Persistent Cycles and Long-run Covariability in Paleoclimate Time Series

Vasco J. Gabriel, University of Victoria, Canada

Luis F. Martins, ISCTE - Instituto Universitário de Lisboa, Portugal and CIMS, UK

Anthoulla Phella, University of Glasgow, UK

This version: November 2023

Abstract

Motivated by the presence of a strong cyclical component in paleoclimate data, this paper considers the problem of conducting cointegration inference when the data contains very large and persistent cycles. Our first contribution is to show, analytically and through Monte Carlos simulations, that while point estimation remains consistent, commonly applied tests no longer are valid when the data contains persistent cycles. Our second contribution is empirical: we propose the use of the long-run covariability approach of Müller & Watson (2018) to quantify low-frequency comovement amongst a range of paleoclimate times series. These methods allow us to focus on the long run properties of the data, bypassing short and medium run fluctuations, while being agnostic regarding the order of integration of the time series. We provide new estimates for the long-run relationship between temperatures and CO_2, concluding that in the long-run a 100 ppm increase in CO_2 levels would raise temperatures by around $1°C$. Finally, we illustrate how joint modelling of this set of paleoclimate time series can be carried out by factor analysis and how long-term projections about temperature increases and ice-sheet retreat can be constructed.

Keywords: Paleoclimate data; Glaciar cycles; Equilibrium climate sensitivity; Low frequency analysis.

JEL Classification: C22, C53, Q54.

1 Introduction

The analysis of paleoclimate data has played an important role in informing the debate around the causes of climate change and in quantifying equilibrium climate sensitivity (i.e., the long-term global temperature change in response to increases in atmospheric CO2 concentrations from pre-industrial levels). Indeed, the correct determination of the time series properties of paleoclimate data is crucial to ascertain the long term drivers of Earth's climate and, through appropriate (formal) statistical procedures, to establish a baseline for its evolution, therefore providing a useful contrast with recent anthropogenic-driven climate change (see Covey *et al.*, 1996, Lea, 2004, Knutti *et al.*, 2017, *inter alia*). In this regard, we argue that Milankovitch cycles (i.e., the role played by Earth's orbital movements, which affects solar radiation), while important in explaining the complex dynamics of glacial/interglacial cycles, may present non-negligible challenges for the study of the long-run features of climate data.

A strand of the literature has focused on the application of cointegrated vector autoregressions, making use of inherent error-correction mechanisms to model long-run relationships between climate and orbital geometry; for example, Kaufmann & Juselius (2013) and Kaufmann & Pretis (2020, 2021) find support for a long-run equilibrium driven by solar insolation, but disturbed by interactions among components of the climate system (see also Miller, 2019 for measurement issues associated with paleoclimate data). However, evidence for the presence of stochastic trends in the relevant time series is mixed at best. Several papers report results that are consistent with the presence of a stochastic trend (Gordon, 1991, Woodward & Gray, 1993, Woodward & Gray, 1995, Gordon *et al.*, 1996, Kärner, 1996). Conversely, many other studies suggest that the data is consistent with the presence of a deterministic trend with possibly highly persistent noise (Bloomfield & Nychka, 1992, Bloomfield, 1992, Baillie & Chung, 2002, Fomby & Vogelsang, 2002). The results in Davidson *et al.* (2016), for example, do not support the hypothesis of integrated behaviour.

The consequences of uncertainty around the order of integration of the series of interest for the purpose of cointegration analysis are well known in the literature.[1] Elliott (1998) discusses the robustness of cointegration methods when the regressors are near-unit processes, finding that while point estimation remains consistent, hypothesis testing based on the assumption of exact unit roots may lead to highly misleading results under slight deviations from the unit root benchmark. Furthermore, Pesaran *et al.* (2001) propose a cointegration test that bypasses pre-testing for unit roots, therefore allowing for uncertainty about whether regressors are stationary or non-stationary. More recently, Dou & Müller (2021) generalize the local-to-unity framework, showing that it can approximate well a large class of models that generate substantial persistence not dissimilar to I(1) processes. On the other hand, Duffy & Simons (2023) characterize cointegration

[1] The literature on near-unit roots is extensive, see the recent survey by Phillips (2021).

based on impulse response functions in a way that long run relationships are identified even the series are not exact unit roots.

