Econometric Game 2019, Team 17: Modeling the dynamics of the carbon cycle through the global carbon budget equation

Abstract

In this paper, we model the main data objects which jointly define the Global Carbon Budget equation. Investigating their time series properties, we find evidence of unit roots and cointegration relationships and demonstrate that a vector-error-correction model (VECM) is able to capture this structure. As a benchmark, we use a state-space model with a stochastic linear trend, and in order to form forecast combinations, we consider a set of simple univariate models. We construct forecasts for the atmospheric CO_2 concentration until the year 2100 by the VECM, the state-space model, and the combination of all models. We find evidence that the pathway of atmospheric CO_2 concentration lies between the two RCP4.5 and RCP6.0 scenarios defined by the Representative Concentration Pathways initiative. Using our model, we synthesize counterfactual scenarios of CO_2 emissions given different pathways for fossil-fuel emission. To reach the RCP2.6, our model suggests that we need to reduce fossil-fuel emissions by 85% annually. Alternatively, global implementation of the European Union's goal to reach zero fossil-fuel emission by 2050 would allow us to meet the RCP2.6 scenario in 2100.

This version: April 11, 2019

I. Introduction

The stock of CO_2 in the global climate system is managed by the carbon cycle. CO_2 is exchanged between biosphere, hydrosphere, pedosphere, and atmosphere. As CO_2 is a key driver of climate change, it is of crucial importance to investigate the mediation of CO_2 between these different earth systems to achieve a precise picture of climate change dynamics. The aim of this paper is therefore to first set up a statistical model of the global carbon cycle and anthropogenic CO_2 emission that reproduces key properties of the global CO_2 time series. Based on the model estimates, we then forecast global carbon concentration paths until the year 2100 and investigate in counterfactual exercises by how much fossil fuel CO_2 emissions need to be reduced in order to comply with specific goals of CO_2 concentration in the future.

The key modeling equation of this paper is the so-called carbon budget. It is defined as

$$G_t^{AMT} = E_t^{ANT} - S_t^{LND} - S_t^{OCN} \tag{1}$$

where E_t^{ANT} is anthropogenically released CO_2 into the atmosphere, G_t^{AMT} is the growth of atmospheric CO_2 emission, S_t^{LND} is the exchange flow of CO_2 from atmosphere and hydrosphere (the land sink) and S_t^{OCN} is the flow of CO_2 from atmosphere to biosphere (the ocean sink). The man-made CO_2 emission can be written as $E_t^{ANT} = E_t^{LUC} + E_t^{FF}$, where E_t^{LUC} and E_t^{FF} denote anthropogenic CO_2 emissions from land changes (such as deforestation) and fossil fuel, respectively. In summary, the carbon budget equation describes the flow of CO_2 between different earth systems, and importantly, the CO_2 that is not observed absorbed by either land or sea must go into the atmosphere. The change in the CO_2 concentration in the atmosphere, G_t, is the man-induced CO_2 emission and corrected for the exchange of CO_2 between atmosphere and land as well as atmosphere and ocean.

The rest of the paper is organized as follows. Section II discusses the science of climate change and the carbon cycle. Section III presents the data and investigates the different time series, whereas Section IV introduces the econometric framework, namely the state-space model and the vector-error-correction model (VCM). In Section V, we forecast the atmospheric CO_2 concentration until 2100 and relate it to the Intergovernmental Panel on Climate Change's (IPCC) greenhouse

gas emission trajectory scenarios. Using our model, we conduct a counterfactual analysis by asking if and by how much fossil fuel emissions have to be reduced to reach certain CO_2 concentration goals. Section VI concludes.

II. Carbon Dioxide, Climate Change and the Carbon Cycle

Much of the focus in both the academic and political debates about climate change has evolved around greenhouse gases like methan or carbon dioxide. Greenhouse gases are a class of chemical objects that are associated with the greenhouse effect of climate change: anthropogenically emissions of such gases increase their concentration level in the atmosphere and disrupt the earth's energy balance by reflecting some of the energy back to the earth's surface that otherwise would be emitted into space. Among greenhouse gases, carbon dioxide (CO_2) is notoriously famous because it is particularly reflective. The reason is that carbon dioxide's eigenmodes - the frequency at which carbon dioxide molecules start to resonate, thus to take up and send back energy - is approximately the same as the frequency with which the earth surface reflects energy back to the atmosphere. With increasing levels of CO_2 concentration, more energy will be reflected back to the earth surface, and vice versa, leading to a feedback mechanism: climate change.

The overall mediation process of carbon dioxide throughout the geosphere is known as the carbon cycle. Carbon dioxide is not only exchanged between atmosphere and biosphere (e.g. through the CO_2 consuming photosynthesis of plants), but is also exchanged between ocean and atmosphere (as e.g. through CO_2 producing plancton). The overall transmission of carbon dioxide between the different geospheres can be modeled through the global carbon budget.

Taking the global carbon cycle into account is crucial for the statistical modeling of the dynamics of atmospheric CO_2 concentration, as not all of anthropogenic carbon dioxide emissions will deposit in the atmosphere. As Bennedsen, Hillebrand, and Koopman (2018) note, only 45% of human-induced CO_2 emission reaches the atmosphere, while 24% flows to the oceans, and 31% to the land. For the analysis of carbon dioxide in the atmosphere and for climate change more general, this means that not all anthropogenic carbon dioxide emissions will directly contribute to climate change (through an atmospheric greenhouse gas channel), but might

do so indirectly by disturbing a global carbon cycle that has been in a dynamic equilibrium for the last 800.000 years [see Hillebrand (2019)].

III. Data and Descriptives

The following section summaries the dataset of our analysis and the Representative Concentration Pathways. Then, we analyzes whether the data objects show signs of stationarity and whether the time series are integrated. Lastly, we test for cointegration.

A. Data & the RCP carbon dioxide projections

The dataset contains four time series of carbon dioxide flow variables - the carbon dioxide emission from fossil fuels, E^{FF}, the carbon dioxide emission from changes in land-use, E^{LUC}, the carbon dioxide sink rate of the ocean, S^{OCN}, and the land sink rate S^{LND}. Anthropogenic carbon dioxide emission from fossil fuels (E^{FF}) and from land-use changes (E^{LUC}) are estimates based on actual data and described in [Boden, Marland, and Andres (2009) & RA Houghton (2017)]. The time series on the sink rates stem from averages of several global ocean biogeochemistry and global vegetation models [see Quere, Friedlingstein, Sitch, Hauck, Pongratz, Pickers, Korsbakken, Peters, Canadell, and Arora (2018)]. Lastly, the time series of the atmospheric carbon dioxide growth rate consists of actual measurements following Dlugokencky and Tans (2018). All five time series cover the time period 1959-2017.

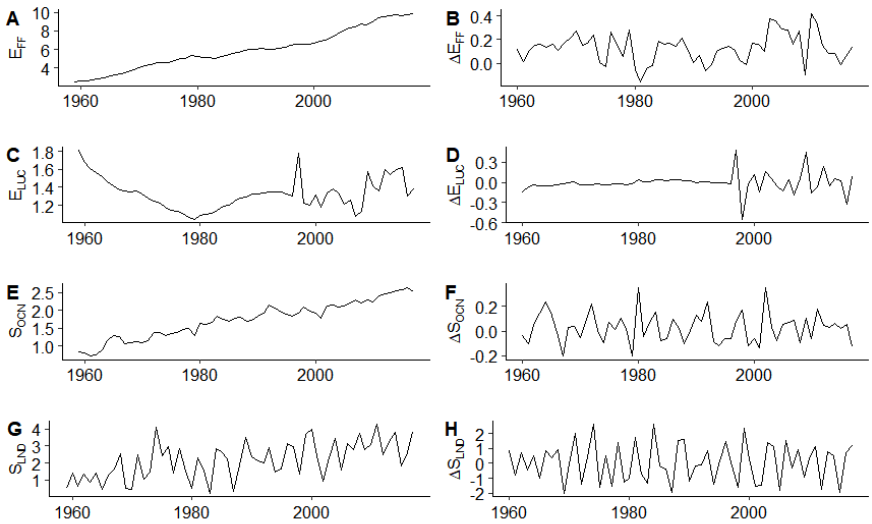
The Representative Concentration Pathways (short RCP's) are a set of four scenario-based projections of global greenhouse gas emission paths (see e.g. Meinshausen, Smith, Calvin, Daniel, Kainuma, Lamarque, Matsumoto, Montzka, Raper, Riahi et al. (2011a)). They are gathered and assessed by climate scientists and published by the Intergovernmental Panel on Climate Change (IPCC), the United Nation's main climate change institution. The four scenarios, RCP2.6, RCP4.5, RCP6 and RCP8.5 correspond to different dynamics and net emissions of greenhouse gas, and they are broadly consistent with several future paths of anthropogenic global greenhouse emission trajectories. It is worth noting that the RCP's focus is on greenhouse gas emission, while the focus of our study is on

carbon dioxide.

