to the effects of the Covid pandemic. This is also reflected in lower forecasts of emissions E and subsequent changes in atmospheric concentrations G_ATM. The SOI is predicted to switch from a La-Niña period (positive numbers) in 2020 and 2021 to an El-Niño period (negative numbers) in 2022. These dynamics are reflected in the forecasts of

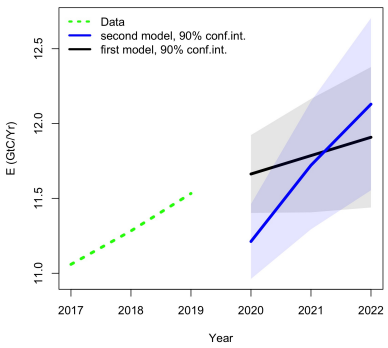
FIGURE 10. **Forecasts for C, G_ATM, E, S_LND, and S_OCN** for Model II (blue) and for Model I (black), with 90% pointwise confidence intervals. Model II employs forecasts of SOI generated from the model described in Appendix B and forecasts of World GDP changes from the World Bank and the IMF.



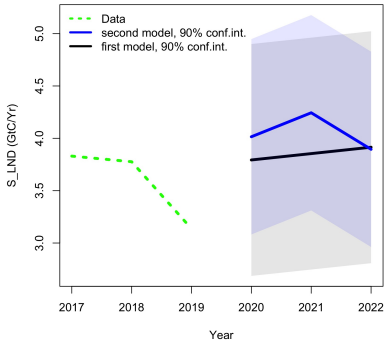
(A) C - IMF



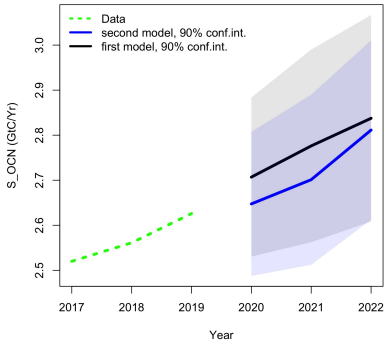
(B) G_ATM* - IMF



(C) E - IMF



(D) S_LND - IMF

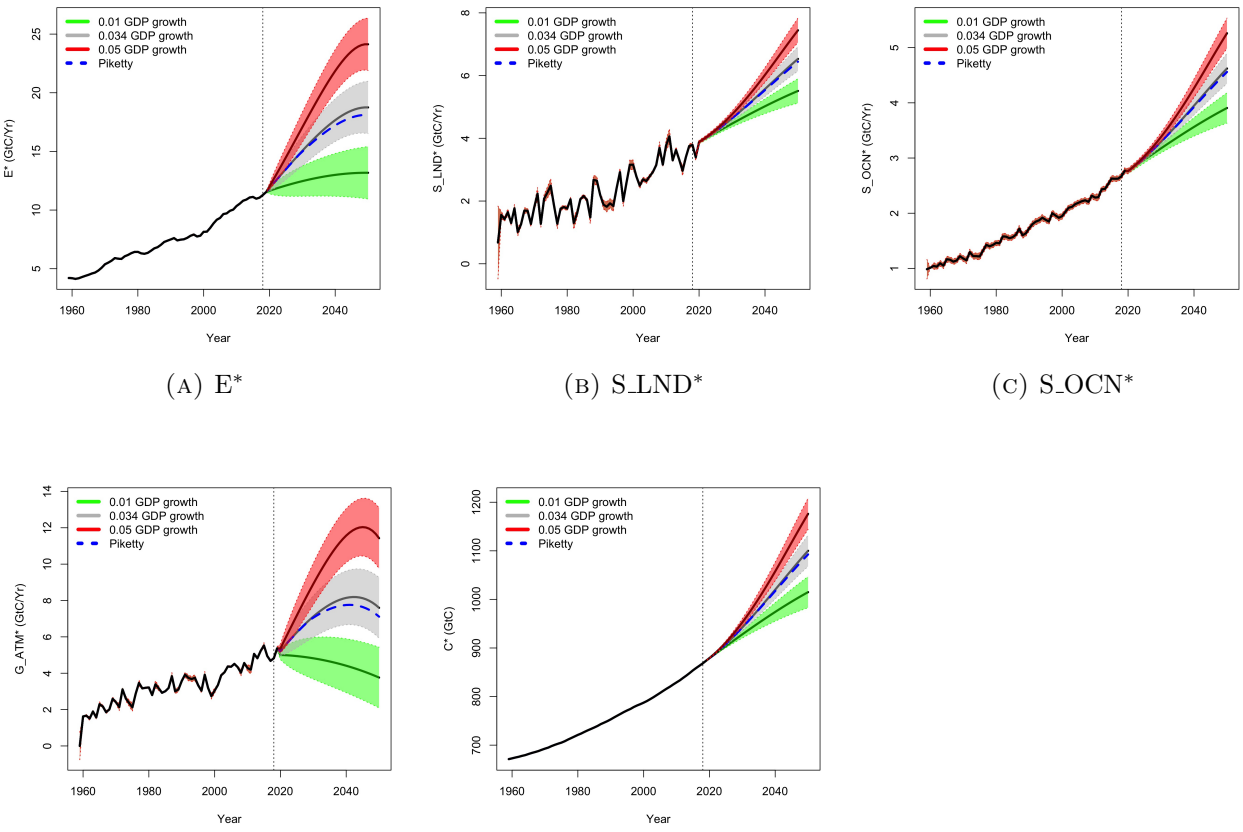


(E) S_OCN - IMF

G_ATM*, E, S_LND, and S_OCN: Model II forecasts G_ATM to decrease in 2020, to increase some in 2021 and then substantially in 2022. The forecast implied by the IMF's GDP forecast reaches 2019 G_ATM levels again in 2022; the forecast implied by the World Bank's numbers falls short of that level by about 0.3 GtC/Year. The sink processes are forecast to react moderately to these changes in the exogenous variables, and the differences from the GDP forecasts are very small. Model I does not take any of this information into account and simply extrapolates from the 1959-2019 dynamics.

6.3. **Scenario analysis.** In this section, we conduct scenario analyses based on assumed paths for World GDP and we present long-term projections of emissions, sinks activity, and atmospheric concentrations for the period 2020–2050. In all scenarios, SOI is set to zero, implying that we are not using scenarios for the ENSO cycle. Given an assumed GDP path for the period up to 2050, we treat the other variables as missing for this future period and obtain estimates of the model variables using the Kalman smoother.

FIGURE 11. Scenario analysis for different World GDP growth paths with decreasing emission intensity of GDP and without additional emission reductions