In a similar vein, and inspired by nature of paleoclimatic time series, we examine the consequences of the presence of large, persistent cycles for cointegration inference. We do so analytically and by means of Monte Carlo simulations, demonstrating that, under a variety of DGPs for persistent cycles in a VECM setting, ML point estimation remains consistent, but converging at rate-$\sqrt{T}$, not at a rate-T as in the standard cointegration approach. Moreover, the trace and max-eigenvalue cointegration tests will also diverge, therefore suggesting that the data is stationary and the variables should be modelled in levels rather than in first-differences. These results are, to some extent, akin to the problem of cointegration with non-unit root regressors.

Indeed, the importance of modelling persistent cycles in climate time series has been long recognized in the climate literature, see Mitchell (1976). Several papers attempt to decompose the variability of the paleoclimate data in the frequency domain, see Mann & Lees (2019), Muller & MacDonald (1997), Wunsch (2008), Meyers *et al.* (2008) and Ditlevsen *et al.* (2020), *inter alia.* More recently, Proietti & Maddanu (2022) proposes an alternative parametric approach through the fractional sinusoidal waveform process.

Given that persistent cycles have the potential to lead to misleading cointegration inference, we propose to study the long-run comovement between temperatures and CO_2 by employing recently developed methods that highlight low-frequency covariability of time series; see Müller & Watson (2018) and Müller & Watson (2021), MW henceforth. Indeed, we suggest that variation in paleoclimate time series is dominated by glaciation cycles and therefore contain only limited information about the very long-run relation between temperatures and CO_2. Using the procedures of MW allows us to isolate a small number of low-frequency trigonometric weighted averages, which are then used to conduct inference about the long-run (co)variability of temperatures and CO_2. Moreover, in order to attenuate "curse-of-dimensionality" issues, we allow for a group of effects to be "partialled out", so that we are able hone in on the long-run relationship between temperatures and CO_2 while controlling for orbital forcing.

This approach is quite novel and distinctive in this literature and there are several advantages in using these low-frequency techniques. First, they allow us to focus on time spans that go beyond cyclical dynamics and thus obtain better estimates of long-run coefficients that are uncontaminated by short or medium run variations. Second, the methodology is flexible in that we can conduct inference on bivariate or multivariate low-frequency features of several climate variables. Third, these methods are fairly robust to the persistence patterns of paleoclimate data, permitting combinations of nonstationary, near-nonstationary or stationary series, thus circumventing the need to (pre) test for unit roots and cointegration. Indeed, the data transformations are approximately Gaussian and therefore standard inference tools and confidence intervals can be employed. Finally, the construction of long horizon forecasts is relatively straightforward, thus

allowing us to complement the scenario analysis in Castle & Hendry (2020).

Put simply, the method consists in obtaining low-frequency weighted averages of the data by using trigonometric projections that ensure approximate Gaussianity of the transformed series. Inference is then based on a (relatively) small number q of low-frequency averages, whereby correlation or regression coefficients (and respective standard errors) can be computed in the usual way. The choice of q determines the cyclicality that the researcher wishes to study. In our case, we are agnostic about the choice of q and suggest selecting q such that covariability is maximized. As it turns out, in our case q ranges from 16 to 20, which indicates spans of data between 80,000 to 100,000 years, consistent with the cyclicality in paleoclimate data.

Thus, we start by obtaining measures of long-run covariability between temperatures and levels of CO_2, with our results suggesting that correlation is indeed strong and quite significant, with a 90% confidence interval of $[0.90, 0.98]$, while the long-run regression coefficient is estimated at around 0.09. Given that the exclusively bivariate focus is likely to lead to overestimation, we use an extension of the MW approach that allows us to control for measures capturing orbital forcing, i.e., a selection of control variables is "partialled out" prior to the application of low-frequency projections. The resulting confidence intervals for the regression coefficient of temperatures on CO_2 range between 0.048 and 0.106, and therefore not too dissimilar to the results found in Castle & Hendry (2020).