B. Descriptives: Testing for Unit Roots and Cointegration

In Figure 1, we have plotted the time series of E^{FF}, E^{LUC}, S^{OCN} and S^{LND} and their first differences. We see that the emission of fossil fuel and the ocean sink seems to be increasing almost linearly for the considered period, whereas the emission from land changes such as deforestation and the flow kept by the landmass appear to be more volatile. Looking at the plots of the first differences, we notice that the series for the emission due to land changes is remarkably stable for the first large part of the sample. However, at the beginning of the new millennium we see a large increase in the volatility of the series.

Figure 1: Time Series in levels and First Differences



This figure shows the time series in levels (left) and first differences (right) for the four considered time series: emissions from fossil fuel, emissions from land changes, the amount of emission collected by the landmass and the amount of emission collected by the ocean.

The left column of Figure 1 indicate some sort of unit root behaviour in the four time series. In order to test this formally, we perform an Augmented Dickey-Fuller (ADF) test (for more details on the test see Fuller (2009)). The ADF test test the null hypothesis that there is a unit root, and it relies on the following regression equation

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p y_{t-p} + \varepsilon_t, \tag{2}$$

and testing for a unit root is equivalently to testing $\gamma = 0$. We consider various versions of the test based on different model specification and provide the summary statistic of the tests in the Appendix. However, we cannot reject the null hypothesis of a unit root. This is robust across all specifications (excluding a constant and a trend and different amount of lags) except for S_t^{OCN} and S_t^{LND} which could potentially be trend stationary at the 5% confidence level. We additionally test for the same relationship using the MacKinnon (1996) test statistic and a Perron (1988) test and obtain the same conclusion. Thus, the results suggests that the processes all contain a stochastic trend. To further test for a common stochastic trend we use the Johansen test for cointegration (see Johansen (1988)). We choose a model where the cointegration relationship includes a constant and test for different numbers of cointegration relationships using the trace test and the long run VECM formulation. The null hypothesis of the model is that that the series contain $r < k$ cointegration relationships versus the alternative that $r = k$ and so the test is done sequentially for $r = 0, 1, 2, 3$ (as we have 4 time series). We consider different lag length for robustness. We cannot reject the null hypothesis of no cointegration relationship for lag length 5 and 7, but reject the same null for all other lag length from 2-10. Depending on the exact number of lags used in the model we get somewhere between 1 and 3 different cointegration relationships, but it seems evident that some relationship exists.

IV. Econometric Models

In this section, we introduce three statistical models describing the global carbon budget and its dynamics. In order to make the global carbon budget equation operational in a statistical model, we need to add a stochastic error term that accounts for the fact that the accounting identity formulated in the carbon equation might not always hold when faced with real data. Therefore, we will from now on refer to the "stochastic" carbon budget equation:

$$G_t^{AMT} = E_t^{ANT} - S_t^{LND} - S_t^{OCN} + e_t, \tag{3}$$

where e_t is an idiosyncratic and weakly stationary error term and all variables inherit their names from the identity equation (1).

In the following section, we present three distinct econometric models of the time series behavior of the elements of the carbon budget equation. First, as all four elements of the carbon budget equation seem to follow a stochastic trend, we model the time series individually in a state space model with linear stochastic trends. In order to account for the observed cointegration of the time series, we then specify a Vector-Error-Correction Model. In a last step, we introduce two simple models of the ARIMA(I, p, q) family and a non-linear AAR(p) model. We then discuss the benefits of model averaging in the context of forecasting exercises with few observations.

A. A State Space Model

Figure 1 suggest an underlying trend in the various time-series of data objects. Thus, we will pursue the route of Bennedsen et al. (2018) and assume that the various time series follow a linear trend model. To that end, for a univariate time series y_t, we assume that y_t follows

$$y_t = T_t + \epsilon_t, \tag{4}$$

where T_t denotes the stochastic trend given by

$$T_{t+1} = T_t + \beta + \eta_t, \tag{5}$$

for some unknown constant coefficient $\beta \in \mathbb{R}$. Here, ϵ_t can be thought of measurements errors or deviations from the trend due to some unexpected weather circumstances. As shown by Bennedsen et al. (2018), this specification allows to interpret T_t as stochastic process that consists of a linear trend with slope β and a random walk component. The advantage of this approach is that it allow us to cast the system in a simple state-space model given by:

$$y_t = Ax_t + \epsilon_t, \tag{6}$$

$$x_{t+1} = Bx_t + \kappa_t, \tag{7}$$

where $x_t = [T_t, \beta]$ is the time-t state vector and:

$$A = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{8}$$

We assume the error structure:

$$\epsilon_t \sim N(0, \sigma_\epsilon^2), \quad \kappa_t \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & 0 \end{pmatrix} \right], \tag{9}$$

and that ϵ_t and κ_t are mutually independent. We estimate this model by Kalman filtering in conjunction with maximum likelihood. The advantage of this approach is that it can handle non-stationarity of the data and it allows for noise in the measurements.

For the Kalman filtering, denote by $(y_1, \ldots, y_T)$ the observed data and write $Y_t = (y_1, \ldots, y_T)$ for observations up to time t. Further, denote the time-t state vector estimate by $\mu_{t|t} = \mathbb{E}(x_t|Y_t)$ and the state vector prediction by $\mu_{t+1|t} = \mathbb{E}(x_{t+1}|Y_t)$. The corresponding state vector variance-covariance matrices are $\Sigma_{t|t} = \text{var}(x_t|Y_t)$ and $\Sigma_{t+1|t} = \text{var}(x_{t+1}|Y_t)$. During the estimation we use the low storage algorithm of Koopman, Shephard, and Doornik (1999). For this, we need an initial condition for the first factor vector and variance-covariance matrix given by $x_{1|0}$ and $\Sigma_{1|0}$, which we treat as unknown coefficients. The innovation in observing y_t is $\zeta_t = y_t - Ax_t$ with variance $\Gamma_t = A\Sigma_{t|t-1}A' + \sigma_\epsilon^2$. Then by the Kalman filter the update step is

$$\begin{aligned} \mu_{t|t} &= \mu_{t|t-1} + \Sigma_{t|t-1}A'\Gamma_t^{-1}\zeta_t, \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}A'\Gamma_t^{-1}A\Sigma_{t|t-1}, \end{aligned}$$

and the prediction step is

$$\mu_{t+1|t} = B\mu_{t|t} = B\mu_{t|t-1} + K_t\zeta_t, \tag{10}$$

$$\Sigma_{t+1|t} = B\Sigma_{t|t}\Phi_1' + \Omega = B\Sigma_{t|t-1}\Phi_1 + \Omega - K_t\Gamma_t K_t'. \tag{11}$$

Here the second equalities in both lines substitute the update step and sets $K_t = A\Sigma_{t|t-1}A'\Gamma_t^{-1}$. The contribution to the log-likelihood of each new observation

is

$$\log p\left(y_t|Y_{t-1}\right) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\log|\Gamma_t| - \frac{1}{2}\zeta_t'\Gamma_t^{-1}\zeta_t,$$

and the prediction-error decomposition of the log-likelihood function is therefore

$$\log L = \sum_{t=1}^{T}\log p\left(y_t|Y_{t-1}\right) = -\frac{mT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|\Gamma_t| + \zeta_t'\Gamma_t^{-1}\zeta_t\right), \tag{12}$$

where m denotes the first dimension of y_t. To calculate the log-likelihood function we only need the series of innovations ζ_t and Γ_t and these are calculated from $\mu_{t|t-1}$ and $\Sigma_{t|t-1}$ alone. The low storage algorithm only stores the series of ζ_t, Γ_t, and K_t and calculates $\mu_{t+1|t}$ and $\Sigma_{t+1|t}$ by (10) and (11) with $\mu_{t|t-1}$ and $\Sigma_{t|t-1}$ stored only from last period. Since this method minimizes the need for storage and avoids calculating the update step, it speeds up the filter which must be calculated many times in the optimization of (12) over parameters.