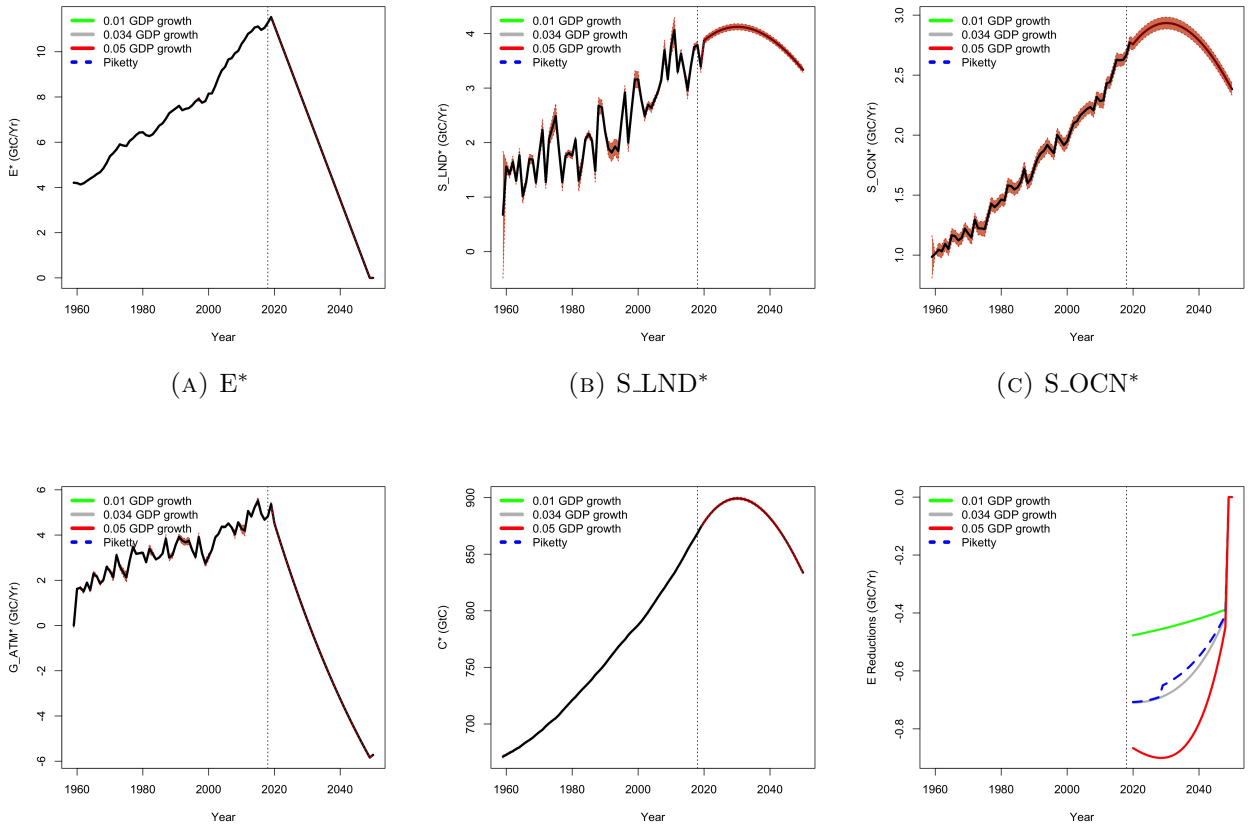
(A) E* (B) S_LND* (C) S_OCN*

(D) G_ATM* (E) C*

We consider four scenarios for World GDP growth 2020–2050: (1) constant annual growth of 0.01, (2) constant annual growth equal to the historical mean of 0.034 during the 1959–2019 period, (3) constant annual growth equal to 0.05, and (4) a growth scenario employed in Piketty (2014) of 0.034 until 2030 and 0.03 until 2050.

The emissions equation (17) is defined to be broadly consistent with the Kaya identity in first differences (Kaya and Yokobori, eds, 1997; Raupach et al., 2007). Therefore, the parameter β_5 can be interpreted as the emission intensity of World GDP. For the purpose of scenario analysis, we assume that this emission intensity declines linearly to zero between 2020 and 2050. This assumption reflects the finding that the emission intensity has been decreasing for many countries (Friedlingstein et al., 2020; Andrew, 2021) and that it is an important policy goal in the transition to more sustainable growth (Haberl et al., 2020). The linear decline to zero by 2050 is motivated by the scenarios discussed in Masson-Delmotte et al. (2018). It is a simplification that abstracts from complexities such as negative emissions or rebound effects (Brockway et al., 2021; Bruns et al., 2021).

FIGURE 12. Scenario analysis for emission reductions consistent with 1.5 degree warming



(A) E* (B) S_LND* (C) S_OCN*

(D) G_ATM* (E) C* (F) Emission reductions

We conduct two different sets of scenario analyses. In the first set, the only process working towards more sustainability is the decreasing emission intensity. In particular, there are no further emission reductions. In the second set of scenarios, in addition to the declining emission intensity we consider emission reductions that are consistent with 1.5 degree global warming targets, similar to the scenarios laid out in Masson-Delmotte et al. (2018) and summarized in UNFCCC (2021).

The results of the first set of scenarios are shown in Figure 11. Emissions follow three clearly separated cones for the three constant growth scenarios, and the Piketty scenario is close to the 0.034 scenario, as to be expected. Towards the end of the projection period close to 2050, emissions level out. This happens because the emission intensity is near zero and World GDP is essentially decoupled from emissions. Changes in atmospheric concentrations G_ATM* reach a maximum around 2045 in the 0.034 and 0.05 scenarios and decline thereafter. This is because emissions level out while the sinks remain highly active according to their linear dependence on the high level of concentrations. The concave shape of G_ATM* with a maximum thus depends on the assumption that the sinks do not saturate. The mean implied atmospheric concentrations in 2050 for the four scenarios are (1)

1015GtC (477ppm) for the 0.01 scenario, (2) 1100GtC (517ppm) for the 0.034 scenario, (3) 1176GtC (552ppm) for the 0.05 scenario, and (4) 1092GtC (514ppm) for the Piketty scenario.

The results of the second set of scenarios are shown in Figure 12. Here, the path for emissions is set to decline from 2020 to 2050 linearly to zero, broadly in line with the recommendations set forth in Masson-Delmotte et al. (2018) to keep global warming to 1.5 degrees over the pre-industrial level. We impose this path on the emissions variable in the measurement equation (8) and run the Kalman recursions. The Kalman smoother gives the implied paths for the sinks and the growth of atmospheric concentrations. Note that the paths are actually plotted for all four growth scenarios with their confidence bands, but since the emission path is given deterministically, there is little room for variation or uncertainty. Most uncertainty remains in the sinks activity, but even this is very limited. The sinks and the atmospheric concentrations peak in 2030, and decline thereafter. Atmospheric concentrations reach a maximum of 899GtC (422ppm) in 2030 and decline to a level of 834GtC (392ppm) by 2050.

Since, in this model, economic activity generates CO_2 emissions, economic growth will result in a greater mitigation burden than that implied by the ostensibly required yearly mitigation target given by $E_t - E_{t-1}$. In effect, as long as the economy has not been fully decoupled from emission generating activities, economic activity will work against other emission mitigation efforts. In this scenario exercise, this is captured by the process $X_t^E = E_t - E_{t-1} - \beta_5 \Delta GDP_t$, which precisely denotes the necessary level of emissions reductions needed to accommodate the emissions path E_t, shown in panel [A] of Figure 12, *after* taking into account emissions from economic activity, here captured by changes in World GDP. Thus, in this scenario exercise, the term X_t^E measures the mitigation burden implied by the imposed E path (panel [A]) *and* the scenario of future economic activity (GDP growth). Similarly to the other state variables, the Kalman smoother can also be applied to extract the implied path of X_t^E on the interval 2020-2050, which is shown in the last panel of Figure 12. It is evident that the magnitude of the required reductions depends heavily on future economic activity, with higher growth rates of World GDP implying the need for larger emissions reductions from 2020 to 2050. For instance, even in the present case, where we optimistically assume that emissions become completely decoupled from economic activity in 2050, there is still an almost 100% increase in the mitigation burden when comparing the low growth to the high growth scenario. These findings highlight the role of future economic activity in the context of emissions mitigation efforts. In particular, we see how economic growth will, at least in the short term, work against mitigation

efforts and thus affect the societal adjustments needed to reach prescribed emissions targets, such as those implied by the Paris Agreement.

7. Conclusions and directions for further research

We proposed a statistical model for the global carbon budget, consisting of the time series variables: atmospheric CO_2 concentrations, anthropogenic CO_2 emissions, and land and ocean CO_2 uptake (sinks). The cornerstone of the model is the budget equation, which ensures that the fraction of emissions that is not absorbed by the terrestrial biosphere and the ocean constitutes an annual flow to the stock of atmospheric concentrations. We discussed the central assumptions of random walk with drift dynamics for anthropogenic emissions and linear dependence of sinks on atmospheric concentrations. The model equations then allow for a closed-form solution for atmospheric concentrations that reveal stochastic integration of order one with a second near unit root. This shows that the deterministic trend in concentrations is linear, but will on finite samples appear quadratic due to the second near unit root. We also discussed that the second root moves closer to unity with increasing atmospheric concentrations.

