2022-13
Working paper. Economics
ISSN 2340-5031

REVISITING GRANGER CAUSALITY OF CO2 ON GLOBAL WARMING: A QUANTILE FACTOR APPROACH.

LIANG CHEN, JUAN J.DOLADO, JESÚS GONZALO and ANDREY RAMOS.

uc3m | Universidad Carlos III de Madrid

Serie disponible en http://hdl.handle.net/10016/11

Web: http://economia.uc3m.es/

Correo electrónico: departamento.economia@eco.uc3m.es

Revisiting Granger Causality of CO_2 on Global Warming: A Quantile Factor Approach *

Liang Chen[1], Juan J. Dolado[2], Jesús Gonzalo[3], and Andrey Ramos[3]

[1] *Peking University HSBC Business School, chenliang@phbs.pku.edu.cn*
[2] *Department of Economics, Universidad Carlos III de Madrid, dolado@eco.uc3m.es*
[3] *Department of Economics, Universidad Carlos III de Madrid, jgonzalo@est-eco.uc3m.es*
[3] *Department of Economics, Universidad Carlos III de Madrid, anramosr@eco.uc3m.es*

March 29, 2022

Abstract

The relationship between global warming and CO_2 is a long-standing question in the climate change literature. In this paper we revisit this topic through the lenses of a new class of factor models for high-dimensional panel data, labeled Quantile Factor Models (QFM). This technique allows us to extract quantile-dependent factors from the distributions of changes in temperatures across a wide range of stable weather stations in the Northern and Southern Hemispheres over a century (1917-2018). In particular, we test whether CO_2 emissions/concentrations Granger-cause the underlying factors of the different quantiles of the distribution of changes in temperature, and find that they exhibit much higher predictive power on large negative and medium (lower and middle quantiles) than on large positive changes (upper quantiles). These findings are novel in this literature and complement recent results by Gadea and Gonzalo (2020) who document the existence of steeper trends in lower temperature levels than in other parts of the distribution.

Keywords: Global warming, CO2 emissions, Quantile factor models, Granger causality.
JEL codes: C31, C33, Q54.

*We are indebted to Lola Gadea for helpful comments. Financial support from the National Natural Science Foundation of China (Grant No.71703089), the Spanish Ministerio de Economía y Competitividad (grant PID2019-104960GB-100), and MadEco-CM (grant S205/HUM-3444) is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

As stressed by world leaders during the recent UN Climate Change Conference of the Parties (COP26) in Glasgow, one of the more pressing issues in the international policy agenda is fighting higher global surface temperatures (global warming; GW hereafter). A proper design of climate policy requires a deep understanding of the relationship between between global carbon dioxide (CO_2) emission and GW prior to assessing the impact of the latter phenomenon on economic activity. This is by all means a long-standing issue in climate change science and economics. (see e.g. Hansen et al. 1981, Hsiang and Kopp 2018, and Castle and Hendry 2020)[1] [2]

For example, a traditional research question in this literature has been the size of the so-called transient climate response (TCR), defined as the change in global mean temperature at the time of doubling of atmospheric concentration increasing at a rate of 1 percent per year (see e.g. Montamat and Stock 2020). As explained in Myhre et al. (2013), the radioactive forcing arising from an increase in CO_2 is proportional to the growth rate of CO2 emissions or the acceleration rate of CO_2 concentrations. Thus, this establishes a linear relationship between temperature changes and the the first (second) difference of (logged) CO_2 emissions (concentrations). There is a plethora of studies estimating this causal relationship through time series techniques (e.g. testing for Granger causality; see Stips et al. 2016 and the references therein). Typically long time series of the change in average temperatures across a large number of weather stations are regressed on their lagged values and those of the growth rate of CO_2 emissions to check whether the coefficients of the latter are statistically significant. The goal of this paper is to revisit this Granger-causality analysis by focusing on time series of the *entire* distributions of temperatures across a wide range of stable weather stations around the world and its main characteristics (moments, quantiles, etc.) rather than focus exclusively on their averages, as is customary in this literature. This novel approach is relevant since, though the average temperature might not display any discernible growth pattern as CO_2 changes, the lower or upper tails might exhibit a clear increase. In other words, even when the average temperature exhibits some growth, having a wide angle picture of the trending behavior of the whole distribution will help in designing effective policies preventing CO_2 emissions and concentrations.

This issue can be clarified by means of the following example. Suppose that data is available on two stationary variables (after suitable transformations, such as first or second differences):

[1] When sunlight reaches Earth, its surface absorbs some of the light's energy and re-radiates it as infrared waves. These waves travel up into the atmosphere and will escape back into space if unimpeded. For example, unlike oxygen and nitrogen which do not interfere with infrared waves in the atmosphere because molecules are picky about the range of wavelengths that they interact with, CO_2 and other greenhouse gases absorb energy at a variety of wavelengths whose ranges do overlap with that of infrared energy. As CO_2 soaks up this infrared energy, it vibrates and re-emits the infrared energy back in all directions. About half of that energy goes out into space, and about half of it returns to Earth as heat, contributing to GW through the "greenhouse effect".

[2] The greenhouse effect was first discovered by Fourier (1824), experimentally verified by Foote (1856) and Tyndall (1863), and quantified by Arrhenius (1896).

$\{X_{it}\}$ which represents temperatures measured at $i = 1, 2, ...N$ weather stations over $t = 1, 2, ...T$ periods, and $\{Z_t\}$ which captures aggregate CO_2 emissions over the same sample period. Typically, if researchers are interested in estimating the heterogeneous effects of Z_t on GW at different parts of the temperatures distribution, they would use average temperatures across stations, i.e. $X_t = N^{-1} \sum_{i=1}^{N} X_{it}$ to run the following quantile regression (QR): $X_t = \alpha(\tau) + \beta(\tau) Z_t + u_t(\tau)$, with quantiles $0 < \tau < 1$, such that $\alpha(\tau) + \beta(\tau) Z_t$ denotes the conditional quantile of X_t given Z_t. Although this represents a step forward with respect to a standard linear regression model, this standard QR estimation procedure throws away all the information on the distribution of temperatures across different stations. An alternative procedure could be to run QR for each unit (weather station) and then average across units for each quantile. However, this is a doubtful statistical procedure since the average of quantiles is not equal to the quantile of the averages. As a result, a simple way to enrich this model would be to consider a factor structure for the panel data on temperatures with $X_{it} = \lambda_i'(\tau) f_t(\tau) + u_{it}(\tau)$, where $f_t(\tau)$ and $\lambda_i(\tau)$ are a $r(\tau) \times 1$ vector of factors and loadings, respectively, which may differ at each τ. Having obtained consistent estimates of the quantile-dependent objects, a natural OLS regression to run for each factor, $f_t(\tau)$ would be the following:

$$f_{jt}(\tau) = \beta_j Z_t + \epsilon_{jt}, \text{ for } j = 1, \ldots, r(\tau), \tag{1}$$

which provides estimates of the effects of CO_2 on the conditional quantile of the corresponding factor at a given τ. Note that, if instead of $\{Z_t\}$ we were to consider its lagged values, a test on the joint statistical significance of its coefficients provides a test of the null hypothesis on Granger-causality of CO_2 on the quantiles of the factors.

To obtain the above-mentioned quantile-dependent objects, we rely on a novel methodology of *Quantile Factor models* (QFM in short) recently proposed by Chen et al. (2021) (CDG henceforth). This approach extends the theory of approximate factor models (AFM), designed to extract common factors at the mean of the distribution of large panel datasets (see Chamberlain and Rothschild 1983), to their quantiles. As is well known, AFM imply that a panel $\{X_{it}\}$ of N variables (units), each with T observations, has the representation $X_{it} = \lambda_i' f_t + \epsilon_{it}$, where $\lambda_i = [\lambda_{i1}, .., \lambda_{ir}]'$ and $f_t = [f_{t1}, .., f_{tr}]'$ are $r \times 1$ vectors of factor loadings and common factors, respectively, with $r \ll N$, and where $\{\epsilon_{it}\}$ are zero-mean weakly dependent idiosyncratic disturbances which are uncorrelated with the factors. The availability of fairly straightforward estimation procedures for AFM, e.g. via Principal Components Analysis (PCA) or similar methods, has led to their widespread use in many fields of economics (see Bai and Ng 2008 and Stock and Watson 2011 for overviews).[3]

In line with the generalization of linear regression to QR models, CDG (2021) argue that the

[3]More recently, PCA has also been used to model interactive fixed-effects models in linear and non-linear models (see Bai 2009 and Chen et al. 2018), common correlated effects (see Pesaran 2006), and for predictive analytics in Big Data (see Athey and Imbens 2019).

standard regression interpretation of static AFM as linear conditional mean models of X_{it} given f_t (i.e. $\mathbb{E}(X_{it}|f_t) = \lambda_i' f_t$), prevents capturing hidden factors that may shift specific characteristics (moments or quantiles). The insight is that neither the loadings λ_i nor the factors f_t are allowed to vary across the distributional characteristics of each unit in the panel.