Under quadratic loss the optimal forecast is the relevant conditional expectation. The optimal forecast for y_T made at time T for time $T+h$ is therefore

$$\hat{y}_{T+h|T} = \mathbb{E}\left[\hat{y}_{T+h}|\hat{y}_{T+h-1|T}, \hat{y}_{T+h-2|T}, \ldots\right] = \mathbb{E}\left[\hat{x}_{T+h}|\hat{y}_{T+h-1|T}, \hat{y}_{T+h-2|T}, \ldots\right]. \tag{13}$$

Ignoring the constant beta coefficient in the state vector we have from the transition dynamics that

$$\mathbb{E}\left[\hat{x}_{T+h}|\hat{y}_{T+h-1|T}, \hat{y}_{T+h-2|T}, \ldots\right] = T_T + \hat{\beta}h, \tag{14}$$

where we have used that the random walk component has mean zero. The forecasting of the state vector (14) translates into forecasting of y_T via (13).

B. Vector Error Correction Model

Let Y_t be our vector of climate variables defined as in the state space formulation above. We then start with a classical vector autoregressive model (VAR)

$$\Theta(L)Y_t = \delta + \epsilon_t, t \in 1, ..., n, \tag{15}$$

where $\Theta(L)$ is a $k \times k$ matrix lag polynomial of degree p

$$\Theta(L) = I_k - \Theta_1 L - \cdot - \Theta_p L^p, \tag{16}$$

δ is a $p \times 1$ vector of intercepts and ϵ_t is a $p \times 1$ vector of weakly stationary and serially uncorrelated white noise error terms. As noted above, there seems to be evidence that the different climate variables share a common stochastic trend or, in other words, that Y_t is cointegrated. Because of this, we extend (15) to a Vector Error Correction Model (VECM) of the form:

$$\Delta Y_t = \mu + \Pi Y_{t-p} + \Gamma_{p-1} \Delta Y_{t-p+1} + \cdot + \Gamma_1 \Delta Y_{t-1} + \varepsilon_t \tag{17}$$

where μ is now our $p \times 1$ vector of intercepts, Π is the $p \times p$ error correction part and Γ_i is the $p \times p$ Vector Autoregressive parts. Choosing the model specification amounts to choosing the number of lags p and the number of cointegration relationships. The latter can be chosen by performing the Johansen test (testing the rank of Π) sequentially.

C. Model Averaging

Empirical studies often find that forecast combinations dominate the ex ante best individual forecasting models in terms of forecasting accuracy. There are several reasons underlying this phenomena (Timmermann, 2006). First, individual forecasting models are most likely - at least to a part - misspecified, and model combinations are less prone to misspecification. Second, instability may hamper identification of a single best model, especially under a state-dependent performance and in an unstable forecasting environment for which past records are unreliable. For time-series data with low frequency, this is particularly pressing. For those reasons, we consider three additional models to our state-space model and VECM. In particular, the models we consider amount to an autoregressive model $AR(p)$, an autoregressive integrated moving average model $ARIMA(p, I, q)$, and an additive nonlinear autoregressive model $AAR(p)$. We briefly state the model equations. The well-known $AR(p)$ is given by

$$\theta(L)\, y_t = \varepsilon_t, \tag{18}$$

where $\theta(L)$ is a polynomial of order p in the lag operater L. As we fail to reject at least one unit root for each of the four series, we estimate the model with first differences. The optimal lag length p is chosen by AIC. An extension that includes a moving average part is the $ARIMA(p,q)$ model that reads

$$\theta(L)\, y_t = \alpha(L)\, \varepsilon_t, \tag{19}$$

where $\theta(L)$ is as before and $\alpha(L)$ is the counterpart of lag order q. We choose lag lengths p and q to minimize the AIC. The last model we consider is the $AAR(p)$ given by

$$y_{t+1} = \theta_0 + \sum_{j=1}^{p} s_j \left(y_{t-(j-1)} \right), \tag{20}$$

where $s_j()$ are non-parametric univariate functions of lagged time series values represented by cubic regression splines. All models produce forecasts as recursive one-step ahead forecasts according to

$$\hat{y}_{T+h|T} = \mathbb{E}\left[\hat{y}_{T+h} | \hat{y}_{T+h-1|T}, \hat{y}_{T+h-2|T}, \ldots \right]. \tag{21}$$

To combine the forecasts, we use an equal-weighted average. Several authors have suggested to use other weighting schemes, for instance the Bates-Granger optimal weights (Bates and Granger, 1969). However, we do not press this further.

V. Results

In this section, we examine the empirical performance of the different models under consideration. First, we present the in-sample fit of our models. But as choosing a model based on in-sample fit is often erroneous, we study out-of-sample performance between 2001-2017, in which periods we know the true time series. Second, we produce forecasts from 2018-2100 for the VECM, the state-space model, and the forecast combinations. Last, we conduct a counterfactual exercise by asking how fossil fuel CO_2 emissions need to be reduced to coincide with certain CO_2 concentration trajectories.

We start our analysis with the VECM where we consider $p = 1,2,3,4,5$ lags and

use the number of cointegration relationships identified by the Johansen test. We combute AIC and BIC for each model specification and show the results in Table 1.

Table 1: Information criteria.

p	r	AIC	BIC
1	3	-724.9592	-661.6246
2	2	-693.8331	-604.7176
3	1	-658.0549	-547.6515
4	2	-637.2530	-486.0902
5	1	-602.4980	-431.0826

This table displays the AIC and BIC for various specifications of the VECM.

From Table 1, a VECM with one lag and three cointegration relationships dominates on all specifications on both AIC and BIC. Thus, we declare this particular VECM as our preferred specification and continue to estimate it. The results are shown in Table 2.

Table 2: Parameter estimates of our preferred VEC model.

Variable	$\widehat{\beta}$	t-stat($\widehat{\beta}$)
E_{FF}:ECT_1	0.0194	0.6357
E_{FF}:ECT_2	0.1205	1.2661
E_{FF}:ECT_3	-0.0895	-0.7755
E_{FF}:$L.E_{FF}$	0.4810	3.2368
E_{FF}:$L.E_{LUC}$	0.0846	0.6570
E_{FF}:$L.S_{OCN}$	0.1790	1.1715
E_{FF}:$L.S_{LND}$	-0.0022	-0.1127
E_{LUC}:ECT_1	0.0447	1.4748
E_{LUC}:ECT_2	-0.2830	-2.9898
E_{LUC}:ECT_3	-0.1476	-1.2860
E_{LUC}:$L.E_{FF}$	0.0400	0.2704
E_{LUC}:$L.E_{LUC}$	-0.2288	-1.7869
E_{LUC}:$L.S_{OCN}$	0.0343	0.2257
E_{LUC}:$L.S_{LND}$	0.0204	1.0389
S_{OCN} :ECT_1	0.0554	2.0452
S_{OCN}:ECT_2	-0.0911	-1.0782
S_{OCN}:ECT_3	-0.2802	-2.7345
S_{OCN}:$L.E_{FF}$	0.0119	0.0902
S_{OCN}:$L.E_{LUC}$	0.1033	0.9037
S_{OCN}:$L.S_{OCN}$	-0.0149	-0.1098
S_{OCN}:$L.S_{LND}$	-0.0260	-1.4840
S_{LND} :ECT_1	0.4019	1.9227
S_{LND}:ECT_2	-0.5222	-0.8005
S_{LND}:ECT_3	0.1186	0.1500
S_{LND}:$L.E_{FF}$	-0.4310	-0.4232
S_{LND}:$L.E_{LUC}$	-0.9546	-1.0821
S_{LND}:$L.S_{OCN}$	-0.3442	-0.3287
S_{LND}:$L.S_{LND}$	0.2937	2.1761

This table displays the coefficient from the specified VECM with three cointegration vectors and one VAR lag. ECT_j is the value for the jth cointegration vector on the coefficients and L is the lag operator.

Looking at the coefficients from the VECM, we see that both E_{FF} and S_{LND} seems to be well explained by their own history, but that the cointegration vectors are important for explaining the remaining paths. Hence, after controlling for the cointegration relationship, the time series appear to be isolated physical processes which do not directly interact with each other. On a 10 percentage

confidence level, more of the series seems to be explained by the cointegration relationship and some relationship could potentially exist between S_{OCN} and S_{LND}. We also demonstrate a short out-of-sample exercise for each of the five VECM specifications. That is, we train the models using data until 2000 and recursively compute out-of-sample forecasts until 2017. Then, we compute the root mean squared errors (RMSE) for each of the series and for each models. The results are provided in Table 3.