In a Monte Carlo study we showed that parameter estimation by maximum likelihood (facilitated by Kalman filter methods) has good properties even on finite samples. In a empirical study for the global carbon budget data set as provided by the Global Carbon Project, we presented the model parameter estimates, residual diagnostics, and projections of the model variables. Based on this model analysis, we decomposed the variation in the budget imbalance into contributions from concentrations, land sink, and ocean sink. We discussed the airborne fraction and sink rate implied by our model.

We presented a number of out-of-sample exercises and pseudo-out-of-sample exercises. First, we re-estimated the model parameters using the first 50 years of data and validated these on the last 10 years of data. Second, we adopted the model, together with external forecasts of World GDP and SOI, to forecast anthropogenic emissions, atmospheric concentrations, land sink, and ocean sink for three years (one year of nowcasts, two years of forecasts). Finally, we conducted two sets of scenario analyses with different assumptions on declining emission intensity of World GDP and emission reductions and show the implied paths for sinks and atmospheric concentrations.

We plan several directions for extending the model. Given the current model structure, it is conceptually straightforward to include individual ensemble members rather than the averages of them for S_LND and S_OCN The model structure is also conducive for increasing the resolution on

the macroeconomic sphere. For example, we can replace World GDP in emissions by factors obtained from large macroeconomic data sets. The mechanics of the global carbon cycle can be extended to include elements of widely used small-scale climate models. The model structure can be extended by energy balance modules to provide a modeling connection to global temperatures.

Appendix A. System matrices

Motivated by the findings in Section 4, which are summarily listed in the beginning of Section 5, we specify the X_i and X^E processes as follows.

$$X_{1,t+1} = \phi_1 X_{1,t} + \eta_{1,t+1}, \tag{19}$$

$$X_{2,t+1} = \eta_{2,t+1}, \tag{20}$$

$$X_{3,t+1} = \phi_3 X_{3,t} + \eta_{3,t+1}, \tag{21}$$

$$X_{4,t+1} = \eta_{4,t+1}, , \tag{22}$$

$$X^E_{t+1} = \phi_E X^E_t + \eta_{5,t+1}, \tag{23}$$

where $\phi_{1,3} \in (-1, 1)$. Defining the measurement vector as $y_t = (C_t, E_t, S\_LND_t, S\_OCN_t)'$ and the state vector as $\alpha_t = (C^*_t, G\_ATM^*_t, S\_LND^*_t, S\_OCN^*_t, X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t}, X^E_t, E^*_t)'$, the model can be represented as

$$\begin{aligned} y_t &= Z\alpha_t, \\ B\alpha_{t+1} &= \tilde{c} + \tilde{T}\alpha_t + \tilde{\eta}_t, \end{aligned} \tag{24}$$

where the matrix B contains the contemporary relations: G_ATM depending on concurrent E, S_LND, and S_OCN, the sinks depending on concurrent C.

Pre-multiplying the state equation with B^{-1} transforms model (24) to standard state space form:

$$\begin{aligned} y_t &= Z\alpha_t, \\ \alpha_{t+1} &= c + T\alpha_t + \eta_t. \end{aligned} \tag{25}$$

Explicitly, (25) is given by the measurement equation

$$\begin{bmatrix} C_t \\ E_t \\ S\_LND_t \\ S\_OCN_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} C^*_t \\ G\_ATM^*_t \\ S\_LND^*_t \\ S\_OCN^*_t \\ X_{1,t} \\ X_{2,t} \\ X_{3,t} \\ X_{4,t} \\ X^E_t \\ E^*_t \end{bmatrix}.$$

The state equation is

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -\frac{\beta_1}{C_0} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\beta_2}{C_0} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C^*_{t+1} \\ G\_ATM^*_{t+1} \\ S\_LND^*_{t+1} \\ S\_OCN^*_{t+1} \\ X_{1,t+1} \\ X_{2,t+1} \\ X_{3,t+1} \\ X_{4,t+1} \\ X^E_{t+1} \\ E^*_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_E & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} C^*_t \\ G\_ATM^*_t \\ S\_LND^*_t \\ S\_OCN^*_t \\ X_{1,t} \\ X_{2,t} \\ X_{3,t} \\ X_{4,t} \\ X^E_t \\ E^*_t \end{bmatrix}$$

$$+ \begin{bmatrix} 0 \\ 0 \\ c_1 \\ c_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ d \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ 0 \end{bmatrix}.$$

It holds that

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ -\frac{\beta_1}{C_0} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{\beta_2}{C_0} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{-(\beta_1^*+\beta_2^*)}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{\beta_1^*}{c} & \frac{\beta_1^*}{c} & \frac{1+\beta_2^*}{c} & \frac{-\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_1^*}{c} \\ \frac{\beta_2^*}{c} & \frac{\beta_2^*}{c} & \frac{-\beta_2^*}{c} & \frac{1+\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_2^*}{c} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where

$$\beta_1^* = \frac{\beta_1}{C_0}, \quad \beta_2^* = \frac{\beta_2}{C_0}, \quad c = 1 + \beta_1^* + \beta_2^*.$$

Then,

$$\begin{bmatrix} \frac{1}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{-(\beta_1^*+\beta_2^*)}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{\beta_1^*}{c} & \frac{\beta_1^*}{c} & \frac{1+\beta_2^*}{c} & \frac{-\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_1^*}{c} \\ \frac{\beta_2^*}{c} & \frac{\beta_2^*}{c} & \frac{-\beta_2^*}{c} & \frac{1+\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_2^*}{c} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_E & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} & \frac{1}{c} \\ \frac{-(\beta_1^*+\beta_2^*)}{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} & \frac{1}{c} \\ \frac{\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta_1^*}{c} & \frac{\beta_1^*}{c} \\ \frac{\beta_2^*}{c} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\beta_2^*}{c} & \frac{\beta_2^*}{c} \\ 0 & 0 & 0 & 0 & \phi_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_E & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} \frac{1}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{-(\beta_1^*+\beta_2^*)}{c} & \frac{1}{c} & \frac{-1}{c} & \frac{-1}{c} & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} \\ \frac{\beta_1^*}{c} & \frac{\beta_1^*}{c} & \frac{1+\beta_2^*}{c} & \frac{-\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_1^*}{c} \\ \frac{\beta_2^*}{c} & \frac{\beta_2^*}{c} & \frac{-\beta_2^*}{c} & \frac{1+\beta_1^*}{c} & 0 & 0 & 0 & 0 & 0 & \frac{\beta_2^*}{c} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \\ 0 \end{bmatrix}.$$

The covariance matrix Q of the errors η_t in the state equation is given by zeros except the entries in rows and columns 5 to 9, which are

$$Q[5:8,5:8] = \begin{bmatrix} \sigma^2_{\eta_1} & r_{12}\sigma_{\eta_1}\sigma_{\eta_2} & r_{13}\sigma_{\eta_1}\sigma_{\eta_3} & 0 & 0 \\ r_{12}\sigma_{\eta_1}\sigma_{\eta_2} & \sigma^2_{\eta_2} & 0 & 0 & 0 \\ r_{13}\sigma_{\eta_1}\sigma_{\eta_3} & 0 & \sigma^2_{\eta_3} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\eta_4} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\eta_5} \end{bmatrix}.$$

Appendix B. Forecast model for Southern Oscillation Index

The forecast model for monthly data on the Southern Oscillation Index (SOI) covering the period Jan-1866 to Dec-2020, obtained from Climatic Research Unit (2021) and Ropelewski and Jones (1987), is a basic structural time series model as outlined in (Durbin and Koopman, 2012, Section 3.2). The model is

$$SOI_t = \psi_t + \sum_{j=1}^{6} \gamma_{j,t} + \varepsilon_t, \tag{26}$$

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t,$$

where ψ_t is a cycle component defined below, $\eta_t \sim \mathsf{N}(0, \sigma^2_\eta)$, and $|\phi| < 1$. The monthly seasonal component is defined as

$$\gamma_{j,t} = \tilde{\gamma}_j \cos\lambda_j t + \tilde{\gamma}^*_j \sin\lambda_j t, \tag{27}$$

$$\gamma_{j,t} = -\tilde{\gamma}_j \sin\lambda_j t + \tilde{\gamma}^*_j \cos\lambda_j t, \tag{28}$$

and

$$\lambda_j = \frac{2\pi j}{12}.$$

A chi-square test for seasonality strongly rejects the null of no seasonality (29.322, with p-value 0.002). Specifying a stochastic version of the seasonal component, where equation (27) contains a random error variable, resulted in an estimated variance indistinguishable from zero, and thus the deterministic form displayed above was adopted here.