As an illustration of the above-mentioned limitations of AFM, consider the factor structure in a standard *location-scale shift* (LSS) model with the following Data Generating Process (DGP): $X_{it} = \alpha_i f_{1t} + \eta_i f_{2t}\epsilon_{it}$, with $f_{1t} \neq f_{2t}$ (both are scalars), $\eta_i, f_{2t} > 0$ and $\mathbb{E}(\epsilon_{it}) = 0\acute{}$. The first factor ($f_{1t}$) shifts location, whereas the second factor (f_{2t}) shifts the scale and therefore governs the volatility of shocks to X_{it}.[4] Such a DGP can be rewritten in QR format as $X_{it} = \lambda_i'(\tau) f_t + u_{it}(\tau)$, with $0 < \tau < 1$, $\lambda_i(\tau) = [\alpha_i, \eta_i \mathsf{Q}_\epsilon(\tau)]'$, where $\mathsf{Q}_\epsilon(\tau)$ represents the quantile function of ϵ_{it}, $f_t = [f_{1t}, f_{2t}]'$, $u_{it}(\tau) = \eta_i f_{2t}[\epsilon_{it} - \mathsf{Q}_\epsilon(\tau)]$, and the conditional quantile $\mathsf{Q}_{u_{it}(\tau)}[\tau|f_t] = 0$.[5] It is easy to check that PCA will only extract the location-shifting factor f_{1t} in this model, but it fails to capture the scale-shifting factor f_{2t}.

To overcome this shortcoming, CDG (2021) propose QFM (whose definition is provided in Section 2), and their estimation procedure, called *Quantile Factor Analysis* or QFA in short (see Section 3). QFA allows to estimate the space spanned by f_{1t} and f_{2t} in the previous DGP and, in general, the loadings, factors and the number of factors (i.e. $\lambda_i(\tau)$, $f_t(\tau)$ and $r(\tau)$ for $\tau \in (0,1)$), which could all vary across quantiles.[6] Hence, as argued earlier, QFM could be thought of as capturing the same type of flexible generalization that QR techniques represent for linear regression models.[7]

The QFA estimation procedure relies on the minimization of the standard *check* function in QR (instead of the quadratic loss function used in AFM) to estimate jointly the common factors $f_t(\tau)$ and the loadings $\lambda_i(\tau)$ at a given quantile τ, once the number of factors has been selected by a consistent criterion. CDG (2021) derive the average rates of convergence of the objects of interest and establish their asymptotic normality based on smoothed QR (see e.g., Galvao and Kato 2016). Lastly, it is noteworthy that, given that QFA estimation captures all quantile-shifting factors (including those affecting the means of observed variables), these asymptotic results provide a natural way to differentiate AFM from QFM. This is specially relevant in the presence of outliers where AFM may not work well, whereas in contrast QFM will render valid estimation and inference. The insight is similar to that underlying the use of robust least median regression when outliers abound as in Huber (1981). This issue is relevant in our dataset of temperatures consisting of the annual changes of temperatures recorded at 441

[4] Notice that the simplifying assumption of a known number of factors in this specific example is later relaxed.

[5] Throughout the paper we use $\mathsf{Q}_W[\tau|Z]$ to denote the conditional quantile of W given Z.

[6] Note that in this particular DGP, since f_{1t} can be consistently estimated by PCA, it is also feasible to estimate f_{2t} by applying PCA to the squared residuals stemming from subtracting the factor structure at the mean from the original variables. However, in practice the DGP is unknown and therefore QFA is needed.

[7] Related to CDG (2021), Ando and Bai (2020) use a similar setup with an unobservable factor structure which is also allowed to be quantile dependent; yet, their assumptions are more restrictive since all the moments of the idiosyncratic errors are required to exist.

weather stations from 1917 to 2018. Figure 1 displays the fraction of outliers in each year of the sample, corresponding to those observations that exceed three standard deviations, which ranges between 2 and 11 percent and it is increasing over the sample period. Hence, as a byproduct of the analysis, we will illustrate the advantages of using QFA in the presence of outliers through several Monte-Carlo simulations.

In parallel with QFA, use is made of the of the statistical procedure proposed by Gadea and Gonzalo (2020) to test for the presence of deterministic and/or stochastic trends in a wide range of moments of the unconditional distribution of temperatures across a large number of weather stations over a long period. Using the time series of distributions at a given frequency as a functional analysis object, these authors design simple tests of this kind of trends at given unconditional quantiles, e.g. the 25th, 50th or 75th percentiles. This allows us to select the specification and the filtering of the variables in a bivariate Granger-causality regression where the dependent variable (selected factors at each quantile) and the CO_2 regressor should be stationary. Our combination of both statistical techniques allows us to extract factors at all relevant quantiles of the distribution of the change in temperatures to then test whether changes in CO_2 Granger cause these factors. Hence, the difference between both approaches is that the findings in Gadea and Gonzalo (2020) refer to a steeper change in temperatures in those locations where it is colder (e.g. in the Artic) whereas, by retrieving those factors which are common to all low temperatures around the world, our QFA approach implies that the growth (acceleration) rate in CO_2 emissions (concentrations) has a global effect on those temperatures.

In relation to the literature on this topic, our main contributions in this paper are threefold (i) we provide Monte Carlo evidence showing that QFA behaves much better than other standard procedures, such as PCA, in retrieving the right number of factors where the observational data at hand exhibits big outliers, as in the case of temperature changes; (ii) we propose a new method for implementing quantile Granger-causality tests based on using the QFA factors as dependent variables that complements the standard mean (PC factors) causality tests in the literature, and (iii) we find that changes in CO_2 emissions have higher predictive power at low than at high quantiles of the distribution of changes in temperature.

It is noteworthy that, in principle, the results in (iii) above are different from those reported by Gadea and Gonzalo (2020) about the the presence of upward sloping trends in the former unconditional quantiles but not in the latter. In effect, they find that lower temperatures in some specific areas of the world (like the Artic region) have grown at a higher pace than high temperatures. By contrast, our main finding is that the growth rate in CO_2 emissions has a higher predictive power during time periods where the temperature is decreasing (or increasing very little) but little or no effect during periods with large positive changes. In other words, higher growth/acceleration in CO_2 emissions/concentrations is bound to help to increase the temperature in periods where this goes down, without any discernible Granger-causality effect

on large positive changes. However, our finding of a non-uniform climate sensitivity to CO_2 go in line with the heterogeneous GW results documented by Gadea and Gonzalo (2020). The latter suggests that CO_2 has stronger warming effects in cold regions, whereas our current results indicate that periods with negative temperature increments (irrespective of the region) are more sensitive to the warming effects of CO_2. We conjecture that, one possibility behind this non-uniform climate sensitivity could be the fact that, only when the increments in temperature are high (upper quantiles of the conditional distribution of changes in temperature), then tropical cloud formation increases. This has two effects. On the one hand, clouds does not allow the arrival of short-wave sun's rays, which would lead to cooling. On the other hand, they do not allow the exit of the long-wave sun's rays that the earth emits when it warms up. Uncertainty arises because it not yet known which of the two effects is the dominating one (see Kamae et al. 2016). Hence, in the absence of panel data on tropical cloud formation over the period under consideration in this study, this phenomenon makes it difficult to detect the effects of the growth (acceleration) rate of CO_2 emissions (concentrations) on large (positive) changes in temperatures.

The rest of the paper is organized as follows. Section 2 defines QFM. In Section 3, we introduce the QFA estimator and its computational algorithm, establish the average rates of convergence of the quantile-dependent factors and factor loadings, propose a consistent selection criterion to choose the number of factors at each quantile, and finally run a Monte Carlo simulation results to highlight the advantages of using QFA instead of PCA in finite samples with big outliers. Section 4 considers an empirical application of causality analysis to estimate the effect of CO_2 emissions and concentrations on GW using a large panel dataset on the annual distributions of temperatures over the last century. Finally, Section 5 concludes.

2 Quantile Factor Model

To motivate our empirical analysis, this section reviews the basic concepts and tools underlying CDG's (2021) QFM approach.

Let $\{X_{it}\}$ be a panel of N observed variables (units), each with T observations. Then, X_{it}, with $i = 1, 2, .., N$ and $t = 1, 2, ..., T$, has the following QFM structure at some $\tau \in (0, 1)$:

$$\mathsf{Q}_{X_{it}}[\tau | f_t(\tau)] = \lambda_i'(\tau) f_t(\tau),$$

where the common factors $f_t(\tau)$ are gathered in a $r(\tau) \times 1$ vector of unobservable random variables, $\lambda_i(\tau)$ is a $r(\tau) \times 1$ vector of non-random factor loadings with $r(\tau) \ll N$. Note that in the QFM defined above, the factors, the loadings, and the number of factors are all allowed to be quantile-dependent.

Alternatively, the above equation implies that

$$X_{it} = \lambda_i'(\tau) f_t(\tau) + u_{it}(\tau), \tag{2}$$

where the quantile-dependent idiosyncratic error $u_{it}(\tau)$ is assumed to satisfy the following quantile restrictions:

$$P[u_{it}(\tau) \leq 0 | f_t(\tau)] = \tau.$$

As mentioned in the Introduction, LSS models provide nice illustrations of potential DGPs with the previous QFM representation. In particular, recall the example given above, i.e. $X_{it} = \alpha_i' f_{1t} + (\eta_i' f_{2t})\epsilon_{it}$, where $\{\epsilon_{it}\}$ are zero-mean i.i.d errors independent of $\{f_{1t}\}$ and $\{f_{2t}\}$, with cumulative distribution function (CDF) F_ϵ such that the median of ϵ_{it} is 0, i.e., $\mathsf{Q}_\epsilon(0.5) = 0$, $\alpha_i, f_{1t} \in \mathbb{R}^{r_1}$, $\eta_i, f_{2t} \in \mathbb{R}^{r_2}$, and $\eta_i' f_{2t} > 0$. Then, when f_{1t} and f_{2t} do not share common elements, this model has a QFM representation in form of (2) with $\lambda_i(\tau) = [\alpha_i', \eta_i' \mathsf{Q}_\epsilon(\tau)]'$, $f_t(\tau) = [f_{1t}', f_{2t}']$ for $\tau \neq 0.5$, and $\lambda_i(\tau) = \alpha_i$, $f_t(\tau) = f_{1t}$ for $\tau = 0.5$. Note that in such a case, the loadings are quantile-dependent objects while the factors are not.