Table 3: Out-of-sample RMSE based on the VECM where p is the number of lags included, and r denotes the number of cointegration relationships.

p	r	E_{FF}	E_{LUC}	S_{OCN}	S_{LND}	AveRMSE	AIC	BIC
1	3	0.1960	0.1856	0.1361	0.9032	0.3552	-724.9592	-661.6246
2	2	0.2035	0.1878	0.1550	1.1548	0.4253	-693.8331	-604.7176
3	1	0.2211	0.2142	0.1693	1.2592	0.4660	-658.0549	-547.6515
4	2	0.2904	0.3909	0.2434	2.3408	0.8164	-637.2530	-486.0902
5	1	0.3875	0.5534	0.3592	4.1155	1.3539	-602.4980	-431.0826

This table shows the RMSE for the VEC model out-of-sample forecasts decomposed into the four considered climate series. p corresponds to the number of VAR lags, r the corresponding cointegration relationships as identified by the Johansen test, column 3-6 the RMSE for the individual series, AveRMSE the simple average, and AIC and BIC are information criterias for the model.

Considering the average RMSE, it is clear the model with a single lag and three cointegration relationships is outperforming the remaining models in addition to obtainig the lowest information criteria. This model specification implies that the interactions between the physical systems have a short memory while all remaining covariation in the series is originating from a common stochastic trend.

Next, we consider state-space models. We estimate those individually on each of four time series. The estimated parameter values are depicted in Table 4. Only the emissions from land-use change are negative trending, while the remaining time-series have a positive estimated β. However, only the parameter value for the emissions from fossil fuels and the absorption from the ocean appear to be significantly different from zero. In a correctly specified model the residuals should be iid and gaussian distributed. Therefore, we also included the Jarque-Bera normalitly test statistic and the Ljung-Box Q-test autocorrelation test. They

indicate some sort of misspecification in the emissions from land-use change and absorption from the ocean. Note that we can use the budget equation to calculate that the flow to atmospheric carbon concentration is increasing with $0.054 = 0.128 - 0.006 - 0.031 - 0.037$, which means that we forecast an increasing atmospheric concentration in the future.

Table 4: Parameter estimates of the state space models.

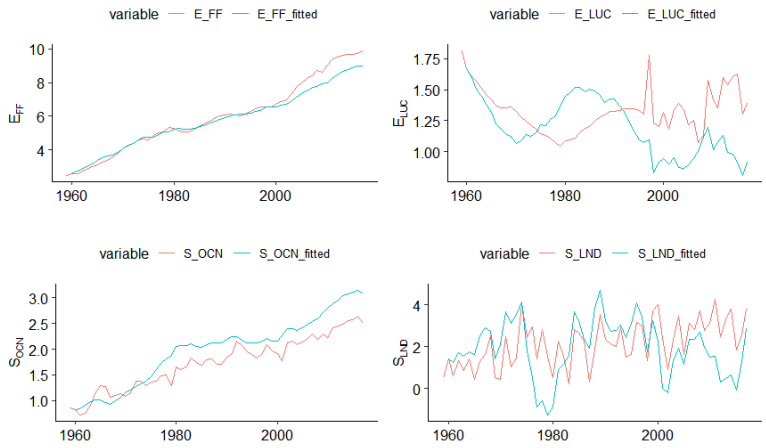
	$\hat{\sigma}_\epsilon$	t-stat($\hat{\sigma}_\epsilon$)	$\hat{\sigma}_\eta$	t-stat($\hat{\sigma}_\eta$)	$\hat{\beta}$	t-stat($\hat{\beta}$)	N	LB
E_{FF}	0.000	0.012	0.120	90.161	0.128	2.692	0.104	26.116
E_{LUC}	0.086	50.403	0.066	33.252	-0.006	-0.141	66.636	22.918
S_{OCN}	0.080	31.110	0.062	16.493	0.031	35.020	0.802	18.438
S_{LND}	0.870	569.341	0.000	0.001	0.037	0.830	0.048	57.171

This tables shows the estimated parameter values together with its t-statistic for each of the four time-series based on the state space models. In addition, the table depicts the Jarque-Beta test statistic for normality and the Ljung-Box Q test statistic for autocorrelation for each of the residuals series, respectivley. The critical value for the former is 5.11 and 31.04 for the latter.

A. In-sample fit

The next objective is to study the in-sample fit. Figure 2 and 3 show the in-sample fits of the VECM and state-space model, respectively. While the state-space model provides a better in-sample fit to both E_{FF} and E_{LUC}, the VECM provides a better fit to the remaining two series. Table 5 shows the in-sample RMSEs for the state-space model, the VECM, and the forecast combination with respect to fitting CO_2 emission and concentration, respectively. For both series, the state-space models outperforms the VECM. However, we find evidence that we may gain better in-sample fit by averaging over all models.

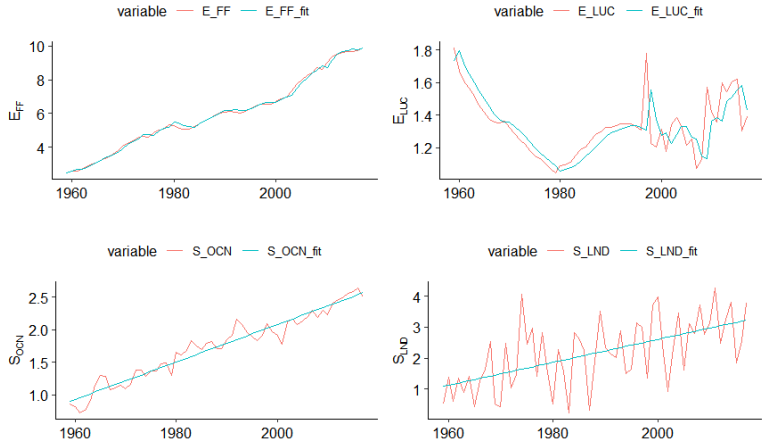
Figure 2: In sample fit of the preferred VECM specification.



.

The figure shows the in sample fit of the VCEM to each of the four considered climate series. Reading clockwise, the depicted series is the flow of fossil fuel CO_2 from fossil fuel, the flow of CO_2 from land changes, the emission absorbed by the landmass as well as for the ocean. The red lines corresponds to the actual series, whereas the blue corresponds to the model fits.

Figure 3: In sample fit of the State Space Model.



.

The figure shows the in sample fit of the state space model to each of the four considered climate series. Reading clockwise, the depicted series is the flow of fossil fuel CO_2 from fossil fuel, the flow of CO_2 from land changes, the emission absorbed by the landmass as well as for the ocean. The red lines corresponds to the actual series, whereas the blue corresponds to the model fits.

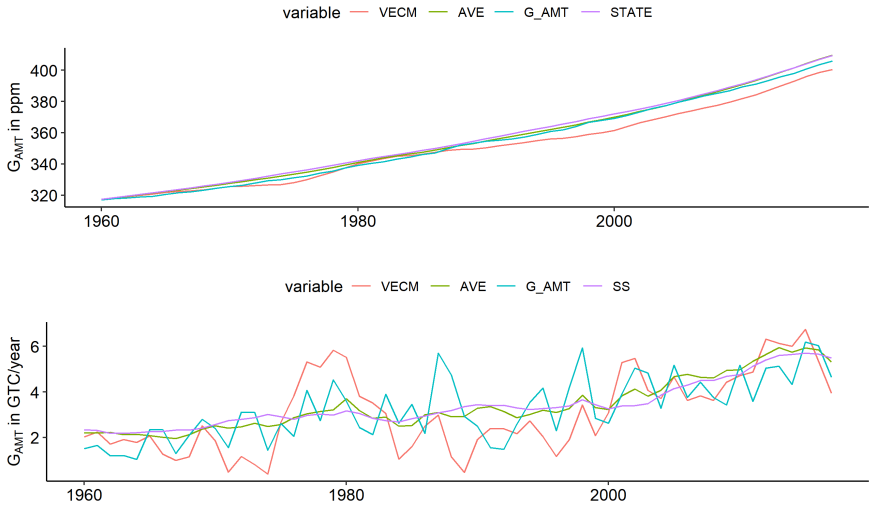
Table 5: In-sample RMSE for the state-space model (SS), the optimal VECM, and the forecast combination

	SS	VECM	AVE
CO2 emission	1.004	1.331	0.953
CO2 concentration	2.709	4.520	1.851

This table shows summary statistics for in-sample model fit for the state-space model, the VECM, and the model averaging. The first row show the RMSEs for CO2 emission, whereas the second row shows the RMSEs for CO2 concentration.