We also pursue the alternative approach of summarizing relevant exogenous information by extracting a common factor driving orbital forcing and CO_2 levels, with these results confirming our previous findings. Finally, emulating Castle & Hendry (2020), we present long-term forecasts for temperatures and ice-volume that can be constructed from low-frequency (multivariate) factor models, conditioning on (anthropogenically determined) current levels of CO_2 concentrations, which are far higher than those typical during previous glaciar cycles. Our results suggest a steady increase in temperatures, coupled with a substantial decline in ice volumes well below historical minima, pointing towards an ice-free planet under current CO_2 levels.

The paper is organised as follows. Section 2 provides a brief description of the data and a set of models that generate persistent cycles, used in section 3 to analyse the implications for cointegration inference within a VECM setup. In section 4, we explore the low-frequency inference procedures of MW using paleoclimate data, presenting baseline results for the simple bivariate relationship between temperatures and CO_2 levels, as well estimations considering additional controls, which can either be "partialled out" or encapsulated into a common factor extracted via principal component analysis. Furthermore, we discuss multivariate low-frequency analysis and forecasting of paleoclimate time series, while section 5 offers some concluding remarks.

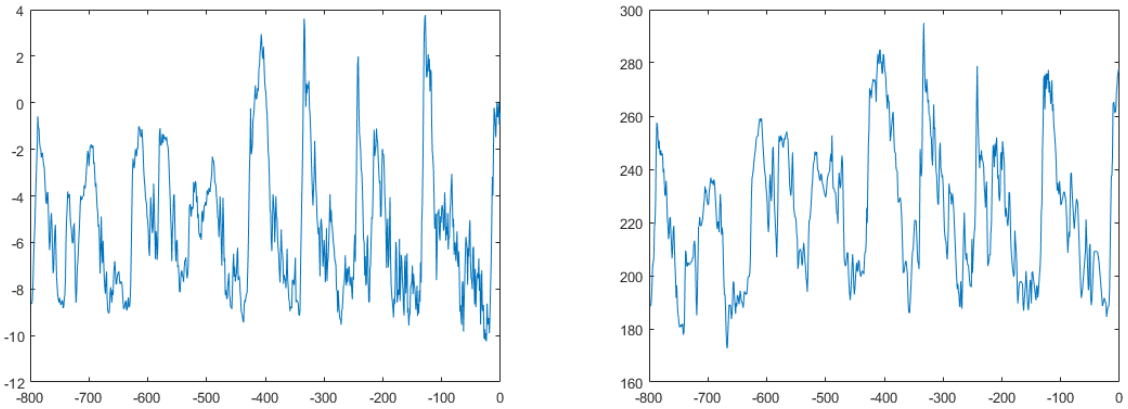
2 Data, Related Literature and Models

To motivate the discussion, we begin this section with an overview of the paleoclimatic data, highlighting the main features, namely persistent cyclicality and comovement. We then consider simple models that generate marked cyclical components, which will be used in section 3 to study their effects on cointegration analysis.

2.1 Data

We focus on three paleoclimate time series recently investigated in Davidson *et al.* (2016), Castle & Hendry (2020, Ch. 6) and Proietti & Maddanu (2022): these are ice cores reconstructions of temperatures, CO_2 atmospheric levels and ice volume. The first two series are sourced from the European Project for Ice Core in Antarctica (EPICA) (Jouzel *et al.*, 2007; Loulergue *et al.*, 2008; Lüthi *et al.*, 2008), with CO_2 measured in ppm (parts per million, where 1 ppm= 7.8 gigatonnes of CO_2, while the ice volume series is from Lisiecki & Raymo (2005). In addition, and following Kaufmann & Juselius (2013), we consider standard ice-age orbital drivers, namely Eccentricity (Ec), Obliquity (Ob) and Precession (Pr).

The period comprises 800,000 years (800 kyr), with all observations adjusted to the common EDC3 time scale and linearly interpolated for missing observations. The total sample size in 1000 year intervals is thus $T = 801$ with the last 100 observations (i.e. 100,000 years, ending 1000 years before the present) used to evaluate the predictive ability of our models (see Kaufmann & Juselius, 2013; Davidson *et al.*, 2016; Castle & Hendry, 2020 for further details on the construction of the data, as well as Miller, 2019 for the consequences of interpolation on subsequent statistical analysis).