Another example is provided by a LSS model where different factors affect the first, second and third moments of the data, i.e. $X_{it} = \alpha_i f_{1t} + f_{2t}\epsilon_{it} + c_i f_{3t}\epsilon_{it}^3$, where ϵ_{it} is a standard normal random variable whose CDF is denoted as $\Phi(\cdot)$. Let f_{2t}, f_{3t}, c_i be positive, then X_{it} has an equivalent representation in form of (2), with $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau), c_i \Phi^{-1}(\tau)^3]'$, $f_t(\tau) = (f_{1t}, f_{2t}, f_{3t})'$ for $\tau \neq 0.5$, and $\lambda_i(\tau) = \alpha_i$, $f_t(\tau) = f_{1t}$ for $\tau = 0.5$. In particular, if $c_i = 1$ for all i and noticing that the mapping $\tau \mapsto \Phi^{-1}(\tau)^3$ is strictly increasing, then we have for $\tau \neq 0.5$, $Q_{X_{it}}[\tau | f_t(\tau)] = \alpha_i f_{1t} + \Phi^{-1}(\tau) \cdot [f_{2t} + f_{3t}\Phi^{-1}(\tau)^2]$, so that there exists a QFM representation as in (2) with $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau)]'$ and $f_t(\tau) = [f_{1t}, f_{2t} + f_{3t}\Phi^{-1}(\tau)^2]'$ for $\tau \neq 0.5$. Notice that in this case, the second factor in $f_t(\tau)$, $f_{2t} + f_{3t}\Phi^{-1}(\tau)^2$, is quantile dependent even for $\tau \neq 0.5$.

Finally, note that applying PCA to the data in the two previous DGPs only yield consistent estimates of the factors shifting the means but will fail to capture those other extra factors shifting quantiles, other than the means, or their corresponding quantile-varying loadings. In the sequel, QFA is therefore proposed as a new estimation procedure of all the quantile-dependent objects in QFM.

3 QFA Estimators and their Asymptotic Properties

To simplify the notations, we suppress hereafter the dependence of $f_t(\tau), \lambda_i(\tau), r(\tau)$ and $u_{it}(\tau)$ on τ, so that the QFM in (2) is rewritten as:

$$X_{it} = \lambda_i' f_t + u_{it}, \quad P[u_{it} \leq 0 | f_t] = \tau, \tag{3}$$

where $\lambda_i, f_t \in \mathbb{R}^r$. Suppose that we have a sample of observations $\{X_{it}\}$ generated by (3) for $i = 1, \ldots, N$, and $t = 1, \ldots, T$, where the true values of $\{f_t\}$ are $\{f_{0t}\}$ and the true values of $\{\lambda_i\}$ are $\{\lambda_{0i}\}$. We take a fixed-effects approach by treating $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ as parameters to be estimated, and our asymptotic analysis is conditional on $\{f_{0t}\}$. In Section 3.1, we consider the estimation of $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ while r is assumed to be known, while the estimation of r at each quantile is discussed in Section 3.3.

3.1 Estimating Quantile Factors and Loadings

It is well known in the literature on factor models that $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ cannot be separately identified without imposing normalizations (see Bai and Ng 2002). Without loss of generality, we choose the following normalizations:

$$\frac{1}{T}\sum_{t=1}^{T} f_t f_t' = \mathbb{I}_r, \quad \frac{1}{N}\sum_{i=1}^{N} \lambda_i \lambda_i' \text{ is diagonal with non-increasing diagonal elements.} \tag{4}$$

Let $M = (N+T)r$, $\theta = (\lambda_1', \ldots, \lambda_N', f_1', \ldots, f_T')'$, and $\theta_0 = (\lambda_{01}', \ldots, \lambda_{0N}', f_{01}', \ldots, f_{0T}')'$ denotes the vector of true parameters, where we also suppress the dependence of θ and θ_0 on M to save notation. Let $\mathcal{A}, \mathcal{F} \subset \mathbb{R}^r$ and define:

$$\Theta^r = \left\{\theta \in \mathbb{R}^M : \lambda_i \in \mathcal{A}, f_t \in \mathcal{F} \text{ for all } i, t, \{\lambda_i\} \text{ and } \{f_t\} \text{ satisfy the normalizations in (4)}\right\}.$$

Further, define:

$$\mathbb{M}_{NT}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda_i' f_t),$$

where $\rho_\tau(u) = (\tau - \mathbf{1}\{u \leq 0\})u$ is the check function. The QFA estimator of θ_0 is defined as:

$$\hat{\theta} = (\hat{\lambda}_1', \ldots, \hat{\lambda}_N', \hat{f}_1', \ldots, \hat{f}_T')' = \arg\min_{\theta \in \Theta^r} \mathbb{M}_{NT}(\theta).$$

It is obvious that the way in which our estimator is related to the PCA estimator studied by Bai and Ng (2002) and Bai (2003) is analogous to how QR is related to standard least-squares regressions. However, unlike Bai (2003)'s PCA estimator, our estimator $\hat{\theta}$ does not yield an analytical closed form. This makes it difficult not only to find a computational algorithm that would yield the estimator, but also the analysis of its asymptotic properties. In the sequel, we briefly review CDG´s (2021) computational algorithm, called *iterative quantile regression* (IQR, hereafter), that can effectively find the stationary points of the object function. In parallel, CDG's (2021) Theorem 1 shows that $\hat{\theta}$ achieves the same convergence rate as the PCA estimators for AFM.

To describe the algorithm, let $\Lambda = (\lambda_1, \ldots, \lambda_N)'$, $F = (f_1, \ldots, f_T)'$, and define the following

averages:

$$\mathbb{M}_{i,T}(\lambda, F) = \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda' f_t) \quad \text{and} \quad \mathbb{M}_{t,N}(\Lambda, f) = \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(X_{it} - \lambda_i' f).$$

Note that we have $\mathbb{M}_{NT}(\theta) = N^{-1}\sum_{i=1}^{N} \mathbb{M}_{i,T}(\lambda_i, F) = T^{-1}\sum_{t=1}^{T} \mathbb{M}_{t,N}(\Lambda, f_t)$. The main difficulty in finding the global minimum of $\mathbb{M}_{NT}$ is that this object function is not convex in θ. However, for given F, $\mathbb{M}_{i,T}(\lambda, F)$ happens to be convex in λ for each i and likewise, for given Λ, $\mathbb{M}_{t,N}(\Lambda, f)$ is also convex in f for each t. Thus, both optimization problems can be efficiently solved by various linear programming methods (see Chapter 6 of Koenker 2005). Based on this observation, CDG (2021) propose the following iterative procedure:

Iterative quantile regression (IQR):
Step 1: Choose random starting parameters: $F^{(0)}$.
Step 2: Given $F^{(l-1)}$, solve $\lambda_i^{(l-1)} = \arg\min_\lambda \mathbb{M}_{i,T}(\lambda, F^{(l-1)})$ for $i = 1, \ldots, N$; given $\Lambda^{(l-1)}$, solve $f_t^{(l)} = \arg\min_f \mathbb{M}_{t,N}(\Lambda^{(l-1)}, f)$ for $t = 1, \ldots, T$.
Step 3: For $l = 1, \ldots, L$, iterate the second step until $\mathbb{M}_{NT}(\theta^{(L)})$ is close to $\mathbb{M}_{NT}(\theta^{(L-1)})$, where $\theta^{(l)} = (\text{vech}(\Lambda^{(l)})', \text{vech}(F^{(l)})')'$.
Step 4: Normalize $\Lambda^{(L)}$ and $F^{(L)}$ so that they satisfy the normalizations in (4).

To see the connection between the IQR algorithm and the PCA estimator of Bai (2003), suppose that $r = 1$, and replace the check function in the IQR algorithm by the least-squares loss function. Then, it is easy to show that the second step of the algorithm above yields $\Lambda^{(l-1)} = (X'F^{(l-1)})/\|F^{(l-1)}\|^2$ and $F^{(l)} = (X\Lambda^{(l-1)})/\|\Lambda^{(l-1)}\|^2 = XX'F^{(l-1)}/C_{l-1}$, where X is the $T \times N$ matrix with elements $\{X_{it}\}$, $C_l = \|F^{(l)}\|^2 \cdot \|\Lambda^{(l)}\|^2$, and $\|\cdot\|$ is the Frobenius norm of a matrix, i.e. the square root of the sum of squares of its elements. Thus, with proper normalizations at each step, the iterative procedure is equivalent to the well-known *power method* of Hotelling (1933), and the sequence $F^{(0)}, F^{(1)}, \ldots$ will converge to the eigenvector associated with the largest eigenvalue of XX'. In the more general case $r > 1$, by replacing the check function in the IQR algorithm by the least-squares loss function and normalize $F^{(l-1)}, \Lambda^{(l-1)}$ to satisfy (4) at step 2, it can be shown that this iterative procedure is similar to the method of *orthogonal iteration* (see Section 7.3.2 of Golub and Van Loan 2013) for calculating the eigenvectors associated with the r largest eigenvalues of XX', which is the PCA estimator of Bai (2003). Therefore, the IQR algorithm and its corresponding QFA estimator can be viewed as an extension of PCA to QFM.