Lastly, we wish to investigate whether averaging across several forecasts helps to improve the model fit. In Figure 4 we have plotted the actual stock and flow of atmospheric concentration against the best performing VECM, the state-space model, and the averaged forecast of the five previous presented models. All forecasts seem to perform equally well in-sample, although the forecast combination seems to slightly outperform the others. Looking at the flow, we see that the averaged forecast does exactly what is expected as it is a much less noisy forecast. It seems to be in the middle of the actual series for the entire considered period. However, in-sample fit does not always translate well into a good out-of-sample performance, for which reason we next consider out-of-sample performance.

Figure 4: In-Sample-Fit of Global Carbon Dioxide Stock and Flow



.

The figure shows the stock (top) and flow (bottom) of the atmospheric CO_2 concentration and the fitted values from the VECM, the state space and from the averaging model. The red line corresponds to the fitted values from the VECM, the green from the averaging forecast and the blue the observed values.

B. Out-of-sample performance

We investigate the out-of-sample performance of the various models under consideration and whether the superiority of the model averaging carries over to out-of-sample forecast accuracy. We perform two forecasting exercises, one being similar to the exercise for the different VECM specifications. First, we split the the data set into an estimation and evaluation sample. The in-sample data covers the period 1959-2000, and the remaining sample is kept to examine the accuracy of the forecast. Second, based on the entire data set up to 2017, we forecast the future paths of the various time series. Those forecasts allows us to explore the scenarios predicted by the different models of the Representative Concentration Pathways initiative, which will be discussed in Section D.

Table 6: Out-of-sample RMSE for all considered models

The table shows the out-of-sample RMSE for all considered models decomposed into the four climate series. Column 1-4 shows the RMSE for each of the individual series while Average is their simple average.

	E_{FF}	E_{LUC}	S_{OCN}	S_{LND}	Average
AR	1.906	0.195	0.450	0.824	0.844
AAR	1.653	0.162	0.467	1.238	0.880
ARIMA	1.906	0.181	0.450	0.854	0.848
StaSpa	1.148	0.251	0.450	1.429	0.820
VECM	0.196	0.186	0.136	0.903	0.355
Combi	1.325	0.177	0.360	0.815	0.669

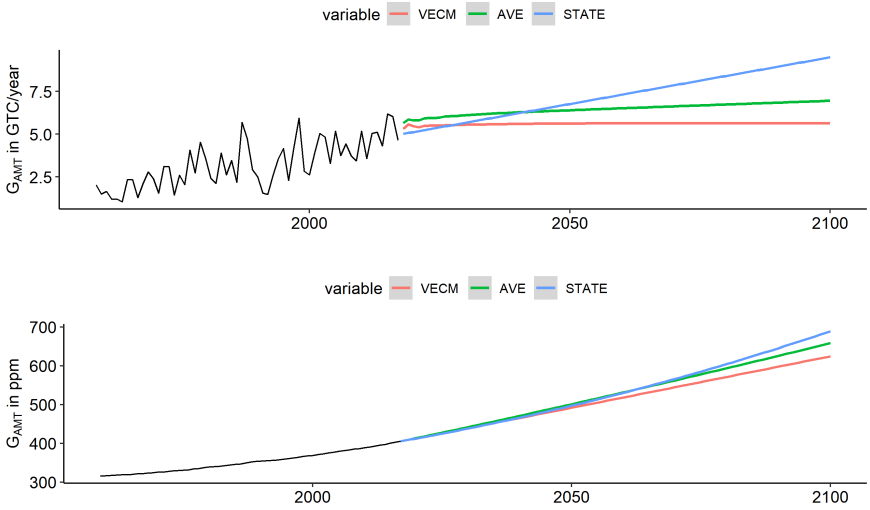
Using the first 42 observations of each time series, we fit the various models and compute out-of-sample forecasts. Based on the actual values, we compute the RMSEs for each series where we evaluate the models on their average RMSEs across the four series. The results are reported in Table 6, where we include RMSE for all models and the combination. They show that the state-space model, in comparison to the VECM, performs worse for all the considered time series.

Based on Table 6, our preferred model for out-of-sample forecasting is a VECM with one lag and three cointegration relationships.

C. Out-of-sample forecasts

We now turn to forecasts of both the emission and concentration of CO_2. We consider forecasts based on the VECM, the state-space model, and the forecast combination. We forecast recursively from 2018 to 2100. The results are shown in Fig. 5.

Figure 5: VECM vs Model-Average Forecast

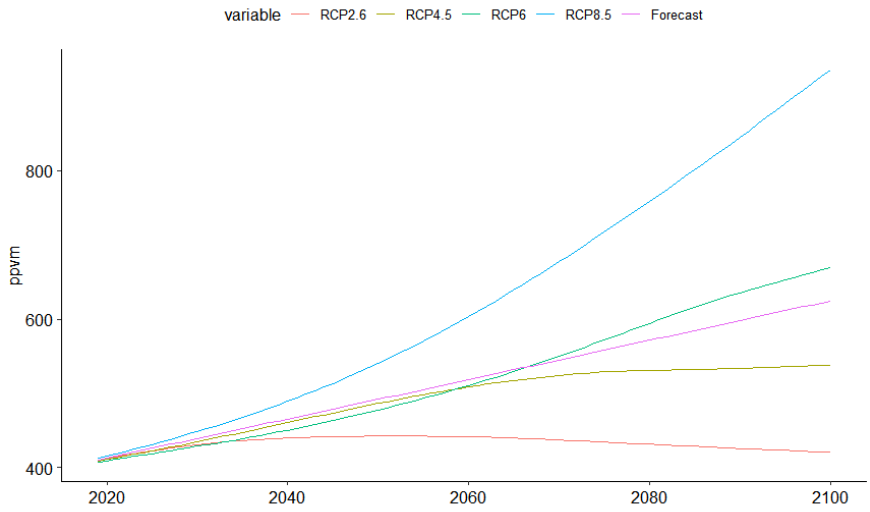


The figure shows the actual and forecasted flow (top) and stock (bottom) of the atmospheric CO_2 concentration for the VECM, the state space model and the averaging forecast.

We observe three patterns from Fig. 5. First, the forecasts of the emissions based on the state-space model have a strong trend, which is a feature of the linear Kalman filter. Second, the corresponding forecast based on the VECM appears to flatten out, which is also a model feature, namely the long-run equilibrium. As expected, the average is somewhere in between. The different paths feed into forecasted concentrations as expected. The model suggests that the CO_2 concentrations will be between 600-700 ppm in 2100 and the emission around 7.5 GtC/year, ceteris paribus.

To relate the forecasted series of both CO_2 emission and concentration to the various RCP scenarios, we plot the forecast from the VECM with the RCP scenarios in Fig. 6. Our model suggests that the CO_2 concentrations in 2100 lie between the RCP4.5 and RCP6 scenarios.

Figure 6: Predicted path for the atmospheric concentration of CO_2 with RCPs



This figure shows the four RCP global hemispheric CO_2 scenarios until the year 2100 and the predicted transmission of the VECM if no change is made. We see that the current path will place the level somewhere between RCP6 and RCP4.5.