(a) Temperature

(b) CO_2

Figure 1: Temperatures and CO_2 concentration levels

The most striking feature of the paleoclimate series is their cyclicality, i.e. significant recurring comovements, such that temperature and CO_2 levels stay below their mean for long periods during

glaciations (see Figure 1), with the opposite pattern for ice volume. As extensively discussed in Davidson *et al.* (2016), this behaviour makes it difficult to reconcile with the cointegration approach used to study long-run relationships and followed in much of the literature. In fact, it could be argued that glacial cycles cycles are the long-run movements in the data, as they play the most prominent role, recurring every 80.000-100.000 years (although not in a completely regular fashion).

Modelling paleoclimatic ice core data is an inherently difficult task because the structural relationship between orbitally-induced insolation and seasons in the paleoclimate rests on semi-quantitative arguments. For example, using filtered data, Lisiecki (2010) and Kawamura *et al.* (2010) present evidence that strongly suggests a relationship between orbital variables based on their frequency spectra. On the other hand, Rial (1999) suggests that the eccentricity signal acts as a frequency modulator for ice volume and argues why therefore certain peaks in the ice volume spectrum are missing. These approaches do not establish a direct relationship between variables by means of a statistical model, but rather present circumstantial evidence and physical mechanisms in favour of one. In contrast, Imbrie & Imbrie (1980) establish a formal, albeit deterministic, model which, to date, offers the cleanest direct relationship between orbital variables and ice volume on Earth, with a broad consensus in the literature regarding the likely non-linear nature of the relationship.

Aside from the role of orbital variation in driving ice ages, considerable attention has been devoted to the interplay of temperatures, ice volume and atmospheric concentrations of CO_2. The treatment of the latter and, in particular, whether it should be treated as a forcing variable or an endogenous response, has been subject of some debate. Lea (2004) and Jaccard *et al.* (2016), for example, suggest that orbital forcings drive variations in temperatures, which in turn affect ice volume and how much trapped CO_2 is released into the atmosphere. In turn, Davidson *et al.* (2016) find that CO_2 and other greenhouse gases are Granger-caused by temperatures.

Consequently, Castle & Hendry (2020) model these variables as a jointly endogenous system, with orbital forcing variables deemed to be strongly exogenous. They find evidence of an endogenous response of CO_2 to orbital forcing, as well as support for a 'weak' form of the Milankovitch hypothesis, i.e. in order to account for all aspects of glacial cycles, one needs to also consider nonlinear interactions between the different orbital components (in line with the physics literature cited above).

Nevertheless, the features in the data remarked above raise the following questions:

1. To what extent the strong dynamics of persistent cycles (such as those displayed in Figure 1) impair cointegration analysis?

2. Once these (exogenous) cycles are taken into account, how can one quantify with precision the relationship between temperatures and CO_2?

This section and the next attempt to provide an answer to the first question, while section 4 deals with 2. Given its importance for accurate climate sensitivity projections, our focus is on inference about the long-run relationship between temperature and gases concentrations, so in order to abstract from the admittedly difficult-to-model cyclical component, we propose using the low-frequency "filtering" procedures of MW.

2.2 Simple Models for Persistent Cycles

As discussed in the preceding sub-section and in the wider literature, orbital variables (and, to a large extent, paleoclimatic variables) possess very persistent cycles. Proietti & Maddanu (2022) recently proposed the fractional sinusoidal waveform process to model the strong cyclical component, such that parameters that regulate the amplitude and the cycle phase follow fractional noise processes. We present alternative, potentially simpler, models, distinguishing deterministic from stochastic laws.

Periodic Deterministic and Stochastic Cycles As a simple way to capture the data properties, we first consider the following model:

$$y_t = \mu + \delta_t \cos\left(\lambda\pi\frac{t}{T}\right) + \theta u_t, \text{ with } u_t = \rho u_{t-1} + \varepsilon_t, \quad t = 2, ..., T, \tag{1}$$

where $\lambda > 2$ is constant, δ_t can be time-varying and ε_t is a white noise (WN) with unit variance.[2] The number of cycles $\lambda/2$, each with periodicity $2T/\lambda$, is assumed to be constant, whereas the expected cycle's range $[\mu - \delta_t, \mu + \delta_t]$ can vary over time, both properties being present in the data reviewed in section 2.1. The parameters θ and ρ control the amount and persistence of noise around the cycle, respectively, and are assumed to follow an AR(1) process, possibly a near-unit root process.