Then, consistency of the QFA estimators can be proved under the following set of assumptions outlined in CDG (2021),

Assumption 1. *(i) $\mathcal{A}$ and $\mathcal{F}$ are compact sets and $\theta_0 \in \Theta^r$. In particular, $N^{-1}\sum_{i=1}^{N} \lambda_{0i}\lambda_{0i}' = diag(\sigma_{N1}, \ldots, \sigma_{Nr})$ with $\sigma_{N1} \geq \sigma_{N2} \cdots \geq \sigma_{Nr}$, and $\sigma_{Nj} \to \sigma_j$ as $N \to \infty$ for $j = 1, \ldots, r$ with $\infty > \sigma_1 > \sigma_2 \cdots > \sigma_r > 0$.*

(ii) The conditional density function of u_{it} given $\{f_{0t}\}$, denoted as f_{it}, is continuous, and satisfies that: for any compact set $C \subset \mathbb{R}$ and any $u \in C$, there exists a positive constant $\underline{\mathsf{f}} > 0$ (depending on C) such that $\mathsf{f}_{it}(u) \geq \underline{\mathsf{f}}$ for all i, t.
(iii) Given $\{f_{0t}, 1 \leq t \leq T\}$, $\{u_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$ is independent across i and t.

Assumptions 1 (i) is the standard *strong factors* assumption in the literature (see Assumption B of Bai 2003) which allows to order the factors. Assumptions 1 (ii) and (iii) are similar to (C1) and (C2) in Ando and Bai (2020), except that moments of u_{it} are not required to exist. Also notice that Assumption (iii), which allows for both cross-sectional and time series heteroskedasticity, requires the idiosyncratic errors to be mutually independent. This assumption looks restrictive in principle but Monte Carlo simulations in CDG (2021) show that it can be relaxed to allow for mild cross-sectional and serial dependence in the error terms.

Write $\hat{\Lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)'$, $\Lambda_0 = (\lambda_{01}, \ldots, \lambda_{0N})'$, $\hat{F} = (\hat{f}_1, \ldots, \hat{f}_T)'$, $F_0 = (f_{01}, \ldots, f_{0T})'$, and let $L_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Then, under Assumption 1, CDG (2021) derive the average rate of convergence of $\hat{\Lambda}$ and $\hat{F}$ as $N, T \to \infty$, which are given by:

$$\|\hat{\Lambda} - \Lambda_0\|/\sqrt{N} = O_P(1/L_{NT}) \quad \text{and} \quad \|\hat{F} - F_0\|/\sqrt{T} = O_P(1/L_{NT}).$$

3.2 Asymptotic Distribution of QFA Estimators

To obtain the asymptotic distributions of the QFA estimators, CDG (2021) address the difficulties raised by the non-smoothness of the indicator functions by proposing a similar estimator of θ_0, denoted as $\tilde{\theta}$, which relies on the following smoothed quantile regressions (SQR):

$$\tilde{\theta} = (\tilde{\lambda}_1', \ldots, \tilde{\lambda}_N', \tilde{f}_1', \ldots, \tilde{f}_T')' = \underset{\theta \in \Theta^r}{\arg\min}\, \mathbb{S}_{NT}(\theta),$$

where

$$\mathbb{S}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tau - K\left(\frac{X_{it} - \lambda_i' f_t}{h}\right)\right](X_{it} - \lambda_i' f_t),$$

$K(z) = 1 - \int_{-1}^{z} k(z)dz$, $k(z)$ is a continuous function with support $[-1, 1]$, and h is a bandwidth parameter that goes to 0 as N, T diverge.

Then, define

$$\Phi_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathsf{f}_{it}(0) f_{0t} f_{0t}' \quad \text{and} \quad \Psi_t = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{f}_{it}(0) \lambda_{0i} \lambda_{0i}'$$

for all i, t, and adopt the following assumption, also borrowed from CDG (2021):

Assumption 2. *Let $m \geq 8$ be a positive integer,*
(i) $\Phi_i > 0$ and $\Psi_t > 0$ for all i, t.

(ii) λ_{0i} *is an interior point of* $\mathcal{A}$ *and* f_{0t} *is an interior point of* $\mathcal{F}$ *for all* i, t.
(iii) $k(z)$ *is symmetric around* 0 *and twice continuously differentiable.* $\int_{-1}^{1} k(z)dz = 1$, $\int_{-1}^{1} z^j k(z)dz = 0$ *for* $j = 1, \ldots, m-1$ *and* $\int_{-1}^{1} z^m k(z)dz \neq 0$.
(iv) f_{it} *is* $m+2$ *times continuously differentiable. Let* $\mathsf{f}_{it}^{(j)}(u) = (\partial/\partial u)^j \mathsf{f}_{it}(u)$ *for* $j = 1, \ldots, m+2$. *For any compact set* $C \subset \mathbb{R}$ *and any* $u \in C$, *there exists* $-\infty < \underline{l} < \bar{l} < \infty$ *such that* $\underline{l} \leq \mathsf{f}_{it}^{(j)}(u) \leq \bar{l}$ *and* $\underline{l} \leq \mathsf{f}_{it}(u) \leq \bar{l}$ *for* $j = 1, \ldots, m+2$ *and for all* i, t.
(v) As $N, T \to \infty$, $N \propto T$, $h \propto T^{-c}$ *and* $m^{-1} < c < 1/6$.

The above conditions are standard in SQR, with the exception of (v). For example, this condition slightly differs from the one adopted by Galvao and Kato (2016) who assume that $m^{-1} < c < 1/3$ (or $m \geq 4$) because the incidental parameters (λ_i and f_t) in QFM enter the model interactively, while there are no interactive fixed-effects in the panel quantile models considered by these authors .

Then, under Assumptions 1 and 2, CDG (2021) show that:

$$\sqrt{T}(\tilde{\lambda}_i - \lambda_{0i}) \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\Phi_i^{-2}) \quad \text{and} \quad \sqrt{N}(\tilde{f}_t - f_{0t}) \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\Psi_t^{-1}\Sigma_\Lambda\Psi_t^{-1})$$

for each i and t, where $\Sigma_\Lambda = \text{diag}(\sigma_1, \ldots, \sigma_r)$.

3.3 Selecting the Number of Factors at Quantiles

While the number of quantile-dependent factors $r(\tau)$ has been so far assumed to be known at each τ, CDG (2021) propose a rank-minimization criterion to select the correct number of factors at each quantile with probability approaching one. This criterion selects the number of factors by rank minimization. Suppressing once again the dependence of the quantile-dependent objects on τ, including $r(\tau)$, to ease notation, the criterion works as follows.

Let k be a positive integer larger than r, and $\mathcal{A}^k$ and $\mathcal{F}^k$ be compact subsets of $\mathbb{R}^k$. In particular, let us assume that $[\lambda_{0i}' \quad \mathbf{0}_{1\times(k-r)}]' \in \mathcal{A}^k$ for all i.

Let $\lambda_i^k, f_t^k \in \mathbb{R}^k$ for all i, t and write $\theta^k = (\lambda_1^{k\prime}, \ldots, \lambda_N^{k\prime}, f_1^{k\prime}, \ldots, f_T^{k\prime})'$, $\Lambda^k = (\lambda_1^k, \ldots, \lambda_N^k)'$, $F^k = (f_1^k, \ldots, f_T^k)'$. Consider the normalizations for factors and loadings discussed above, define $\hat{\Lambda}^k = (\hat{\lambda}_1^k, \ldots, \hat{\lambda}_N^k)'$ and write

$$(\hat{\Lambda}^k)'\hat{\Lambda}^k/N = \text{diag}\left(\hat{\sigma}_{N,1}^k, \ldots, \hat{\sigma}_{N,k}^k\right).$$

The rank minimization criterion to estimate the number of factors r is defined as:

$$\hat{r}_{\text{rank}} = \sum_{j=1}^{k} \mathbf{1}\{\hat{\sigma}_{N,j}^k > P_{NT}\},$$

where P_{NT} is a sequence that goes to 0 as $N, T \to \infty$. In other words, $\hat{r}_{\text{rank}}$ is equal to the number of diagonal elements of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$ that are larger than the threshold P_{NT}. Note that it can can be interpreted as a rank estimator of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$ since this average converges to a matrix with rank r, where P_{NT} can be viewed as a cutoff value to choose the asymptotic rank of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$. In particular, CDG (2021) find that the choice

$$P_{NT} = \hat{\sigma}^k_{N,1} \cdot \left(\frac{1}{L^2_{NT}}\right)^{1/3}$$

works well as long as $\min\{N, T\}$ is 100. This is also the value used in our empirical application below.

Then, under Assumption 1, CDG (2021) show that:

$P[\hat{r}_{\text{rank}} = r] \to 1$ as $N, T \to \infty$ if $k > r$, $P_{NT} \to 0$ and $P_{NT} L^2_{NT} \to \infty$.

3.4 Relative Performance of PCA and QFA in a DGP with Outliers

As pointed out in the discussion of Assumption 1 above, the consistency of the QFA estimator does not require the moments of the idiosyncratic errors to exist. Thus, at $\tau = 0.5$, QFA can be viewed as a robust QR alternative to the PCA estimators commonly used in practice. By the same token, our estimator of the number of factors should also be more robust to outliers and heavy tails than the IC-based method of Bai and Ng (2002). In this subsection we confirm these two claims by means of Monte Carlo simulations.