The cycle component ψ_t is defined as $\psi_t = \psi^{(2)}_t$, where

$$\begin{bmatrix} \psi^{(j)}_{t+1} \\ \psi^{*(j)}_{t+1} \end{bmatrix} = \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi^{(j)}_t \\ \psi^{*(j)}_t \end{bmatrix} + \begin{bmatrix} \psi^{(j-1)}_t \\ \psi^{*(j-1)}_t \end{bmatrix}, j = 1, 2,$$

with

$$\psi_t^{(0)} = \kappa_t \sim \mathsf{N}(0, \sigma_\kappa^2) \quad \text{and} \quad \psi_t^{*(0)} = \kappa_t^* \sim \mathsf{N}(0, \sigma_\kappa^2).$$

This is a second-order stochastic trigonometric cycle component with period $2\pi/\lambda$ (Azevedo et al., 2006). The frequency is estimated as $\hat{\lambda} = 0.129$, translating to an estimated period of 48.7 months or about 4 years.

Including a local linear form of a trend and intercept term in (26) in addition to the AR-error and the seasonal structure,

$$\mu_t = \mu_{t-1} + \nu_{t-1} + \xi_{t-1}, \quad \xi_t \sim \mathsf{N}(0, \sigma_\xi^2),$$
$$\nu_t = \nu_{t-1} + \zeta_{t-1}, \quad \zeta_t \sim \mathsf{N}(0, \sigma_\zeta^2),$$

resulted in estimated variances σ_ξ^2 and σ_ζ^2 indistinguishable from zero, and thus the deterministic form $\mu + \nu t$ was tested next. The estimation resulted in insignificant coefficients μ and ν, consistent with the cyclical nature of El-Niño and La-Niña phases, and so the specification (26) with cycle, season, and AR(1) error was chosen.

Figure 13 shows the SOI data in the upper left panel, the smoothed cycle component $\mathbb{E}\left(\hat{\psi}_t \middle| Y_T\right)$ in the upper right panel, the smoothed seasonal component $\mathbb{E}\left(\sum_{j=1}^{6} \hat{\gamma}_{j,t} \middle| Y_T\right)$ in the middle left panel, the smoothed AR(1) component $\mathbb{E}\left(\hat{\varepsilon}_t \middle| Y_T\right)$ in the middle right panel, and the smoothed residual $\mathbb{E}\left(\hat{\eta}_t \middle| Y_T\right)$ in the lower left panel. Here, Y_T is the data on the entire sample.

FIGURE 13. Fitted forecast model for SOI

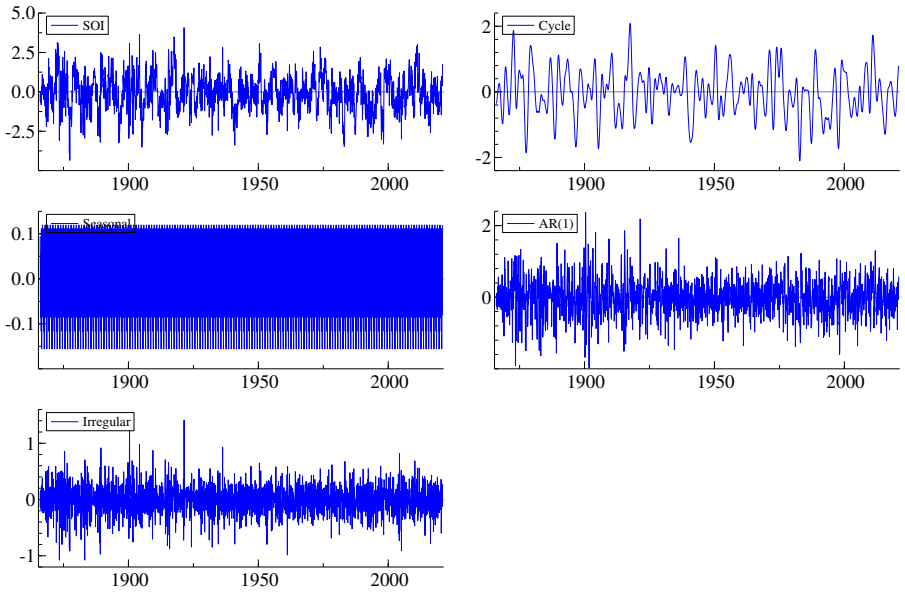
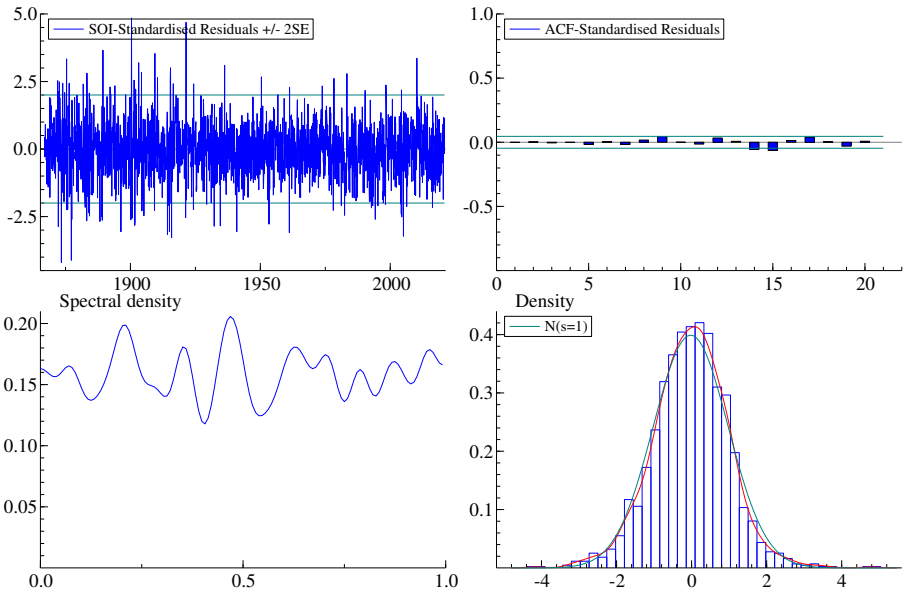


Figure 14 shows the standardized residuals from the one-step-ahead prediction in the upper left panel, the estimated sample autocorrelation function of the residuals in the upper right panel, a smoothed periodogram of the residuals in the bottom left and a histogram of the residuals in the bottom right panel.

FIGURE 14. Residuals from forecast model for SOI



The Durbin-Watson statistic of the residuals is 2.00; the R^2 of the regression is 0.26. The Ljung-Box statistics for the first 12 lags of residuals are not significant at the 95% confidence level, despite the high number of observations ($T = 1860$).

The estimated parameters of the forecast model are reported in Table 11. The forecast values for Jan-2021 through Dec-2022 are shown in Table 12. These forecast values were used to produce the annual forecasts of SOI used in the forecast of the global carbon budget system in Model II displayed in Figures 10 and Table 10. These annual forecasts are the averages of the monthly values and are 0.598 for 2021 and -0.124 for 2021.

TABLE 11. Estimated parameters for model (26).