The component that is left with the most general specification as possible is δ_t, which controls the cycle's range. This can be deterministic or stochastic, and with different orders of convergence. For data that remains relatively stable around μ, such that δ_t is of the same order of u_t, it is assumed that, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor}\delta_t \to \widetilde{\delta}(r)$, i.e., $\sum_{t=1}^{\lfloor rT \rfloor}\delta_t = O(\sqrt{T})$ or $O_p(\sqrt{T})$ for any $r \in [0,1]$. In this case, the scaled demeaned partial sum

$$T^{-1/2}\sum_{t=1}^{\lfloor rT \rfloor}(y_t - \mu) \Longrightarrow \cos(\lambda\pi r)\,\widetilde{\delta}(r) + (\lambda\pi r)\int_0^1 \sin(\lambda\pi r s)\,\widetilde{\delta}(s)\,ds + \sigma_u W(r), \tag{2}$$

as $T \to \infty$, for $r \in [0,1]$ (see Bierens, 1994, Lem. 9.6.3, p. 200). If δ_t is deterministic, one can consider $\delta_t = \frac{\delta}{\sqrt{T}}$, where $\widetilde{\delta}(r) = \delta\lfloor r \rfloor$, or a linear time trend $\delta_t = \delta\frac{t}{T^{3/2}}$, with $\widetilde{\delta}(r) = \delta\frac{1}{2}r^2$. If δ_t

[2] Alternative models for u_t could have been considered. For the purpose of this paper, we assume the AR(1) specification but, in general, u_t is such that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor}u_t \Longrightarrow \sigma_u W(r), r \in [0,1]$, where σ_u^2 is the long-run covariance of u_t and $W(r)$ is a standard Wiener process, with $[\cdot]$ denoting the integer part of $\cdot$.

is stochastic, it can be assumed, for example, that $\delta_t = \frac{1}{T}\sum_{j=1}^{t}\epsilon_j$, where ϵ_t is a $\text{WN}\left(0,\sigma_\varepsilon^2\right)$, so that $\widetilde{\delta}(r) = \sigma_\varepsilon \int_0^r W(s)\,ds$.

The paleoclimatic data possesses cycles with wider ranges and therefore $\sum_{t=1}^{\lfloor rT\rfloor}\delta_t = O(\sqrt{T})$ or $O_p(\sqrt{T})$ is not a reasonable assumption to make in this case. For example, if δ_t is a constant ($\delta_t = \delta$), we have $\sum_{t=1}^{T}\delta_t = O_p(T)$ and $T^{-1/2}\sum_{t=1}^{T}(y_t - \mu) = O(\sqrt{T}) + T^{-1/2}\sum_{t=1}^{T}u_t$ because

$$\frac{1}{T}\sum_{t=1}^{T}\cos\left(\frac{\lambda\pi}{T}t\right) \to \frac{\sin(\lambda\pi)}{\lambda\pi}, \text{ as } T\to\infty. \tag{3}$$

For integer $\lambda > 2$, $\frac{1}{T}\sum_{t=1}^{T}\cos\left(\frac{\lambda\pi}{T}t\right) \to 0$, as $T\to\infty$. Hence, $\delta_t\cos\left(\lambda\pi\frac{t}{T}\right)$ dominates over θu_t and the normalization is the first moment of the series

$$\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu) \xrightarrow{p} \delta\frac{\sin(\lambda\pi)}{\lambda\pi}. \tag{4}$$

Similarly, for deterministic $\delta_t = \delta_0\left(1 - \delta_1\frac{t}{T}\right)$ we have $\frac{1}{T}\sum_{t=1}^{T}\delta_t \to \delta_0\left(1 - \delta_1\frac{1}{2}\right)$, and for stochastic (scaled random walk) $\delta_t = \delta_0 + \frac{1}{T}\sum_{j=1}^{t}\epsilon_j, \delta_0 \neq 0$, we have $\frac{1}{T}\sum_{t=1}^{T}\delta_t \to \delta_0$.