In particular, the following DGP is considered:

$$X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it},$$

where $f_{1t} = 0.8 f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t-1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt}$ are all independent draws from $\mathcal{N}(0,1)$, and $u_{it} \sim i.i.d\ B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$, where B_{it} are i.i.d Bernoulli random variables with means equal to 0.98 and Cauchy$(0,1)$ denotes the standard Cauchy distribution. In this way, approximately 2% of the idiosyncratic errors are generated as outliers.

We consider four estimators of the number of factors r: two estimators based on PC_{p1}, IC_{p1} of Bai and Ng (2002), the Eigenvalue Ratio (ER) estimator proposed by Ahn and Horenstein (2013) and CDG's (2021) rank-minimization estimator with P_{NT} chosen as in section 3.3. We set $k = 8$ for all four estimators, and consider $N, T \in \{50, 100, 200, 500\}$.

Table 1 reports the following fractions:

$$[\text{proportion of } \hat{r} < 3 \ , \ \text{proportion of } \hat{r} = 3 \ , \ \text{proportion of } \hat{r} > 3 \ ]$$

for each estimator having run 1000 replications.

It becomes evident from Table 1 that PC_{p1} and IC_{p1} almost always overestimate the number factors, while the ER estimator tends to underestimate them, though to a lesser extent than what PC_{p1} and IC_{p1} overestimate them. By contrast, the rank-minimization estimator selects accurately the right number of factors as long as $\min\{N, T\} \geq 100$.

Next, to compare the PCA and QFA estimators of the common factors in the previous DGP, we assume that $r = 3$ is known. We first get the PCA estimator (denoted as $\hat{F}_{PCA}$), and then obtain the QFA estimator at $\tau = 0.5$ (denoted $\hat{F}^{0.5}_{QFA}$) using the IQR algorithm. Next, we regress each of the true factors on $\hat{F}_{PCA}$ and $\hat{F}^{0.5}_{QFA}$ separately, and report the average (adjusted) R^2 from 1000 replications in Table 2 as an indicator of how well the space of the true factors is spanned by the estimated factors. As shown in the first three columns of Table 2, while the PCA estimators are not very successful in capturing the true common factors, the QFA estimators approximate them very satisfactorily, even when N, T are not too large.

Thus, this simulation exercise provides strong evidence about the substantial gains that can be achieved by using QFA rather than PCA in those cases where the idiosyncratic error terms in AFM exhibit heavy tails and outliers.

4 Climate Change and CO_2 Emissions

As stressed in the Introduction, QFA provides a useful tool for causal analysis regarding the effects of CO_2 on GW. Specifically, we focus on applying our proposed QFA methodology to investigate how CO_2 emissions affect changes in temperatures. Our dataset (coined Climate for short) includes annual changes in temperature from 441 stations from 1917 to 2018 ($N = 441, T = 102$), drawn from the Climate Research Unit at the University of East Anglia, where information about global temperatures across different stations in the Northern and Southern Hemisphere is collected. Annual global CO_2 emissions and concentration (akin to cumulative emissions but excluding natural sinks) are downloaded from Our World in Data.

4.1 Testing for Trends in Climate Data

Gadea and Gonzalo (2020) provide a novel methodology to test for the existence of trends in the unconditional distributional characteristics (moments, quantiles, etc.) of global temperatures. Treating temperatures as a functional stochastic process, their distributional characteristics

can be thought of as time series objects to which one could apply existing methods in time series literature. For example, the proposed robust linear trend-test is based in the statistical significance of the β coefficient in the following least-squares regression:

$$C_t = \alpha + \beta t + u_t, t = 1, ...T, \tag{5}$$

where C_t denotes the distributional characteristic of interest (e.g. a given quantile).The asymptotic properties of the OLS estimator in the previous equation depend on the summability order of the unknown trend component, $\delta \geq 0$, defined as follows. Let $C_t = h(t) + v_t$, where v_t is an $I(0)$ process and $h(t)$ is the unknown trend polynomial process of order k, then its summability order becomes $S_T = \frac{1}{T^{1+\delta}} \sum_{t=1}^{T} h(t)$ where δ is chosen such that S_T is $O(1)$, and so $\delta = k$. Thus, the OLS estimated coefficient in (5) is given by

$$\hat{\beta} = \frac{\sum tC_t - T\bar{t}\bar{C}}{\sum t^2 - T\bar{t}^2}, \tag{6}$$

where $\sum tC_t = T^{\frac{5}{2}+0.5\delta} \frac{1}{T} \sum (\frac{t}{T})(\frac{C_t}{T^{0.5(1+\delta)}})$ and $\sum t^2 = T^3 \frac{1}{T} \sum (\frac{t}{T})^2$, so that

$$T^{0.5(1-k)} \hat{\beta} = O_p(1), \tag{7}$$

implying consistency if $\delta = k = 0$. However, if C_t is regressed on a polynomial of order s it holds that

$$\hat{\beta} \approx \frac{\frac{1}{\sqrt{T}} \sum (\frac{c_t}{T^{0.5(1+k)}})(\frac{t}{T})^s}{\frac{1}{T} \sum (\frac{t}{T})^{2s}} \frac{T^{1+s+0.5k}}{T^{1+2s}}. \tag{8}$$

Then it is easy to show that

$$T^{s-0.5k} \hat{\beta} = O_p(1). \tag{9}$$

Thus, a necessary condition for consistency in the estimation of the trend slope is $k < 2s$. As a result, if $k = 1$, then $s \geq 1$, and if $k = 0$, $s \geq 1$. In such cases, it is easy to show that $t_{\beta=0} \to N(0, 1)$.

Gadea and Gonzalo (2020) apply the trend-test to two different data sets: i) daily temperatures in Central England for the period 1772-2017, and ii) global temperatures across different stations in the Northern and Southern Hemispheres for the period 1880-2015. The latter dataset in the one more closely related to ours though they only have have available 241 stations. Since the number of units needs to be larger for QFA to be valid, our dataset is shorter (1917-2018) in exchange for almost doubling the number of stations (441). At any rate, both applications in

Gadea and Gonzalo (2020) lead to similar results. First, a linear trend component is detected in most of the distributional characteristics, suggesting the existence of GW. Second, the dynamics of GW is heterogeneous across the different unconditional quantiles: the slope of the trend in the lower quantiles is steeper than those in the mean, median, and upper quantiles.

In line with this approach, we first use the above-mentioned tests to study the statistical properties of temperatures in our dataset. The evidence reported in Table 3 suggests the presence of linear trends in the levels of the temperature data, consistent with the findings in Gadea and Gonzalo (2020). Additionally, we apply a standard ADF test to each individual station and each distributional characteristic, including a constant and a linear trend component in the test equation. The results in Table 4 show that the null of unit root is widely rejected (with the exception of three stations) for almost all distributional characteristics (except for max, q75, and q99). Based on this evidence it is possible to assert that the temperature levels are trend-stationary process or, equivalently, that their annual changes are integrated of order zero, I(0), with a drift. Thus, first differences (rather than the levels) of the temperatures are taken in the QFA application to achieve stationarity which recall is a key requirement to extract both PCA and QFA factors.

4.2 Quantile Factor Analysis

QFA is used to estimate the quantile factors for the changes of the temperature data in its standardized format. Note that the lower quantiles capture large negative changes in temperature (the smallest values) while the upper quantiles refer to the larger positive changes (the highest values). The number of factors are selected according to the rank-minimization criterion explained in section 3.3 for a fine grid of quantile levels, τ, ranging from 0.01 to 0.99. As pointed out in the previous section, the number of factors varies across quantiles in the sense that it decreases as we move away from the median. In particular, the number of factors are: 1 (at $\tau =$ 0.01, 0.05, 0.95, and 0.99), 2 (at $\tau = 0.1$ and 0.9), 4 (at $\tau = 0.25$ and 0.75), and 6 (at $\tau = 0.5$). For illustrative purposes, in Figure 2 the estimated factors for the quantiles 0.05 and 0.95 are plotted to shown that they are fairly different. In addition, PCA is used to estimate the factors at the mean, with the number of factors selected according to the PC_{p1} criterion of Bai and Ng (2002). The PC_{p1} selects 8 factors, which is the maximum number imposed in the algorithm.

To compare the QFA factors (denoted as $\hat{F}^{\tau}_{QFA}$) and the PCA factors (denoted as $\hat{F}_{PCA}$), we regress each element of $\hat{F}^{\tau}_{QFA}$ on the 8 $\hat{F}_{PCA}$ and compute the R^2 in these regressions as a measure of correlation.[8] The results are shown in the upper panel of Table 6. It becomes clear that for the quantiles at the center of the distribution ($\tau = 0.25$, 0.5, and 0.75) the estimated factors are highly correlated with the PCA factors, with all the R^2s above 0.90. By contrast, the QFA factors at the upper and lower quantiles (e.g. $\tau = 0.01$, 0.05, 0.95, and 0.99) exhibit

[8] We choose the number of PCA factors estimated by PC_{p1} in these regressions to play conservative.

much lower correlations with the PCA factors, with R^2s around 0.6. Thus, there seems to be room for using QFA in this application.