	ϕ	λ	σ^2_κ	γ_1	γ^*_1	γ_2	γ^*_2	γ_3	γ^*_3	γ_4	γ^*_4	γ_5	γ^*_5	γ_6
Estimate	0.39	0.129	0.01	-0.03	1e-4	0.02	-0.03	-0.03	0.04	0.05	-0.02	0.07	-0.03	0.03
std.err.				0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02

FIGURE 15. Forecasts for SOI

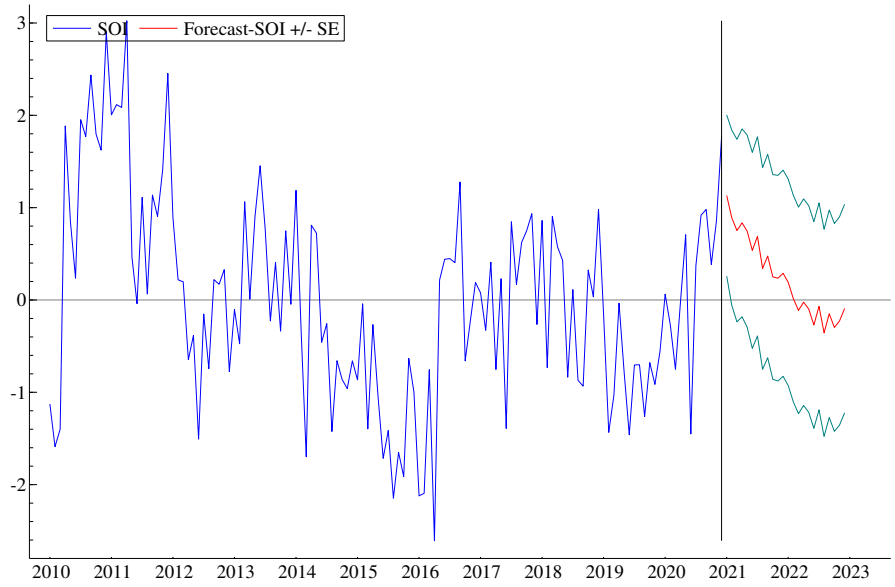


TABLE 12. Forecasts of SOI Aug-2020 through Dec-2021.

Jan 2021	1.131	Jan 2022	0.191
Feb 2021	0.890	Feb 2022	0.015
Mar 2021	0.752	Mar 2022	-0.113
Apr 2021	0.835	Apr 2022	-0.024
May 2021	0.745	May 2022	-0.097
Jun 2021	0.537	Jun 2022	-0.271
Jul 2021	0.689	Jul 2022	-0.070
Aug 2021	0.341	Aug 2022	-0.355
Sep 2021	0.476	Sep 2022	-0.149
Oct 2021	0.250	Oct 2022	-0.297
Nov 2021	0.236	Nov 2022	-0.228
Dec 2021	0.289	Dec 2022	-0.092
2021	0.598	2022	-0.124

References

Andrew, R., "Figures from the Global Carbon Budget," `https://folk.universitetetioslo.no/roberan/GCB2020.shtml` 2021. Accessed: 2021-03-03.

Angert, A., S. Biraud, C. Bonfils, W. Buermann, and I. Fung, "CO2 seasonality indicates origins of post-Pinatubo sink," *Geophysical Research Letters*, 2004, *31* (11).

Aumont, O., C. Éthé, A. Tagliabue, L. Bopp, and M. Gehlen, "PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies," *Geoscientific Model Development Discussions*, 2015, *8* (2).

Azevedo, J.V., S.J. Koopman, and A. Rua, "Tracking the Business Cycle of the Euro Area," *Journal of Business and Economic Statistics*, 2006, *24*, 278–290.

Bacastow, R. and C. D. Keeling, "Atmospheric Carbon Dioxide and radiocarbon in the natural cycle: II. Changes from A. D. 1700 to 2070 as deduced from a geochemical model," in "Carbon and the biosphere conference proceedings; Upton, New York, USA," Brookhaven Symposia in Biology, 1973, pp. 86–135.

Ballantyne, A. P., R. Andres, R. Houghton, B. D. Stocker, R. Wanninkhof, W. Anderegg, L. A. Cooper, M. DeGrandpre, P. P. Tans, J. B. Miller, C. Alden, and J. W. C. White, "Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty," *Biogeosciences*, 2015, *12* (8), 2565–2584.

Barsky, R.B. and L. Kilian, "Oil and the Macroeconomy since the 1970s," *Journal of Economic Perspectives*, 2004, *18* (4), 115–134.

Bennedsen, M., E. Hillebrand, and S.J. Koopman, "Trend analysis of the airborne fraction and sink rate of anthropogenically released CO2," *Biogeosciences*, 2019b, *16* (18), 3651–3663.

———, ———, and ———, "Modeling, Forecasting, and Nowcasting U.S. CO2 Emissions Using Many Macroeconomic Predictors," *Energy Economics*, 2021, *96*, 105118.

Berthet, S., R. Séférian, C. Bricaud, M. Chevallier, A. Voldoire, and C. Ethé, "Evaluation of an online grid-coarsening algorithm in a global eddy-admitting ocean biogeochemical model," *Journal of Advances in Modeling Earth Systems*, 2019, *11* (6), 1759–1783.

Bousquet, Philippe, Philippe Peylin, Philippe Ciais, Corinne Le Quéré, Pierre Friedlingstein, and Pieter P Tans, "Regional changes in carbon dioxide fluxes of land and oceans since 1980," *Science*, 2000, *290* (5495), 1342–1346.

Brockway, Paul E., Steve Sorrell, Gregor Semieniuk, Matthew Kuperus Heun, and Victor Court, "Energy efficiency and economy-wide rebound effects: A review of the evidence and its implications," *Renewable and Sustainable Energy Reviews*, 2021, *141*, 110781.

Bruns, Stephan B., Alessio Moneta, and David I. Stern, "Estimating the economy-wide rebound effect using empirically identified structural vector autoregressions," *Energy Economics*, 2021, p. 105158.

Buitenhuis, E.T., T. Hashioka, and C. Le Quéré, "Combined constraints on global ocean primary production using observations and models," *Global Biogeochemical Cycles*, 2013, *27* (3), 847–858.

Canadell, J.G., C. Le Quéré, M.R. Raupach, C.B. Field, E.T. Buitenhuis, P. Ciais, T.J. Conway, N.P. Gillett, R. A. Houghton, and G. Marland, "Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks," *Proceedings of the National Academy of*

Sciences, 2007, *104* (47), 18866–18870.

_____, **D. Pataki, R. Gifford, R. Houghton, Y. Luo, M. Raupach, P. Smith, and W. Steffen**, "Saturation of the Terrestrial Carbon Sink," *Terrestrial Ecosystems in a Changing World*, 01 2007, pp. 59–78.

Castruccio, Stefano and Joseph Guinness, "An evolutionary spectrum approach to incorporate large-scale geographical descriptors on global processes," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2017, *66* (2), 329–344.

_____, **Michael L Stein et al.**, "Global space–time models for climate ensembles," *The Annals of Applied Statistics*, 2013, *7* (3), 1593–1611.

Delire, Christine, Roland Séférian, Bertrand Decharme, Ramdane Alkama, Jean-Christophe Calvet, Dominique Carrer, Anne-Laure Gibelin, Emilie Joetzjer, Xavier Morel, Matthias Rocher et al., "The global land carbon cycle simulated with ISBA-CTRIP: Improvements over the last decade," *Journal of Advances in Modeling Earth Systems*, 2020, *12* (9), e2019MS001886.

Denvil-Sommer, A., M. Gehlen, M. Vrac, and C. Mejia, "LSCE-FFNN-v1: the reconstruction of surface ocean pCO2," *Geoscientific Model Development*, 2019, *12*, 2091–2105.

Dickey, D.A. and W.A. Fuller, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 1981, *49* (4), 1057–1072.

Dlugokencky, E. and P. Tans, "Trends in atmospheric carbon dioxide," Technical Report, National Oceanic and Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL) 2020.

Doney, S.C., I. Lima, R.A. Feely, D.M. Glover, K. Lindsay, N. Mahowald, J.K. Moore, and R. Wanninkhof, "Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea CO2 fluxes: Physical climate and atmospheric dust," *Deep Sea Research Part II: Topical Studies in Oceanography*, 2009, *56* (8-10), 640–655.