Persistent Cycle-Near-Unit Root Decomposition Gabriel *et al.* (2023) propose an alternative model to capture persistent cycles based on a simple cycle/noise decomposition where, instead of a cosine function, the cyclical component follows a latent AR(2) process with complex unit roots. Specifically,

$$y_t = \mu + z_t + \gamma u_t, \quad t = 1,...,T, \tag{5}$$

with

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \eta v_t, \tag{6}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{7}$$

where v_t and ε_t are two mutually independent $\text{WN}(0,1)$. Here, μ controls for the level and γ for the amount of noise in the data. z_t is a nonstationary AR(2) process where ϕ_1 and ϕ_2 control for the (time-invariant) frequency of the cycles (following Bierens, 2001, $\phi_2 = -1$ and $\phi_1 = 2\cos(\phi)$ so that the cycle lasts $2\pi/\phi$ periods). The parameter η controls for the time-variability of the cycle's range (if $\eta = 0$ the cycle has constant range; the larger η, the larger variability the cycle exhibits) and z_1, z_2 are constants with probability 1 – z_2 controls for the cycle's range after, for the sake of simplicity, assuming $P(z_1 = 0) = 1$. Finally, u_t is a stationary near-unit root AR(1) process that smoothly fluctuates around the cycle component.[3]

From Bierens (2001), imposing $\phi_2 = -1$, $\phi_1 = 2\cos(\phi)$ and $\eta = 1$, Equation (6) implies

$$z_t = \frac{1}{\sin(\phi)}\sum_{j=1}^{t}\sin(\phi(t+1-j))\,\varepsilon_j + d_t; \tag{8}$$

$$d_t = A\cos(\phi t) + B\sin(\phi t) + C, \tag{9}$$

[3]The pair z_1, z_2 can also be random $O_p(1)$ processes, but in that case the cycle's range is random. As earlier, alternative models for u_t can be considered. Notice that if $\rho = 1$, u_t is a nonstationary process and therefore y_t itself would resemble a I(1) process (the cycle z_t would not be clearly defined in the data).

where A, B, C depend on initial conditions. Thus, y_t is explained through

$$y_t = \mu + d_t + \frac{1}{\sin(\phi)} \sum_{j=1}^{t} \sin(\phi(t+1-j))\varepsilon_j + \gamma \left(\rho^t u_0 + \sum_{j=0}^{t-1} \rho^j v_{t-j} \right), \tag{10}$$

which may include a trend d_t (see Gabriel *et al.* (2023) for details about the properties of the model and the estimation and inference of parameters).

3 Cointegration Analysis in the Presence of Persistent Cycles

In this Section, we study the properties standard cointegration inference for models with I(1) variables when the true data includes persistent cycles. We use Monte Carlo simulations for a variety of models and derive analytical results for a toy model.

3.1 Monte Carlo Simulations

For the sake of simplicity, and following the existing literature, we consider the bivariate vector error correction model

$$\begin{pmatrix} \triangle y_t \\ \triangle x_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (1, \beta_2) \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, \quad t = 2, ..., T, \tag{11}$$

the Maximum Likelihood Estimator (MLE) for $\alpha_1, \alpha_2, \beta_2$ and the trace eigenvalue test for cointegration. We study the bias of the MLE (empirical mean and standard deviation) and for $\widehat{\beta}_2$ we compute distributional quantiles ($\tau \in [0,1]$ with empirical quantiles $\tau = 0.05, 0.50$, and 0.95) for two convergence rates, $T(\widehat{\beta}_2 - \beta_2)$ and $\sqrt{T}(\widehat{\beta}_2 - \beta_2)$, to address the typical cases of nonstationary and stationary data. For the trace test, we compute the percentage of times there is evidence of stationary data, $r = 2$ (rejections of spurious and cointegration regressions). The sample sizes are $T = 50, 100, 800, 10000$ to cover the cases of small samples, our paleoclimate application, as well as large samples. The number of replications is set to 5000.

To better understand the consequences of estimating the model described in (11) when the data $\{x_t, y_t\}_{t=1}^{T}$ includes persistent cycles, we begin by presenting the results for the standard cointegration case where variables are nonstationary unit root processes. Let $x_t = x_{t-1} + u_{xt}$, where u_{xt} is a Gaussian WN(0,1), and $y_t = x_t + v_t$, where v_t is also a Gaussian WN(0,1), independent of u_{xt}. That is, the true coefficients are $\alpha' = (\alpha_1, \alpha_2) = (-1, 0)$ and $\beta' = (1, \beta_2)$ with $\beta_2 = -1$. The results are in Table 1.