4.3 Global warming Granger-causality

Bivariate tests are implemented in this section in order to establish Granger causality from CO_2 emissions to the QFA estimated factors from the changes in temperature. We consider two different but related time series of CO_2 emissions: the annual global CO_2 emissions (a flow variable) and the global CO_2 concentration levels in the atmosphere (a stock variable). As seen in Figure 3, logged CO_2 emissions is not an I(0) variable and exhibits an increasing pattern that seems to be steeper after 1950. The growth rate of CO_2 emissions, in turn, has a fairly constant drift, though its variance looks higher during 1917-1958. Regarding CO_2 concentrations, Figure 4 shows that, as expected, both the levels and first differences of its (logged) series are not I(0), while the second difference seems to behave similarly to the growth rate of emissions.[9] In line with the visual analysis, the ADF-tests previously reported in Table 5 reject the presence of a unit root both in growth rate of CO_2 emissions and the acceleration rate of CO_2 concentrations.

Since it is conventional to use Granger-causality tests for variables that are $I(0)$, the QFA factors at each relevant quantile are located in the left-hand side of the equation, while in the right hand side includes p own lags of the QFA factors and q lags of either the growth rate of CO_2 emissions ($\Delta log(CO2)_{emi}$) or the second difference of the logged CO_2 concentrations, ($\Delta^2 log(CO2)_{cum}$). The specific equations that we estimate are given by:

$$\hat{F}^{\tau}_{QFA,t} = \alpha(\tau) + \sum_{i=1}^{p} \beta_i \times \hat{F}^{\tau}_{QFA,t-i} + \sum_{j=1}^{q} \gamma_j \times \Delta log(CO_2)_{emi,t-j} + u_t, \tag{10}$$

or

$$\hat{F}^{\tau}_{QFA,t} = \alpha(\tau) + \sum_{i=1}^{p} \beta_i \times \hat{F}^{\tau}_{QFA,t-i} + \sum_{j=1}^{q} \gamma_j \times \Delta^2 log(CO_2)_{cum,t-j} + u_t, \tag{11}$$

where the lag lengths p and q in (10) and (11) have been selected according to the AIC criterion.

Note that that similar arguments to those used in Bai and Ng (2006) (Theorem 2) can be used to justify replacing the true quantile factors by the QFA estimated ones in (10) and (11). As discussed by these authors, when applying PCA to estimate an AFM, the condition $\sqrt{T}/N \to 0$ is required for the above substitution to be valid. Given the rate of convergence of the QFA factors in section 3.1, a similar condition holds here, which is easily verified in our finite sample since $\sqrt{102}/441 = 0.023$. Thus, the test looks at the joint significance of the γ_j coefficients,

[9] The correlation coefficient between the first difference of the logged CO_2 emissions and the second difference of the logged CO_2 concentrations is around 0.98.

$j = 1, ..., q$, by means of an F-statistic.

Table 7 reports the p-values of the Granger-causality tests when the growth rate of CO_2 emissions and the acceleration rate of CO_2 concentrations are considered as regressors. Our findings indicate that the growth rate of CO_2 emissions strongly Granger causes the QFA factors at the lower quantiles ($\tau = 0.01, 0.05$), with p-values below 0.001. Quantile factors closely related to the mean according to Table 6 (τ from 0.1 to 0.9) are also Granger caused by the growth rate of CO_2 emissions in line with a broad literature using standard Granger-causality methods with climate and CO_2 data. (see e.g. Stips (2016)). As expected, the results look similar when using the second differences of the CO_2 concentrations since concentration is a stock while emission is a flow. A strong finding using both series of CO_2 is that Granger non-causality is not rejected at the extreme upper quantiles, that is, for $\tau = 0.95$ and 0.99.

To check the robustness of these findings, we make a comparison between the results obtained for the median ($\tau = 0.5$) and the 8 PCA mean factors selected with the PC_{p1} criterion. As noted in Section 4.2, both sets of factors are highly correlated. Additionally, we test Granger causality from the CO_2 to the first difference of the mean curve as it is commonly estimated in the literature. Table 8 shows the p-values for these Granger-causality tests. In agreement with the results for the median, Granger-causality is detected for the first difference of the mean curve, as well as for some of the PCA factors.

From the physical science of climate change, CO_2 emissions and concentrations lead to GW. The previous results for the lower quantiles of the conditional distribution of temperature changes illustrate the non-uniformity in the causes of GW. From the climatology point of view, we have not found yet clear strong reasons behind this lack of uniform climate sensitivity. One possibility is that when the increments in temperature are high (upper quantiles of the conditional increment temperature distribution) the probability of clouds formation increases and it is well known (see Kamae et al. 2016) that clouds augment the uncertainty around climate sensitivity due to the different feedback effects on temperature. This would make it difficult to detect Granger-causality in the upper part of the conditional distribution of those changes. Unfortunately, the literature on this topic has made use of a a few historical tropical cloud formations which prevents incorporating them into our analysis.

Finally, in spite of addressing a different research question related to GW, our results seem to be in line with the evidence reported in Gadea and Gonzalo (2020) where GW is also found to be non-uniform: lower temperatures increase much more than the medium and higher ones. The lower unconditional quantiles in their study correspond to the Arctic region. However, an increase of CO_2 concentrations will have unforeseen consequences (that is, whatever happens in the Poles does not remain there): ice melting, sea level increases, floods, migrations, extreme events, etc. All this is increased even more by its own feedback effects caused by the reduction in the surface albedo (less solar energy is reflected out to space) and by the release of more

greenhouse gasses (CO_2, and Methane) from the permafrost melting. In this respect, we highlight that non-uniform climate sensitivity is not regionally located but affects all the regions around the Globe. In particular, the growth rate of CO_2 emissions affects (positively) the periods where the temperature decreases or does not increase much. To mitigate this warming problem, CO_2 emissions should be reduced until a a situation of net-zero emission is reached. In order to be efficient, mitigation policies should take the asymmetries documented in this paper into account.

5 Conclusions

In this paper we test for Granger-causality of CO_2 emissions/concentration on annual changes in temperature from 441 weather stations over the period 1917-2018 in the Northern and Southern hemispheres. Using the QFA methodology proposed by CDG (2021) to retrieve common factors and their number from those climate changes at different quantiles, we apply Granger-causality tests of different CO_2 measures on those factors. The specification of the corresponding dynamic predictive equations is helped by the methodology proposed by Gadea and Gonzalo (2020) to detect deterministic and stochastic trends in different moments/quantiles of the distribution of changes in temperature. We further show that QFA is a much more robust estimation method that PCR in the presence of outliers, as is the case in our sample.

Our main finding is that CO_2 Granger causes factors at the lower quantiles of the distribution of changes in temperature (i.e. the more negative changes) much more strongly than those at the middle and upper quantiles. We stress that this novel result is not captured by the use the PCA factors (or the mean) since they capture common features of all temperatures whereas QFA factors capture common features at each quantile. Further research should go in the direction of discovering the climatology reasons for this asymmetric climate sensitivity. As pointed out earlier, one potential reason could be the different cloud feedback effects on temperatures which are only present when changes in the latter are large and positive.

A Tables

Table 1: AFM with Outliers in the Idiosyncratic Errors: Estimating the Number of Factors

N	T	PC_{p1} of BN			IC_{p1} of BN			Eigenvalue Ratio			Rank Estimator		
50	50	[0.00	0.04	0.96]	[0.00	0.14	0.86]	[0.26	0.30	0.44]	[0.47	0.53	0.00]
50	100	[0.00	0.02	0.98]	[0.00	0.05	0.95]	[0.33	0.19	0.48]	[0.40	0.60	0.00]
50	200	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.41	0.12	0.47]	[0.33	0.67	0.00]
50	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.56	0.07	0.37]	[0.29	0.71	0.00]
100	50	[0.00	0.02	0.98]	[0.00	0.05	0.95]	[0.34	0.18	0.48]	[0.39	0.61	0.00]
100	100	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.41	0.13	0.46]	[0.10	0.90	0.00]
100	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.48	0.07	0.45]	[0.06	0.94	0.00]
100	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.65	0.05	0.30]	[0.02	0.98	0.00]
200	50	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.45	0.10	0.45]	[0.37	0.63	0.00]
200	100	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.48	0.08	0.44]	[0.10	0.90	0.00]
200	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.63	0.06	0.31]	[0.00	1.00	0.00]
200	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.76	0.08	0.16]	[0.00	1.00	0.00]
500	50	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.57	0.08	0.35]	[0.36	0.64	0.00]
500	100	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.68	0.06	0.26]	[0.05	0.95	0.00]
500	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.76	0.08	0.16]	[0.00	1.00	0.00]
500	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.80	0.10	0.10]	[0.00	1.00	0.00]

Note: The DGP considered in this Table is: $X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it}$, where $f_{1t} = 0.8 f_{1,t\ 1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t\ 1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t\ 1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt} \sim i.i.d\, \mathcal{N}(0,1)$, $u_{it} \sim i.i.d\, B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$ where $B_{it} \sim i.i.d$ Bernoulli(0.98). For each estimation method, the [proportion of $\hat{r} < 3$, proportion of $\hat{r} = 3$, proportion of $\hat{r} > 3$] is reported from 1000 replications.