D. Managing atmospheric CO2 concentration - Counterfactual Evidence

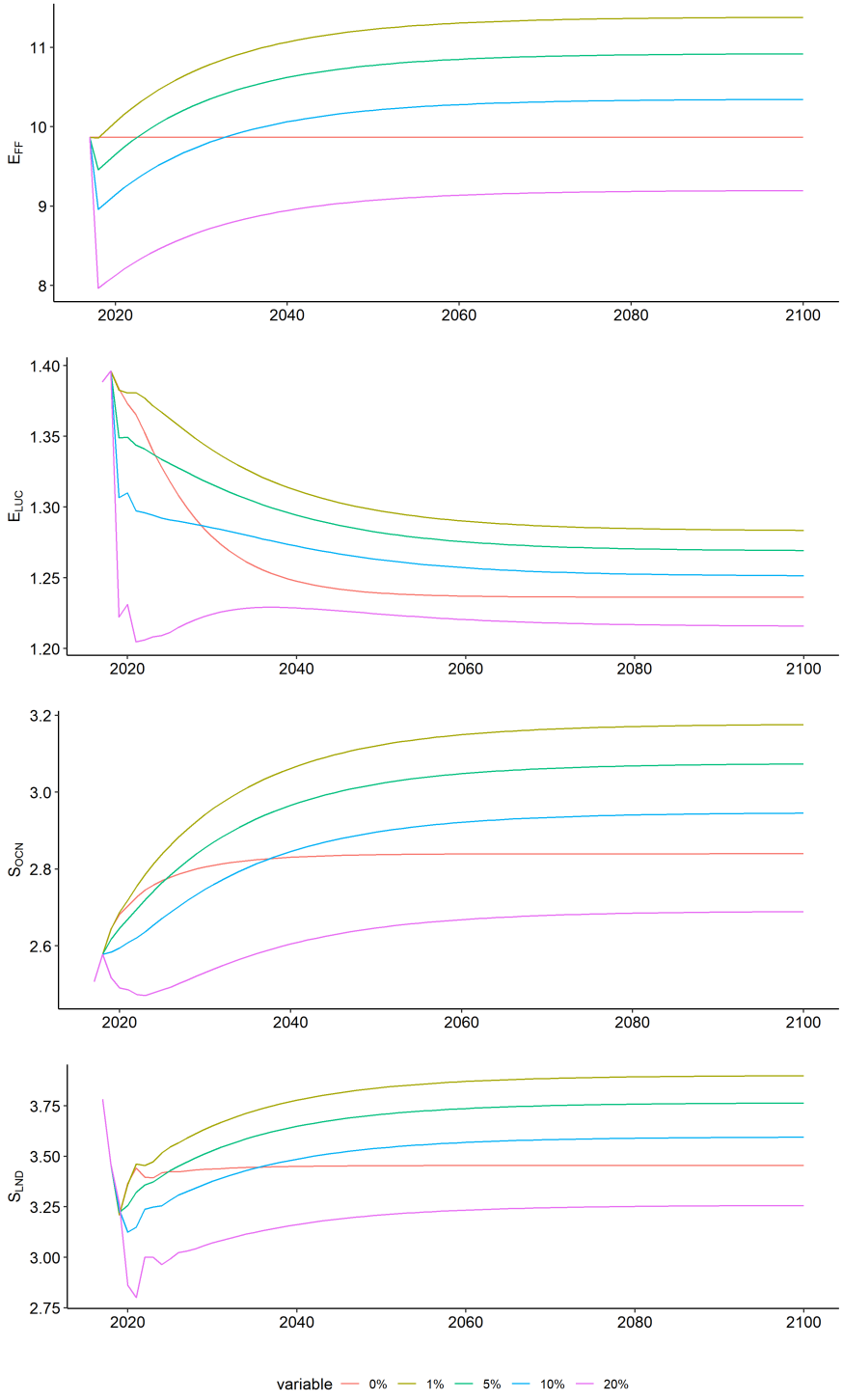
In the last section, we consider the scenarios of the Representative Concentration Pathways initiative (Meinshausen, Smith, Calvin, Daniel, Kainuma, Lamarque, Matsumoto, Montzka, Raper, Riahi et al., 2011b), denoted as RCP2.6, RCP4.5, RCP6, and RCP8.5. The RCPs consider global surface temperature scenarios between 1.5 degrees warming above pre-industrial levels (RCP2.6) to 4.5 degrees warming (RCP8.5) by the year 2100. Above, we have shown that the forecasted carbon dioxide might lie in between the second and third RCP scenario. Using the VECM model, we now approach an important policy question related to what the RCP concentration scenarios imply for admissible global emission paths. In addition, we formulate requirements for global emission reductions corresponding to the RCP2.6.

First, however, we restrict our focus to emissions from fossil fuel burning, cement production, and gas flaring (E_t^{FF}) relative to emissions from land-use change (E_t^{LUC}), the ocean sink (S_t^{OCN}), and the land sink (S_t^{LND}). We do this for two reasons. First, we believe that emissions from fossil fuel burning, cement production, and gas flaring are to a larger extent driven by human production. In

Fig. 1, we observe that this component has been the main driver of the increase in CO_2 emissions. Second, we may synthesize different pathways for (E_t^{FF}), which will impact the forecasts for the remaining three series due to their cointegrating relationships. This is a clear advantage of the VECM.

We start by specifying five different pathways for (E_t^{FF}). First, we synthesize a pathway of status quo such that (E_t^{FF}) stays at the 2017 level. Second, we consider four different reduction pathways, namely, where we reduce the emissions from fossil fuel burning, cement production, and gas flaring from the forecasted level by 1 %, 5 %, 10 %, and 20 %. We feed the different pathways into the VECM model and forecast the remaining three series given the pathway in force. Consequently, the forecasts take cointegrating into account. We plot the forecasts of the series in Fig. 7.

Figure 7: Emission forecasts using different pathways for (E_t^{FF})

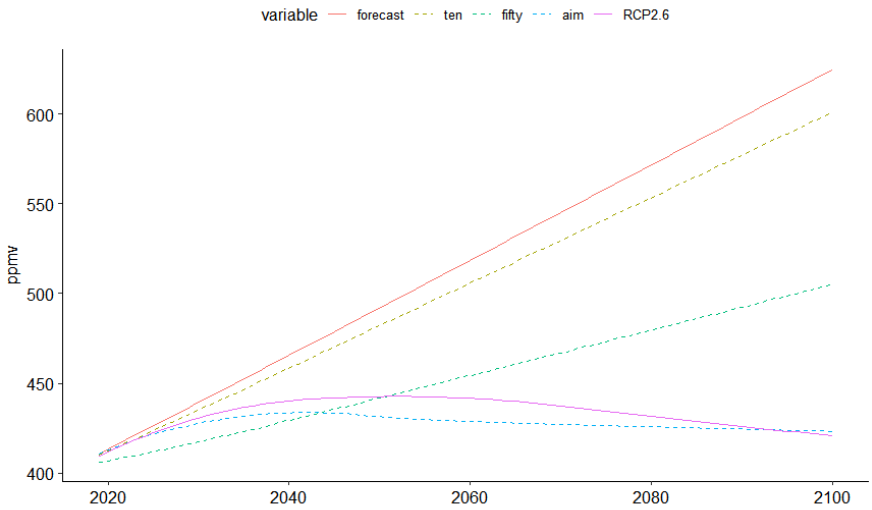


The plot shows the effect of a 0 to 20 percentage reduction in the flow of fossil fuel (i.e. a reduction in E_{FF}) and its forecasted effect on the remaining climate series. Reading from top to bottom, we have the series for E_{FF}, E_{LUC}, S_{OCN} and S_{LND}.

Fig. 7 gives an overview of the paths for the remaining three components in the carbon budget equation for different pathways of (E_t^{FF}). The various percentage reductions imply monotonic reactions in the pathways for the three forecasted series. That is, if we are able to reduce the fossil fuel emission every year by 1 % relative to our forecast, then the remaining series will all be on a higher pathway compared to the case, in which we are able to reduce the fossil fuel emission by any percentage greater than 1 %. The status quo case, however, implies that the forecasted series intersect the other pathways and, for every series, end up between the scenario of 10 % and 20 % reduction, respectively.

Our objective, however, is to compare different emission pathways to the RCP scenarios, for which reason we convert the forecasted emissions to concentrations (see Raupach, Gloor, Sarmiento, Canadell, Frölicher, Gasser, Houghton, Le Quéré, and Trudinger (2014) for details). We extend the pathways and consider in addition yearly emission reductions by 50 %, 80 % and 90 %. Finally, we include a pathway that roughly corresponds to the declared goal of the EU-commission to become CO_2 neutral by 2050. We simulate a similar pathway for fossil fuel emission, in which we reach zero fossil fuel emission by 2050. In the intermediate years, we reduce the emission linearly on an annual basis. The results are given in Table 7, while we plot the pathways corresponding to a 10 % and 50 % yearly reduction, respectively, along with the EU2050-pathway in Fig. 8.

Figure 8: Fuel-based Carbon dioxide emission reduction simulation



. The figure shows simulated paths for the atmospheric concentration of CO_2 for different levels of reduction in the emission of fossil fuel (E_{FF}). The solid lines are the forecasted path (red) and the best possible RCP (purple). The dotted lines show the paths for ten and fifty percent reduction along with a path generated to hit the target precisely. The estimated reduction in fossil fuel emission is 85 percent in this case.