Now, we consider distinct models for x_t with persistent cycles, keeping the same specification as before, where $y_t = x_t + v_t$, so that $\alpha' = (-1, 0)$ and $\beta_2 = -1$. In the first case, $x_t = \delta_t \cos\left(\lambda\pi\frac{t}{T}\right) + \varepsilon_t$, where ε_t is a Gaussian WN(0,σ_ε^2), with a deterministic δ_t. Regarding the model in Equation (1), we assume $\mu = \rho = 0$ and $\theta = 1$. For constant $\delta_t = \delta$, by visual inspection of temperatures and CO2 emissions, where T is around 800, we consider the following

Table 1: Estimation of the Bivariate ECM - Nonstationary Unit Root Processes

	trace	β_2		$T(\beta_2+1)$			$\sqrt{T}(\beta_2+1)$			α_1		α_2	
T	mean	mean	stdev	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	mean	stdev	mean	stdev
50	0.049	-1.000	0.051	-3.901	-0.012	3.896	-0.551	-0.001	0.551	-0.534	0.872	-0.008	0.145
100	0.047	-1.000	0.024	-3.965	0.003	3.862	-0.396	3.8e-04	0.386	-0.483	0.890	-0.010	0.101
800	0.051	-1.000	0.003	-3.837	0.063	3.708	-0.135	0.002	0.131	-0.397	0.917	-0.003	0.035
10000	0.054	-1.000	2.3e-04	-3.839	-0.028	3.729	-0.038	-2.8e-04	0.037	-0.330	0.944	-0.001	0.009

Notes: Model (11); *trace* is the percentage of rejections of spurious and cointegration regressions using the trace test.

parameterization: $\delta = 7$ or $\delta = 80$, $\lambda = 6$ or $\lambda = 16$, and $\sigma_\varepsilon = 1$ or $\sigma_\varepsilon = 7$. For time-varying δ_t, the scaled linear trend is $\delta_t = \delta_0\left(1 - \delta_1 \frac{t}{T}\right)$ for $\delta_0 = 7$, and $\delta_1 = -1$ or $\delta_1 = -2$, for fixed $\lambda = 16$ and $\sigma_\varepsilon = 1$.

The results are in Table 2. There are three main findings that stand out. First, the MLE is not biased, with $\widehat{\beta}_2$ converging faster to the true value than $\widehat{\alpha}_1, \widehat{\alpha}_2$. Second, the rate at which $\left(\widehat{\beta}_2 - \beta_2\right)$ converges in distribution is not the T-rate, but rather the $\sqrt{T}$-rate, and, as expected, the 5^{th} and 95^{th} quantiles depend on the model's specification. Putting together these two findings and contrasting with the results in Table 1, we conclude that fitting a standard VECM with data that follows the model in Equation (1) leads to unbiased estimates, but with erroneous standard errors, confidence sets, and therefore inference. The third finding pertains to the trace cointegration test: clearly, by testing for cointegration in the standard VECM, one will conclude that the data is stationary (the exception being the case of $T = 50, 100$ when $\delta = 7, \lambda = 6, \sigma_\varepsilon = 1$, where there is evidence for cointegrated I(1) data). Therefore, if the data is stationary, one should model variables in levels instead of first-differences. In general, these main conclusions are robust to the particular specification of the deterministic δ_t.

In Table 3, we present results for the same model for x_t, but with $\sigma_\varepsilon = 1$ and with a stochastic δ_t. We consider a scaled random walk $\delta_t = \gamma + \delta_{t-1} + \frac{1}{T}\xi_t$, with $\xi_t \sim \text{WN}\left(0, \sigma^2_\xi\right)$, both with no drift ($\gamma = 0$) and with a drift ($\gamma = 0.01$). Moreover, from the AR(1) specification $\delta_t = \phi_0 + \phi_1\delta_{t-1} + \xi_t$, the cases of a $\text{WN}\left(\phi_0, \sigma^2_\xi\right)$, a stationary AR(1) with $\phi_1 = 0.5$, and a near-unit root process, where $\phi_1 \equiv \phi_{1,T} = 1 - \frac{1}{T}$. The conclusions are the same as those obtained for a deterministic δ_t, with the exceptions of the scaled random walk with a drift and of the near-unit root process – in these two cases, the range of the cycles is not stable, which goes against what is observed in the paleoclimatic data.