Doornik, J.A., "Autometrics," in Jennifer Castle and Neil Shephard, eds., *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, Oxford University Press, 2009, pp. 88–121.

Durbin, J. and S.J. Koopman, *Time series analysis by state space methods*, Oxford University Press, 2012.

Durbin, James and Siem Jan Koopman, "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 2002, *89* (3), 603–616.

Feely, R. A., R. Wanninkhof, T. Takahashi, and P. Tans, "Influence of El Niño on the equatorial Pacific contribution to atmospheric CO_2 accumulation," *Nature*, 1999, *398* (6728), 597–601.

Friedlingstein, P., M. Jones, M. O'Sullivan, R. Andrew, J. Hauck, G. Peters, W. Peters, J. Pongratz, S. Sitch, C. Le Quéré et al., "Global carbon budget 2019," *Earth System Science Data*, 2019, *11* (4), 1783–1838.

_____, **M. O'Sullivan, M. W. Jones, R. M. Andrew, J. Hauck, A. Olsen, G. P. Peters, W. Peters, J. Pongratz, S. Sitch, C. Le Quéré, J. G. Canadell, P. Ciais, R. B. Jackson, S. Alin, L. E. O. C. Arag ao, A. Arneth, V. Arora, N. R. Bates, M. Becker, A. Benoit-Cattin, H. C. Bittig, L. Bopp, S. Bultan, N. Chandra, F. Chevallier, L. P. Chini, W. Evans, L. Florentie, P. M. Forster, T. Gasser, M. Gehlen, D. Gilfillan, T. Gkritzalis, L. Gregor, N. Gruber, I. Harris, K. Hartung, V. Haverd, R. A. Houghton, T. Ilyina, A. K. Jain, E. Joetzjer, K. Kadono, E. Kato, V. Kitidis, J. I. Korsbakken, P. Landschützer,**

N. Lefèvre, A. Lenton, S. Lienert, Z. Liu, D. Lombardozzi, G. Marland, N. Metzl, D. R. Munro, J. E. M. S. Nabel, S.-I. Nakaoka, Y. Niwa, K. O'Brien, T. Ono, P. I. Palmer, D. Pierrot, B. Poulter, L. Resplandy, E. Robertson, C. Rödenbeck, J. Schwinger, R. Séférian, I. Skjelvan, A. J. P. Smith, A. J. Sutton, T. Tanhua, P. P. Tans, H. Tian, B. Tilbrook, G. van der Werf, N. Vuichard, A. P. Walker, R. Wanninkhof, A. J. Watson, D. Willis, A. J. Wiltshire, W. Yuan, X. Yue, and S. Zaehle, "Global Carbon Budget 2020," *Earth System Science Data*, 2020, *12* (4), 3269–3340.

Fuller, W.A., *Introduction to Statistical Time Series*, John Wiley & Sons, 1996.

Gasser, Thomas, Léa Crepin, Yann Quilcaille, Richard A Houghton, Philippe Ciais, and Michael Obersteiner, "Historical CO2 emissions from land use and land cover change and their uncertainty," *Biogeosciences*, 2020, *17* (15), 4075–4101.

Gifford, R.M., "Implications of CO2 effects on vegetation for the global carbon budget," in M. Heimann, ed., *The global carbon cycle*, Springer, 1993, pp. 159–199.

Gloor, M., J. L. Sarmienti, and N. Gruber, "What can be learned about carbon cycle climate feedbacks from the CO2 airborne fraction?," *Atmospheric Chemistry and Physics*, 2010, *10*, 7739 – 7751.

Gregor, Luke, Alice D Lebehot, Schalk Kok, and Pedro M Scheel Monteiro, "A comparative assessment of the uncertainties of global surface ocean CO 2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a)–have we hit the wall?," *Geoscientific Model Development*, 2019, *12* (12), 5113–5136.

Guinness, Joseph and Dorit Hammerling, "Compression and conditional emulation of climate model output," *Journal of the American Statistical Association*, 2018, *113* (521), 56–67.

Haberl, Helmut, Dominik Wiedenhofer, Doris Virág, Gerald Kalt, Barbara Plank, Paul Brockway, Tomer Fishman, Daniel Hausknost, Fridolin Krausmann, Bartholomäus Leon-Gruchalski, Andreas Mayer, Melanie Pichler, Anke Schaffartzik, Tânia Sousa, Jan Streeck, and Felix Creutzig, "A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part II: synthesizing the insights," *Environmental Research Letters*, 2020, *15* (6), 065003.

Hamilton, J.D., "Oil and the macroeconomy since World War II," *Journal of Political Economy*, 1983, *91* (2), 228–248.

———, "This is what happened to the oil price-macroeconomy relationship," *Journal of Monetary Economics*, 1996, *38* (2), 215–220.

———, "What is an oil shock?," *Journal of Econometrics*, 2003, *113* (2), 363–398.

Hansis, E., S.J. Davis, and J. Pongratz, "Relevance of methodological choices for accounting of land use change carbon fluxes," *Global Biogeochemical Cycles*, 2015, *29* (8), 1230–1246.

Harvey, Andrew C, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press, 1989.

Hauck, Judith, Moritz Zeising, Corinne Le Quéré, Nicolas Gruber, Dorothee CE Bakker, Laurent Bopp, Thi Tuyet Trang Chau, Özgür Gürses, Tatiana Ilyina, Peter Landschützer et al., "Consistency and challenges in the ocean carbon sink estimate for the Global Carbon Budget," *Frontiers in Marine Science*, 2020, *7*, 852.

Haverd, V., B. Smith, L. Nieradzik, P.R. Briggs, W. Woodgate, C.M. Trudinger, J.G. Canadell, and M. Cuntz, "A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis," *Geoscientific Model Development*, 2018, *11*, 2995–3026.

Holden, P.B., N.R. Edwards, P.H. Garthwaite, and R.D. Wilkinson, "Emulation and interpretation of high-dimensional climate model outputs," *Journal of Applied Statistics*, 2015, *42* (9), 2038–2055.

Houghton, R.A. and A.A. Nassikas, "Global and regional fluxes of carbon from land use and land cover change 1850–2015," *Global Biogeochemical Cycles*, 2017, *31* (3), 456–472.

IMF, "International Monetary Fund World Economic Outlook Update," `https://www.imf.org/en/Publications/WEO/Issues/2021/01/26/2021-world-economic-outlook-update` Jan 2021. Accessed: 2021-02-15.

———, "International Monetary Fund World Economic Outlook Update: Still Sluggish Global Growth," `https://www.imf.org/en/Publications/WEO/Issues/2019/07/18/WEOupdateJuly2019` July 2019. Accessed: 2020-10-13.

———, "International Monetary Fund World Economic Outlook: A Crisis Like No Other, An Uncertain Recovery," `https://www.imf.org/en/Publications/WEO/Issues/2020/06/24/WEOUpdateJune2020` June 2020. Accessed: 2020-10-13.

Joos, F., I.C. Prentice, S. Sitch, R. Meyer, G. Hooss, G.-K. Plattner, S. Gerber, and K. Hasselmann, "Global warming feedbacks on terrestrial carbon uptake under the Intergovernmental Panel on Climate Change (IPCC) emission scenarios," *Global Biogeochemical Cycles*, 2001, *15* (4), 891–907.

———, M. Bruno, R. Fink, U. Siegenthaler, T.F. Stocker, C. Le Quere, and J.L. Sarmiento, "An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake," *Tellus B*, 1996, *48* (3), 397–417.

Kato, E., T. Kinoshita, A. Ito, M. Kawamiya, and Y. Yamagata, "Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model," *Journal of Land Use Science*, 2013, *8* (1), 104–122.

Kaya, Yoichi and Keiichi Yokobori, eds, *Environment, energy, and economy: strategies for sustainability*, United Nations University Press Tokyo, 1997.

Kilian, L., "The economic effects of energy price shocks," *Journal of Economic Literature*, 2008, *46* (4), 871–909.