Table 2: AFM with Outliers in the Idiosyncratic Errors: Estimation of the Factors

		Regress F on $\hat{F}_{PCA}$			Regress F on $\hat{F}^{0.5}_{QFA}$		
N	T	f_1	f_2	f_3	f_1	f_2	f_3
50	50	0.939	0.810	0.686	0.987	0.975	0.968
50	100	0.931	0.718	0.578	0.987	0.975	0.968
50	200	0.890	0.589	0.412	0.987	0.975	0.968
50	500	0.807	0.405	0.252	0.988	0.975	0.968
100	50	0.928	0.738	0.595	0.993	0.986	0.984
100	100	0.921	0.630	0.441	0.994	0.988	0.984
100	200	0.857	0.479	0.285	0.994	0.988	0.985
100	500	0.713	0.294	0.138	0.994	0.988	0.984
200	50	0.890	0.657	0.513	0.997	0.994	0.992
200	100	0.858	0.514	0.333	0.997	0.994	0.993
200	200	0.779	0.358	0.178	0.997	0.994	0.992
200	500	0.530	0.131	0.051	0.997	0.994	0.992
500	50	0.819	0.501	0.371	0.998	0.997	0.996
500	100	0.725	0.327	0.196	0.999	0.998	0.997
500	200	0.546	0.165	0.062	0.999	0.998	0.997
500	500	0.273	0.036	0.018	0.999	0.998	0.997

Note: The DGP considered in this Table is: $X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it}$, where $f_{1t} = 0.8 f_{1,t\ \ 1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t\ \ 1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t\ \ 1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt} \sim i.i.d\ \mathcal{N}(0,1)$, $u_{it} \sim i.i.d\ B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$ where $B_{it} \sim i.i.d$ Bernoulli(0.98). For each estimation method, we report the average R^2 in the regression of (each of) the true factors on the estimated factors by PCA and QFA (assuming the number of factors to be known).

Table 3: Gadea-Gonzalo Trend test (1916-2018).

Characteristic	Test-statistic	p-value
mean	0.0134	0.0000
sd	-0.0026	0.0000
min	0.0676	0.0027
max	0.0184	0.0000
iqr	0.0004	0.5970
kur	-0.0026	0.0156
skw	0.0010	0.0001
q01	0.0222	0.0000
q05	0.0194	0.0000
q10	0.0136	0.0000
q25	0.0125	0.0000
q50	0.0143	0.0000
q75	0.0129	0.0000
q90	0.0148	0.0000
q95	0.0106	0.0000
q99	0.0108	0.1030

Note: Annual distributional characteristics are estimated using the cross-sectional distribution at each year (1916-2018). OLS estimates and HAC $t_{\beta=0}$ standard errors from regression $C_t = \alpha + \beta t + u_t$.

Table 4: ADF unit root tests.

ADF test by stations		
Percentage of rejections	99.32 %	
Number of rejections	3	
ADF test by characteristics		
Characteristics	Test-statistic	p-value
mean	-5.5723	0.0001
sd	-8.6852	0.0000
min	-3.6772	0.0285
max	-2.5997	0.2815
iqr	-9.1764	0.0000
kur	-8.4206	0.0000
skw	-9.6469	0.0000
q01	-5.2689	0.0020
q05	-8.3412	0.0000
q10	-7.1825	0.0000
q25	-7.2190	0.0000
q50	-6.5730	0.0000
q75	-2.1098	0.5338
q90	-8.3862	0.0000
q95	-7.4222	0.0000
q99	-3.0563	0.1227

Note: Annual distributional characteristics estimated using the cross-sectional distribution at each year (1916-2018). ADF-test equations include intercept and trend. Lag-selection conducted using SBIC criterion.

Table 5: ADF unit root tests for CO2 emissions/concentrations time series.

Transformation	Test-statistic	p-value
Logged CO2 emissions:		
Levels	-1.898	0.6485
First differences	-10.231	0.0000
Logged CO2 concentration:		
Levels	-2.089	0.5454
First differences	-2.163	0.5042
Second differences	-10.615	0.0000

Note: ADF-test equations include intercept and trend. Lag-selection conducted using SBIC criterion.

Table 6: Comparison of $\hat{F}_{QFA}$ and $\hat{F}_{PCA}$

		Elements of $\hat{F}^{\tau}_{QFA}$					
Dataset	τ	1	2	3	4	5	6
Climate	0.01	0.599					
	0.05	0.623					
	0.10	0.759	0.848				
	0.25	0.939	0.961	0.965	0.941		
	0.50	0.995	0.995	0.992	0.988	0.980	0.970
	0.75	0.950	0.961	0.966	0.933		
	0.90	0.755	0.905				
	0.95	0.629					
	0.99	0.567					

Note: This Table reports the R^2 of regressing each element of $\hat{F}_{QFA}$ on $\hat{F}_{PCA}$. For $\hat{F}_{QFA}$, the numbers of estimated factors is obtained using the rank-minimization criterion while, for $\hat{F}_{PCA}$, the numbers of estimated factors are 8 for all datasets.

Table 7: P-values of Granger non-causality tests for the QFA factors

		Elements of $\hat{F}^{\tau}_{QFA}$					
Regressor	τ	1	2	3	4	5	6
	0.01	**0.000**					
	0.05	**0.000**					
	0.10	0.199	**0.008**				
	0.25	0.531	**0.052**	**0.056**	0.967		
$\Delta \log(CO_2)_{emi}$	0.50	0.819	**0.000**	0.124	0.120	0.694	**0.012**
	0.75	0.737	**0.021**	**0.004**	0.542		
	0.90	0.478	**0.010**				
	0.95	0.888					
	0.99	0.137					
	0.01	**0.000**					
	0.05	**0.000**					
	0.10	**0.000**	**0.013**				
	0.25	0.438	**0.009**	0.101	0.809		
$\Delta^2 \log(CO_2)_{cum}$	0.50	0.663	**0.000**	**0.002**	**0.081**	0.459	0.115
	0.75	**0.000**	**0.024**	**0.003**	0.639		
	0.90	0.466	0.286				
	0.95	0.804					
	0.99	0.578					

Note: This Table reports the p-values of Granger non-causality tests, where each of the QFA factors is regressed on their own lags and the lags of the corresponding regressor. The lag lengths are chosen according to AIC setting 12 as the maximum number of lags.

Table 8: P-values of Granger non-causality tests using the mean curve estimate and PCA factors

	Regressor	
Elements of $\hat{F}_{PCA}$	$\Delta \log(\mathrm{CO}_2)_{emi}$	$\Delta^2 \log(\mathrm{CO}_2)_{cum}$
1	0.948	0.897
2	**0.038**	**0.012**
3	0.290	0.394
4	0.526	0.548
5	0.362	0.226
6	**0.068**	0.552
7	0.213	0.225
8	0.796	**0.044**
Δ Mean	**0.007**	**0.013**

Note: This Table reports the p-values of Granger non-causality tests, where each of the PCA factors and the dirst difference of the mean curve are regressed on their own lags and the lags of the corresponding regressor. The lag lengths are chosen according to AIC setting 12 as the maximum number of lags.

B Figures

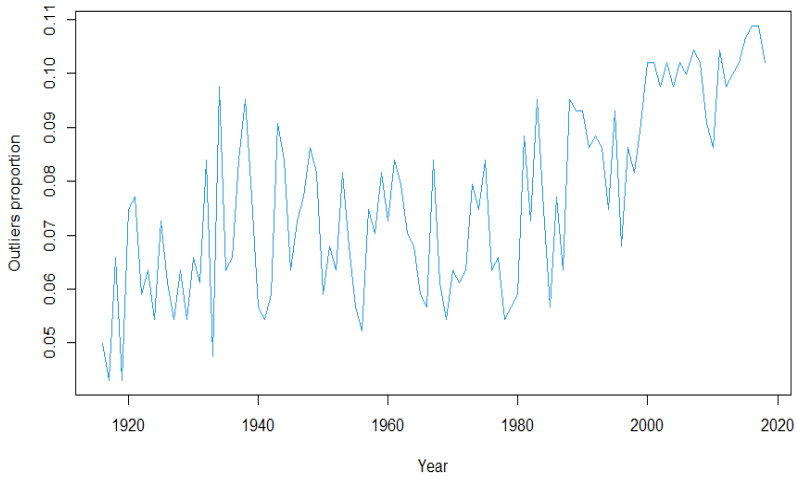

Figure 1: Proportion of outliers in temperature changes

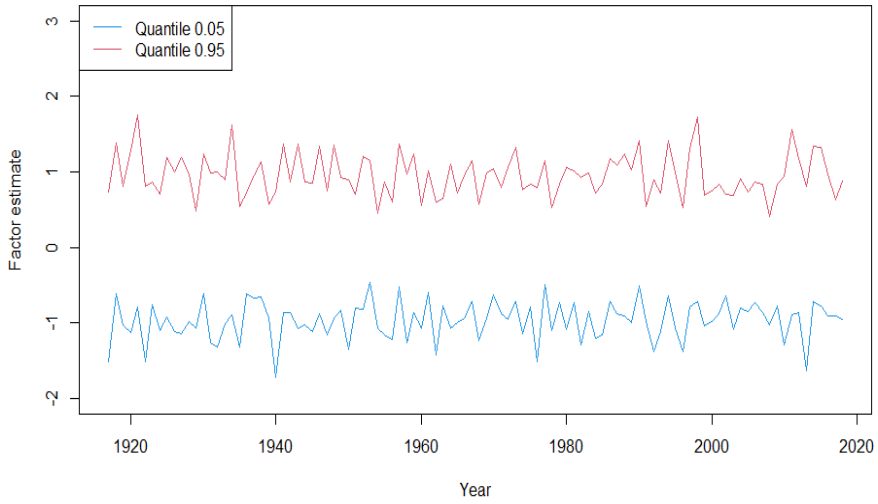

Figure 2: Estimated quantile factors at q25 and q75.