Table 7: VECM based carbon dioxide concentration simulations.

year	RCP2.6	forecast	0%	1%	5%	10%	20%	50%	80%	90%	EU2050
2020	412	413	413	413	413	412	411	407	403	402	413
2025	423	426	425	426	425	423	421	412	403	400	422
2030	431	439	437	439	437	435	431	417	404	400	428
2035	437	452	448	452	450	447	441	423	406	400	432
2040	440	466	460	465	462	458	451	429	407	400	434
2045	442	479	471	478	474	470	461	435	409	401	433
2050	443	492	482	491	487	482	472	442	412	402	431
2055	443	505	494	504	500	494	482	448	414	402	430
2060	442	518	505	517	512	506	493	454	416	403	429
2065	440	532	516	530	525	518	503	461	418	404	428
2070	437	545	527	543	537	529	514	467	420	405	427
2075	435	558	539	557	550	541	524	473	422	405	427
2080	432	572	550	570	562	553	535	480	424	406	426
2085	429	585	561	583	575	565	545	486	427	407	425
2090	426	598	573	596	587	577	556	492	429	408	425
2095	423	611	584	609	600	589	566	499	431	408	424
2100	421	625	595	622	613	601	577	505	433	409	423

The table shows the atmospheric carbon concentration for with 5 year intervals for different levels of reduction in fossil fuel emission (E_{FF}). The far right column shows the path set out in the EU 2050 goals, continued until 2100.

From Fig. 8, we derive three main insights. First, continuing on the forecasted trajectory implies large increases in the atmospheric CO_2 concentration in the range of 50 % higher than the RCP2.6. Second, even if fossil fuel emissions are reduced by 50 % yearly, the trajectory do not coincide with the RCP2.6. Last, if fossil fuel emissions reach zero in 2050 and maintain that level, the forecasted CO_2 concentration coincides with RCP2.6. Another way to match the RCP2.6 is by reducing the annual fossil fuel emission by 85 %.

VI. Conclusion

In this paper, we have estimated several statistical models of the global carbon cycle. We find that a Vector-Error-Correction Model (VECM) describes the global carbon cycle best, in the sense that it produces the most reliable out of sample forecasts. We carry out several forecasting exercises and find that the future path

of carbon dioxide concentration will lie in between the two middle RCP pathways (RCP4.5 and RCP6.0) if no steps to change the current path is taken. Our results suggest that drastic reductions in global anthropogenic carbon dioxide emissions are necessary in order to achieve atmospheric CO_2 concentrations in the year 2100 that are in line with the more ambitious RCP2.6 scenario.

Overall, we suggest to approach our results with great caution. Forecasting as far in the future as the year 2100 is not a trivial thing to do; economists regularly fail in predicting the state of the economy only a very few years ahead. Forecasting carbon dioxide emission broadly shares some key features of the main difficulties in macroeconomic forecasting. First and foremost, our estimates suffer from short sample sizes. In the climate science and climate econometrics literature, researchers are e.g. struggeling to reconcile how the decrease in global surface temperature between 1996 and 2008 matches with the evidence of anthropogenically caused climate change (Kaufmann, Kauppi, Mann, and Stock (2011)). Having longer time series would clearly facilitate the detection of long climatic cycles that could otherwise be missclassified as structural breaks in climate models.

In a recent article, Pretis (2017) stresses that the climate econometrics literature suffers from plausibly unreasonable exogeneity assumptions on the interaction between climate and economy. He argues that one strand of the literature takes the economy as fixed and models climate change as a response to given economic activity (e.g. Kaufmann et al. (2011) or Estrada, Perron, and Martínez-López (2013)), while another stresses human and social responses to climate change (see Dell, Jones, and Olken (2014), Jones and Olken (2010)). The models and forecasts in our paper are in the spirit of the first fixed-economy literature, as we do not properly consider endogenous feedback mechanism between climate change and human behavior.

Implicitly, our model forecast also assume that the economic policy space will be stable. If political organizations e.g. decided to drastically tax carbon dioxide emission, economic theory predicts that the economy should substitue away from carbon dioxide based economic activity.

As a last point, our forecasts completely neglect the role of economic innovation.

Nordhaus (2010) argues that neglecting economic innovation will most likely lead to overestimates of climate change, but admits that the economic profession has little sense of the magnitude of the bias.

References

BATES, J. M. AND C. W. GRANGER (1969): "The combination of forecasts," *Journal of the Operational Research Society*, 20, 451–468.

BENNEDSEN, M., E. HILLEBRAND, AND S. J. KOOPMAN (2018): "Trend analysis of the airborne fraction and sink rate of anthropogenically released CO2," *Biogeosicences Discuss.*

BODEN, T. A., G. MARLAND, AND R. J. ANDRES (2009): "Global, regional, and national fossil-fuel CO2 emissions," *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA doi*, 10.

DELL, M., B. F. JONES, AND B. A. OLKEN (2014): "What do we learn from the weather? The new climate-economy literature," *Journal of Economic Literature*, 52, 740–98.

DLUGOKENCKY, E. AND P. TANS (2018): "Trends in atmospheric carbon dioxide," *National Oceanic Atmospheric Administration*.

ESTRADA, F., P. PERRON, AND B. MARTÍNEZ-LÓPEZ (2013): "Statistically derived contributions of diverse human influences to twentieth-century temperature changes," *Nature Geoscience*, 6, 1050.

FULLER, W. A. (2009): *Introduction to statistical time series*, vol. 428, John Wiley & Sons.

HILLEBRAND, E. (2019): "Case for the Econometric Games 2019," .

JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, 12, 231–254.

JONES, B. F. AND B. A. OLKEN (2010): "Climate shocks and exports," *American Economic Review*, 100, 454–59.

KAUFMANN, R. K., H. KAUPPI, M. L. MANN, AND J. H. STOCK (2011): "Reconciling anthropogenic climate change with observed temperature 1998–2008," *Proceedings of the National Academy of Sciences*, 108, 11790–11793.

KOOPMAN, S. J., N. SHEPHARD, AND J. A. DOORNIK (1999): "Statistical algorithms for models in state space using SsfPack 2.2," *The Econometrics Journal*, 2, 107–160.

MACKINNON, J. G. (1996): "Numerical distribution functions for unit root and cointegration tests," *Journal of applied econometrics*, 11, 601–618.

MEINSHAUSEN, M., S. J. SMITH, K. CALVIN, J. S. DANIEL, M. KAINUMA, J.-F. LAMARQUE, K. MATSUMOTO, S. MONTZKA, S. RAPER, K. RIAHI, ET AL. (2011a): "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300," *Climatic change*, 109, 213.

——— (2011b): "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300," *Climatic change*, 109, 213.

NORDHAUS, W. D. (2010): "Modeling induced innovation in climate-change policy," in *Technological change and the environment*, Routledge, 188–215.

PERRON, P. (1988): "Trends and random walks in macroeconomic time series: Further evidence from a new approach," *Journal of economic dynamics and control*, 12, 297–332.

PRETIS, F. (2017): "Exogeneity in climate econometrics," .

QUERE, A. FRIEDLINGSTEIN, SITCH, HAUCK, PONGRATZ, PICKERS, KORSBAKKEN, PETERS, A. CANADELL, AND ARORA (2018): "Global Carbon Budget 2018," *Earth Syst. Sci. Data*.

RA HOUGHTON, A. N. (2017): "Global and regional fluxes of carbon from land use and land cover change 1850–2015," *lobal Biogeochemical Cycles*.

RAUPACH, M. R., M. GLOOR, J. L. SARMIENTO, J. CANADELL, T. L. FRÖLICHER, T. GASSER, R. A. HOUGHTON, C. LE QUÉRÉ, AND C. M. TRUDINGER (2014): "The declining uptake rate of atmospheric CO 2 by land and ocean sinks," *Biogeosciences*, 11, 3453–3475.

TIMMERMANN, A. (2006): "Chapter 4 Forecast Combinations," in *Handbook of Economic Forecasting*, Elsevier, 135–196.