———, "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 2009, *99* (3), 1053–69.

Knorr, W., "Is the airborne fraction of anthropogenic CO2 emissions increasing?," *Geophysical Research Letters*, 2009, *36*.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," *Journal of Econometrics*, 1992, *54* (1-3), 159–178.

Landschützer, P., N. Gruber, and D.C.E. Bakker, "Decadal variations and trends of the global ocean carbon sink," *Global Biogeochemical Cycles*, 2016, *30* (10), 1396–1417.

Law, R.M., T. Ziehn, R.J. Matear, A. Lenton, M.A. Chamberlain, L.E. Stevens, W. Ying-Ping, J. Srbinovsky, D. Bi, H. Yan et al., "The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1)–Part 1: Model description and pre-industrial simulation," *Geoscientific Model Development*, 2017, *10* (7), 2567.

Lawrence, D.M., R.A. Fisher, C.D. Koven, K.W. Oleson, S.C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy et al., "The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty," *Journal of Advances in Modeling Earth Systems*, 2019, *11* (12), 4245–4287.

Lee, C.-C., "Energy consumption and GDP in developing countries: A cointegrated panel analysis," *Energy Economics*, 2005, *27* (3), 415 – 427.

Liao, Enhui, Laure Resplandy, Junjie Liu, and Kevin W Bowman, "Amplification of the ocean carbon sink during El Niños: role of poleward Ekman transport and influence on atmospheric CO2," *Global Biogeochemical Cycles*, 2020, *34* (9), e2020GB006574.

Lienert, S. and F. Joos, "A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions," *Biogeosciences*, 2018, *15* (9), 2909–2930.

Marland, G. and R.M. Rotty, "Carbon dioxide emissions from fossil fuels: a procedure for estimation and results for 1950-1982," *Tellus B: Chemical and Physical Meteorology*, 1984, *36* (4), 232–261.

Masson-Delmotte, V., P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, T. Maycock E. Lonnoy, M. Tignor, and T. Waterfield, "Global warming of 1.5 degree C. An IPCC Special Report on the impacts of global warming of 1.5 degree C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty," `https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Full_Report_High_Res.pdf` 2018. Accessed: 2021-03-01.

Mauritsen, T., J. Bader, T. Becker, J. Behrens, M. Bittner, R. Brokopf, V. Brovkin, M. Claussen, T. Crueger, M. Esch et al., "Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2," *Journal of Advances in Modeling Earth Systems*, 2019, *11* (4), 998–1038.

Meinshausen, M., S.C.B. Raper, and T.M.L. Wigley, "Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6–Part 1: Model description and calibration," *Atmospheric Chemistry and Physics*, 2011, *11* (4), 1417–1456.

Meiyappan, P., A.K. Jain, and J.I. House, "Increased influence of nitrogen limitation on CO2 emissions from future land use and land use change," *Global Biogeochemical Cycles*, 2015, *29* (9), 1524–1548.

Melton, Joe R, Vivek K Arora, Eduard Wisernig-Cojoc, Christian Seiler, Matthew Fortier, Ed Chan, and Lina Teckentrup, "CLASSIC v1. 0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM)–Part 1: Model framework and site-level performance," *Geoscientific Model Development*, 2020, *13* (6), 2825–2850.

of East Anglia Climatic Research Unit, University, "Southern Oscillation Index," `https://crudata.uea.ac.`

uk/cru/data/soi/ 2021. Accessed: 2020-02-15.

Oh, W. and K. Lee, "Causal relationship between energy consumption and GDP revisited: the case of Korea 1970–1999," *Energy Economics*, 2004, *26* (1), 51 – 59.

Ozturk, I., "A literature survey on energy–growth nexus," *Energy Policy*, 2010, *38* (1), 340 – 349.

Paulsen, H., T. Ilyina, K.D. Six, and I. Stemmler, "Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC," *Journal of Advances in Modeling Earth Systems*, 2017, *9* (1), 438–464.

Perron, P., "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 1989, pp. 1361–1401.

Peters, Glen P., Corinne Le Quéré, Robbie M. Andrew, Josep G. Canadell, Pierre Friedlingstein, Tatiana Ilyina, Robert B. Jackson, Fortunat Joos, Jan Ivar Korsbakken, Galen A. McKinley, Stephen Sitch, and Pieter Tans, "Towards real-time verification of CO2 emissions," *Nature Climate Change*, 2017, *7* (12), 848–850.

Phillips, P.C.B. and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, 1988, *75* (2), 335–346.

Piketty, T., *Capital in the 21st Century*, Belknap/Harvard, 2014.

Poulter, B., D.C. Frank, E.L. Hodson, and N.E. Zimmermann, "Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO2 airborne fraction," *Biogeosciences*, 2011, *8*, 2027–2026.

Pretis, F., J. Reade, and G. Sucarrat, "Automated General-to-Specific (GETS) regression modeling and indicator saturation methods for the detection of outliers and structural breaks," *Journal of Statistical Software*, 2018, *86* (3).

Quéré, C. Le, M.R. Raupach, J.G. Canadell, G. Marland, L. Bopp, P. Ciais, T.J. Conway, S.C. Doney, R.A. Feely, P. Foster, P. Friedlingstein, K. Gurney, R.A. Houghton, J.I. House, C. Huntingford, P.E. Levy, M.R. Lomas, J. Majkut, N. Metzl, J.P. Ometto, G.P. Peters, I.C. Prentice, J.T. Randerson, S.W. Running, J.L. Sarmiento, U. Schuster, S. Sitch, T. Takahashi, N. Viovy, G.R. van der Werf, and F.I. Woodward, "Trends in the sources and sinks of carbon dioxide," *Nature Geoscience*, 2009, *2*, 831 – 836.

_____**, R.M. Andrew, P. Friedlingstein, S. Sitch, J. Hauck, J. Pongratz, P. Pickers, J.I. Korsbakken, G.P. Peters, J.G. Canadell et al.**, "Global carbon budget 2018," *Earth System Science Data*, 2018, *10*, 2141–2194.

_____, _____, _____, _____**, J. Pongratz, A.C. Manning, J.I. Korsbakken, G.P. Peters, J.G. Canadell, R.B. Jackson et al.**, "Global carbon budget 2017," *Earth System Science Data Discussions*, 2017, pp. 1–79.

Raupach, M. R., "The exponential eigenmodes of the carbon-climate sytem, and their implications for ratios of responses to forcings," *Earth System Dynamics*, 2013, *4*, 31 – 49.

_____**, J. G. Canadell, and C. Le Quéré**, "Anthropogenic and biophysical contributions to increasing atmospheric CO2 growth rate and airborne fraction," *Biogeosciences*, 2008, *5*, 1601 – 1613.

_____**, M. Gloor, J. L. Sarmiento, J. G. Canadell, T. L. Frölicher, T. Gasser, R. A. Houghton,**

C. Le Quéré, and C. M. Trudinger, "The declining uptake rate of atmospheric CO_2 by land and ocean sinks," *Biogeosciences*, 2014, *11* (13), 3453–3475.

Raupach, M.R., G. Marland, P. Ciais, C. Le Quéré, J.G. Canadell, G. Klepper, and C.B. Field, "Global and regional drivers of accelerating CO2 emissions," *Proceedings of the National Academy of Sciences*, 2007, *104* (24), 10288–10293.

Rayner, P. J., A. Stavert, M. Scholze, A. Ahlström, C. E. Allison, and R. M. Law, "Recent changes in the global and regional carbon cycle: analysis of first-order diagnostics," *Biogeosciences*, 2015, *12* (3), 835–844.

Rödenbeck, C., D.C.E. Bakker, N. Metzl, A. Olsen, C. Sabine, N. Cassar, F. Reum, R.F. Keeling, and M. Heimann, "Interannual sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme," *Biogeosciences*, 2014, *11*, 4599–4613.

Ropelewski, C.F. and P.D. Jones, "An extension of the Tahiti-Darwin Southern Oscillation Index," *Monthly Weather Review*, 1987, *115*, 2161–2165.