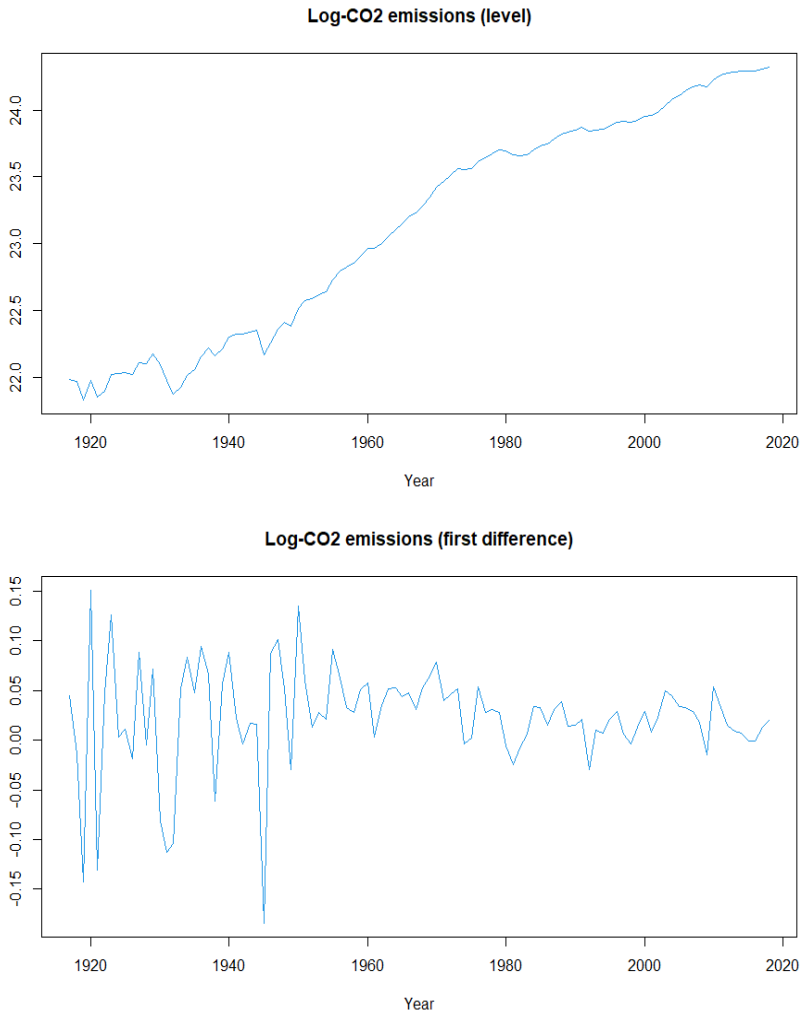

Figure 3: CO2 emissions in levels and first differences.

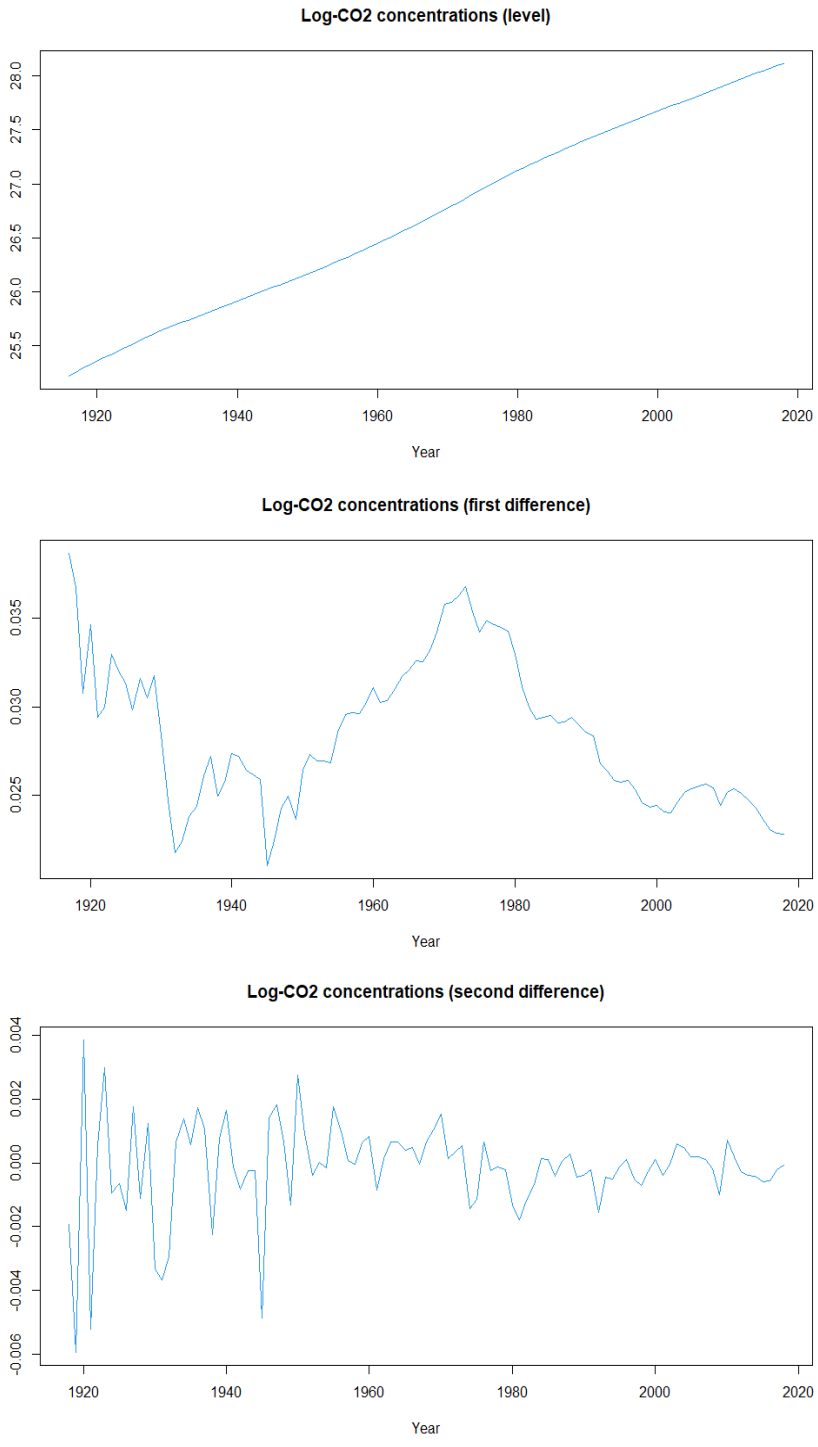

Figure 4: CO2 concentrations in levels, first differences and second differences.

References

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Ando, T. and J. Bai (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association 115*(529), 266–279.

Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine and Journal of Science 41*(5), 237–276.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*(1), 685–725.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*(4), 1229–1279.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008). *Large dimensional factor analysis*. Now Publishers Inc.

Castle, J. and D. Hendry (2020). Climate econometrics: An overview. *Foundations and Trends in Econometrics 10*(3-4).

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–304.

Chen, L., J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Chen, M., I. Fernández-Val, and M. Weidner (2018). Nonlinear factor models for network and panel data. *Working paper, UCL*.

Foote, E. (1856). Circumstances affecting the heat of the sun's rays. *The American Journal of Science and Arts 22*(2), 382–383.

Fourier, J. (1824). On the temperatures of the terrestrial sphere and interplanetary space. *Chimie et de Physique 27*, 136–167.

Gadea, M. D. and J. Gonzalo (2020). Trends in distributional characteristics: Existence of global warming. *Journal of Econometrics 214*(1), 153–174.

Galvao, A. F. and K. Kato (2016). Smoothed quantile regression for panel data. *Journal of Econometrics 193*(1), 92–112.

Golub, G. H. and C. F. Van Loan (2013). *Matrix Computations*, Volume 3. JHU Press.

Hansen, J., D. Johnson, A. Lacis, S. Lebedeff, P. Lee, D. Rind, and G. Russell (1981). Climate impact of increasing atmospheric carbon dioxide. *Science 213*(4511), 957–966.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology 24*(6), 417.

Hsiang, S. and R. E. Kopp (2018). An economist's guide to climate change science. *Journal of Economic Perspectives 32*(4), 3–32.

Huber, P. (1981). *Robust Statistics*. John Wiley and Sons.

Kamae, Y., T. Ogura, H. Shiogama, and M. Watanabe (2016). Recent progress toward reducing the uncertainty in tropical low cloud feedback and climate sensitivity: a review. *Geoscience Letters 3*(17).

Koenker, R. (2005). *Quantile Regression*. Number 38. Cambridge University Press.

Montamat, G. and J. Stock (2020). Quasi-experimental estimates of the transient climate response using observational data. *Climatic Change 160*, 361–371.

Myhre, G., D. Shindell, F. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang (2013). Anthropogenic and natural radiative forcing. In T. Stocker, D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Eds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 659–740. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*(4), 967–1012.

Stips, A., D. Macias, C. Coughlan, E. Garcia-Gorriz, and X. Liang (2016). On the causal structure between co2 and global temperature. *Scientific Reports 6*(21691).

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting 1*, 35–59.

Tyndall, J. (1863). On radiation through the earth's atmosphere. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 25*(167), 200–206.