A. Appendix

Table 8: VECM-based carbon dioxide simulations.

year	RCP2.6	forecast	0%	1%	5%	10%	20%	50%	80%	90%	EU2050
2020	412	413	413	413	413	412	411	407	403	402	413
2022	417	419	418	418	418	417	415	409	403	401	417
2024	421	424	423	424	422	421	419	411	403	400	420
2026	424	429	427	429	427	426	423	413	403	400	423
2028	428	434	432	434	432	430	427	415	404	400	426
2030	431	439	437	439	437	435	431	417	404	400	428
2032	433	445	441	444	442	440	435	420	405	400	430
2034	436	450	446	449	447	444	439	422	405	400	431
2036	438	455	450	454	452	449	443	424	406	400	432
2038	439	460	455	460	457	454	447	427	407	400	433
2040	440	466	460	465	462	458	451	429	407	400	434
2042	441	471	464	470	467	463	455	432	408	400	434
2044	442	476	469	475	472	468	459	434	409	401	434
2046	442	481	473	481	477	473	464	437	410	401	433
2048	442	487	478	486	482	477	468	439	411	401	432
2050	443	492	482	491	487	482	472	442	412	402	431
2052	443	497	487	496	492	487	476	444	412	402	431
2054	443	503	491	501	497	491	480	447	413	402	430
2056	442	508	496	507	502	496	484	449	414	402	430
2058	442	513	500	512	507	501	489	452	415	403	429
2060	442	518	505	517	512	506	493	454	416	403	429
2062	441	524	509	522	517	510	497	457	417	403	428
2064	440	529	514	528	522	515	501	459	418	404	428
2066	440	534	518	533	527	520	505	462	418	404	428
2068	439	540	523	538	532	525	510	464	419	404	427
2070	437	545	527	543	537	529	514	467	420	405	427
2072	436	550	532	549	542	534	518	469	421	405	427
2074	435	556	536	554	547	539	522	472	422	405	427
2076	434	561	541	559	552	544	526	475	423	405	426
2078	433	566	546	564	557	548	531	477	424	406	426
2080	432	572	550	570	562	553	535	480	424	406	426
2082	431	577	555	575	567	558	539	482	425	406	426
2084	429	582	559	580	572	563	543	485	426	407	425
2086	428	587	564	585	577	567	547	487	427	407	425
2088	427	593	568	591	582	572	552	490	428	407	425
2090	426	598	573	596	587	577	556	492	429	408	425
2092	425	603	577	601	593	582	560	495	430	408	424
2094	424	609	582	606	598	586	564	497	431	408	424
2096	423	614	586	612	603	591	568	500	431	409	424
2098	422	619	591	617	608	596	573	502	432	409	424
2100	421	625	595	622	613	601	577	505	433	409	423

Table 9: Result from ADF test for a unit root

Table 10

lag	ADF_1	p.value	ADF_2	p.value	ADF_3	p.value
0	7.483	0.990	0.593	0.988	-0.974	0.935
1	3.463	0.990	0.253	0.973	-1.534	0.760
2	2.822	0.990	0.007	0.954	-1.632	0.719
3	2.139	0.990	-0.099	0.942	-1.924	0.598
0	-0.870	0.365	-3.978	0.010	-3.961	0.018
1	-0.806	0.387	-3.109	0.034	-3.176	0.100
2	-0.751	0.407	-2.541	0.122	-2.747	0.269
3	-0.560	0.475	-2.547	0.120	-2.746	0.269
0	1.409	0.957	-1.104	0.657	-4.162	0.010
1	1.581	0.970	-1.140	0.644	-4.267	0.010
2	2.074	0.990	-1.311	0.585	-4.065	0.013
3	2.415	0.990	-1.298	0.589	-3.869	0.022
0	-1.783	0.075	-5.646	0.010	-7.965	0.010
1	-1.029	0.308	-4.177	0.010	-7.206	0.010
2	-0.222	0.577	-2.606	0.099	-4.457	0.010
3	0.158	0.685	-1.893	0.370	-3.365	0.070

Model 1 is a model without a constant nor a trend, model 2 is for a model with a constant and model 3 is with both a constant and a trend. Reading from the top, the four tests is for E_{FF}, E_{LUC}, S_{OCN} and S_{LND} .

Table 11: Result from Dickey Fuller's test for unit root with the Mckinnons test statistic

Table 12

lag	DF	p.value
0	7.483	1
1	3.463	1.000
2	2.822	0.999
3	2.139	0.992
0	-0.870	0.335
1	-0.806	0.362
2	-0.751	0.387
3	-0.560	0.470
0	1.409	0.959
1	1.581	0.971
2	2.074	0.990
3	2.415	0.996
0	-1.783	0.071
1	-1.029	0.270
2	-0.222	0.602
3	0.158	0.728

The table shows the result of the Fuller (2009) test using the MacKinnon (1996) test statistics. Reading from the top, the four tests is for E_{FF}, E_{LUC}, S_{OCN} and S_{LND} .

Table 13: Result from Phillips-Perron test for unit root

Table 14

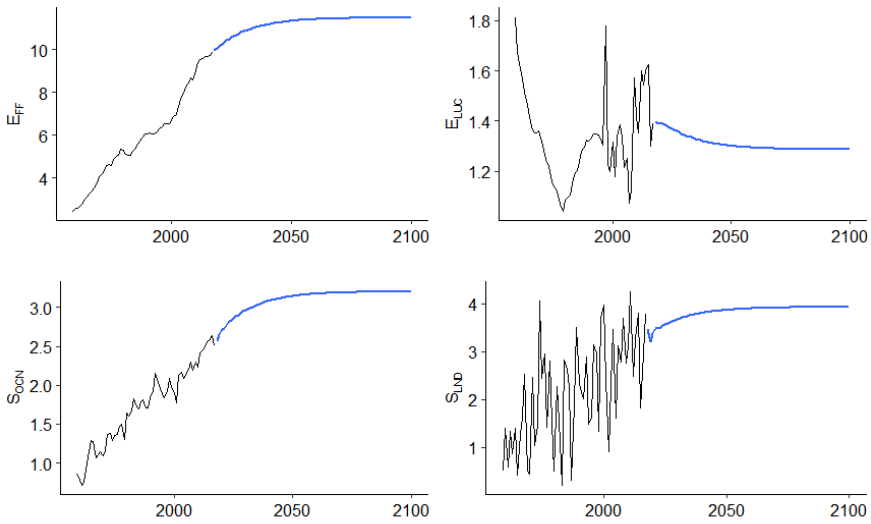
lag	DF	p.value
0	7.483	1
1	3.463	1.000
2	2.822	0.999
3	2.139	0.992
0	-0.870	0.335
1	-0.806	0.362
2	-0.751	0.387
3	-0.560	0.470
0	1.409	0.959
1	1.581	0.971
2	2.074	0.990
3	2.415	0.996
0	-1.783	0.071
1	-1.029	0.270
2	-0.222	0.602
3	0.158	0.728

The table shows the result of the Perron (1988) test using. Reading from the top, the four tests is for E_{FF}, E_{LUC}, S_{OCN} and S_{LND} .

Table 15: Jarque–Bera test for normality of the VECM.

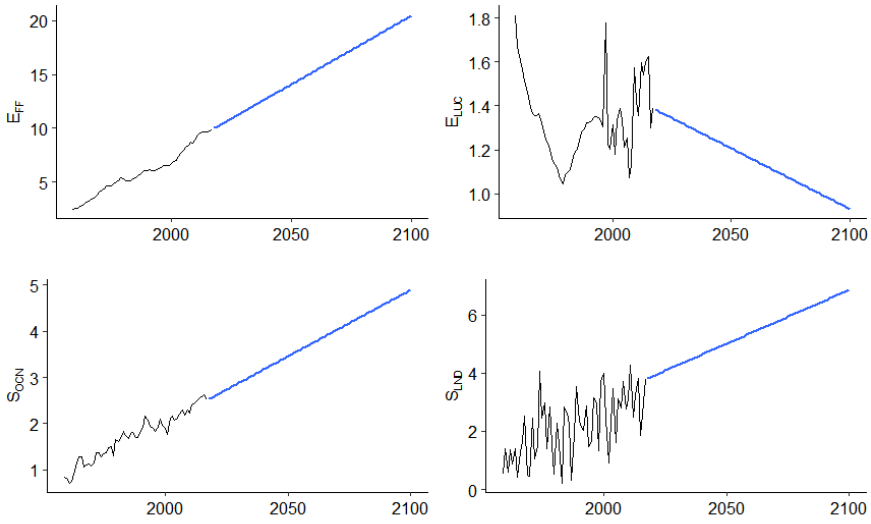
	E_{FF}	E_{LUC}	S_{OCN}	S_{LND}
test statistic	9.381	1.207	2.644	3.285
p_value	0.009	0.547	0.267	0.194

Figure 9: Out-of-sample predictions from the VECM



This figure shows out-of-sample predicted values for E_{FF}, E_{LUC}, S_{LND} and S_{OCN} in clockwise order from the VECM.The black line corresponds to the observed values and the blue the estimated out of sample series.

Figure 10: Out-of-sample Predictions of the Kalman filter.



The plot shows the out of sample forecast from the four considered climate series when using the state space model. The black line is the observed observed series and the blue line is the out-of-sample forecast till 2100. Reading clockwise, the depicted series are: E_{FF}, E_{LUC}, S_{LND} and S_{OCN}.