Schwinger, J., N. Goris, J.F. Tjiputra, I. Kriest, M. Bentsen, I. Bethke, M. Ilicak, K.M. Assmann, and C. Heinze, "Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1)," *Geoscientific Model Development*, 2016, *9*, 2589–2622.

Sellar, A.A., C.G. Jones, J.P. Mulcahy, Y. Tang, A. Yool, A. Wiltshire, F.M. O'connor, M. Stringer, R. Hill, J. Palmieri et al., "UKESM1: Description and evaluation of the UK Earth System Model," *Journal of Advances in Modeling Earth Systems*, 2019, *11* (12), 4513–4558.

Shephard, N., "Distribution of the ML Estimator of an MA(1) and a Local Level Model," *Econometric Theory*, 1993, *9*, 377–401.

Smith, B., D. Warlind, A. Arneth, T. Hickler, P. Leadley, J. Siltberg, and S. Zaehle, "Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model," *Biogeosciences*, 2014, *11*, 2027–2054.

Stern, D.I., "Energy and economic growth in the USA: A multivariate approach," *Energy Economics*, 1993, *15* (2), 137 – 150.

———, "A multivariate cointegration analysis of the role of energy in the US macroeconomy," *Energy Economics*, 2000, *22* (2), 267 – 283.

——— **and A. Kander**, "The role of energy in the industrial revolution and modern economic growth," *The Energy Journal*, 2012, *33* (3).

Stock, J. H. and M. W. Watson, "Evidence on Structural Instability in Macroeconomic Time Series Relation," *Journal of Business and Economic Statistics*, 1996, *14* (1), 11–30.

Tian, H., G. Chen, C. Lu, X. Xu, D.J. Hayes, W. Ren, S. Pan, D.N. Huntzinger, and S.C. Wofsy, "North American terrestrial CO2 uptake largely offset by CH4 and N2O emissions: toward a full accounting of the greenhouse gas budget," *Climatic Change*, 2015, *129* (3-4), 413–426.

UNFCCC, "United Nations Framework Convention on Climate Change: National Determined Contributions Synthesis Report," `https://unfccc.int/process-and-meetings/the-paris-agreement/nationally-determined-contributions-ndcs/nationally-determined-contributions-ndcs/`

ndc-synthesis-report#eq-5 2021. Accessed: 2021-03-01.

Vuichard, Nicolas, Palmira Messina, Sebastiaan Luyssaert, Bertrand Guenet, Sönke Zaehle, Josefine Ghattas, Vladislav Bastrikov, and Philippe Peylin, "Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production," *Geoscientific Model Development*, 2019, *12* (11), 4751–4779.

Walker, A.P., T. Quaife, P.M. van Bodegom, M.G. De Kauwe, T.F. Keenan, J. Joiner, M.R. Lomas, N. MacBean, C. Xu, X. Yang et al., "The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production," *New Phytologist*, 2017, *215* (4), 1370–1386.

Watson, Andrew J, Ute Schuster, Jamie D Shutler, Thomas Holding, Ian GC Ashton, Peter Landschützer, David K Woolf, and Lonneke Goddijn-Murphy, "Revised estimates of ocean-atmosphere CO 2 flux are consistent with ocean carbon inventory," *Nature communications*, 2020, *11* (1), 1–6.

World Bank, "World GDP (constant 2010 US$)," https://data.worldbank.org/indicator/NY.GDP.MKTP.KD 2021. Accessed: 2021-02-15.

_____, "Global Economic Prospects," https://openknowledge.worldbank.org/handle/10986/34710 Jan 2021. Accessed: 2021-02-15.

_____, "Global Economic Prospects," https://openknowledge.worldbank.org/handle/10986/31655 June 2019. Accessed: 2020-10-13.

_____, "Global Economic Prospects," https://openknowledge.worldbank.org/handle/10986/33748 June 2020. Accessed: 2020-10-13.

Yuan, Wenping, Dan Liu, Wenjie Dong, Shuguang Liu, Guangsheng Zhou, Guirui Yu, Tianbao Zhao, Jinming Feng, Zhuguo Ma, Jiquan Chen, Yang Chen, Shiping Chen, Shijie Han, Jianping Huang, Linghao Li, Huizhi Liu, Shaoming Liu, Mingguo Ma, Yanfeng Wang, Jiangzhou Xia, Wenfang Xu, Qiang Zhang, Xinquang Zhao, and Liang Zhao, "Multiyear precipitation reduction strongly decreases carbon uptake over northern China," *Journal of Geophysical Research: Biogeosciences*, 2014, *119* (5), 881–896.

Yue, X and N Unger, "The Yale Interactive terrestrial Biosphere model version 1.0: description, evaluation and implementation into NASA GISS ModelE2," *Geoscientific Model Development*, 2015, *8* (8), 2399–2417.

Zaehle, S. and A.D. Friend, "Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates," *Global Biogeochemical Cycles*, 2010, *24* (1).

Zhang, X.-P. and X.-M. Cheng, "Energy consumption, carbon emissions, and economic growth in China," *Ecological Economics*, 2009, *68* (10), 2706 – 2712.

(M. Bennedsen) Center for Research in Econometric Analysis of Time Series (CREATES), Department of Economics and Business Economics, Aarhus University, Fuglesangsallé 4, 8210 Aarhus V, Denmark.

Email address: mbennedsen@econ.au.dk

(E. Hillebrand) CREATES, Department of Economics and Business Economics, Aarhus University, Fuglesangsallé 4, 8210 Aarhus V, Denmark.

Email address: ehillebrand@creates.au.dk

(S.J. Koopman) Vrije Universiteit Amsterdam, Department of Econometrics, and CREATES, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.

Email address: s.j.koopman@vu.nl

Research Papers 2020

2020-03: Daniel Borup, Bent Jesper Christensen, Nicolaj N. Mühlbach and Mikkel S. Nielsen: Targeting predictors in random forest regression

2020-04: Nicolaj N. Mühlbach: Tree-based Synthetic Control Methods: Consequences of moving the US Embassy

2020-05: Juan Carlos Parra-Alvarez, Olaf Posch and Mu-Chun Wang: Estimation of heterogeneous agent models: A likelihood approach

2020-06: James G. MacKinnon, Morten Ørregaard Nielsen and Matthew D. Webb: Wild Bootstrap and Asymptotic Inference with Multiway Clustering

2020-07: Javier Hualde and Morten Ørregaard Nielsen: Truncated sum of squares estimation of fractional time series models with deterministic trends

2020-08: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Adaptive Inference in Heteroskedastic Fractional Time Series Models

2020-09: Daniel Borup, Jonas N. Eriksen, Mads M. Kjær and Martin Thyrsgaard: Predicting bond return predictability

2020-10: Alfonso A. Irarrazabal, Lin Ma and Juan Carlos Parra-Alvarez: Optimal Asset Allocation for Commodity Sovereign Wealth Funds

2020-11: Bent Jesper Christensen, Juan Carlos Parra-Alvarez and Rafael Serrano: Optimal control of investment, premium and deductible for a non-life insurance company

2020-12: Anine E. Bolko, Kim Christensen, Mikko S. Pakkanen and Bezirgen Veliyev: Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures

2020-13: Morten Ørregaard Nielsen and Antoine L. Noël: To infinity and beyond: Efficient computation of ARCH(∞) models

2020-14: Charlotte Christiansen, Ran Xing and Yue Xu: Origins of Mutual Fund Skill: Market versus Accounting Based Asset Pricing Anomalies

2020-15: Carlos Vladimir Rodríguez-Caballero and J. Eduardo Vera-Valdés: Air pollution and mobility in the Mexico City Metropolitan Area, what drives the COVID-19 death toll?

2020-16: J. Eduardo Vera-Valdés: Temperature Anomalies, Long Memory, and Aggregation

2020-17: Jesús-Adrián Álvarez, Malene Kallestrup-Lamb and Søren Kjærgaard: Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans

2020-18: Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman: A statistical model of the global carbon budget