Finally, we study the properties of the standard VECM (11) with $\alpha' = (-1, 0)$ and $\beta' = (1, -1)$ when x_t follows the persistent cycle-near-unit root decomposition. The cycle component is defined by Equation (6) with x_t evolving as a nonstationary AR(2) process with complex unit roots and, for comparability, we also consider x_t as following the fractional sinusoidal waveform process of Proietti & Maddanu (2022). In the first case, $\mu = 0, z_1 = 0, z_2 = 0.25, \gamma = 1, \eta = 0.001, \rho = 0.95$ and $\phi_1 = 1.9960534, \phi_2 = -1$.[4] In the AR(2) model, $x_t = 0.02 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + 0.01u_t$, where

[4]Results are very similar for $\phi_1 = 1.6355457, \phi_2 = -0.7$ or for $\phi_1 = 1.411423, \phi_2 = -0.5$.

Table 2: Estimation of the Bivariate ECM - Periodic Deterministic Cycle

	trace	β_2		$T(\beta_2+1)$			$\sqrt{T}(\beta_2+1)$			α_1		α_2	
T	mean	mean	sd	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	mean	sd	mean	sd
$\delta=7;\lambda=6;\sigma_\varepsilon=1$													
50	0.005	-1.000	0.029	-2.488	0.030	2.413	-0.351	0.004	0.341	-0.572	0.899	-0.053	0.335
100	0.012	-1.000	0.020	-3.484	0.019	3.376	-0.348	0.002	0.337	-0.583	0.837	-0.042	0.168
800	1	-1.000	0.007	-9.346	-0.125	9.082	-0.330	-0.004	0.321	-0.623	0.783	-0.017	0.048
10000	1	-1.000	0.002	-34.035	-0.086	34.259	-0.340	-8.6e-04	0.342	-0.924	0.381	-0.002	0.014
$\delta=7;\lambda=16;\sigma_\varepsilon=1$													
50	1	-0.999	0.031	-2.581	0.050	2.599	-0.365	0.007	0.367	-0.614	1.041	-0.075	0.669
100	1	-1.000	0.020	-3.489	-0.016	3.412	-0.348	-0.001	0.341	-0.551	0.886	-0.055	0.284
800	1	-1.000	0.007	-9.280	0.090	9.420	-0.328	0.003	0.333	-0.648	0.763	-0.017	0.049
10000	1	-1.000	0.002	-32.642	-0.471	33.478	-0.326	-0.004	0.334	-0.923	0.384	-0.002	0.014
$\delta=80;\lambda=16;\sigma_\varepsilon=7$													
50	1	-0.999	0.002	-0.219	0.004	0.221	-0.031	6.4e-04	0.031	-0.895	7.298	0.024	7.290
100	1	-1.000	0.001	-0.303	-0.002	0.295	-0.030	-2.0e-04	0.029	-0.809	3.046	0.008	2.997
800	1	-1.000	6.1e-04	-0.800	0.005	0.817	-0.028	2.0e-04	0.028	-0.576	0.896	-0.092	0.362
10000	1	-1.000	1.7e-04	-2.837	-0.042	2.905	-0.028	-4.2e-04	0.029	-0.481	0.883	-0.040	0.091
$\delta_0=7;\delta_1=-1;\lambda=16;\sigma_\varepsilon=1$													
50	1	-0.999	0.019	-1.611	0.031	1.612	-0.227	0.004	0.228	-0.595	1.256	-0.028	0.960
100	1	-1.000	0.013	-2.255	-0.010	2.166	-0.225	-0.001	0.216	-0.475	0.969	-0.023	0.401
800	1	-1.000	0.004	-5.971	0.102	6.091	-0.211	0.003	0.215	-0.492	0.872	-0.018	0.049
10000	1	-1.000	0.001	-21.236	-0.161	21.748	-0.212	-0.001	0.217	-0.731	0.682	-0.004	0.013
$\delta_0=7;\delta_1=-2;\lambda=16;\sigma_\varepsilon=1$													
50	1	-0.999	0.014	-1.168	0.018	1.152	-0.165	0.002	0.163	-0.632	1.499	-0.008	1.279
100	1	-1.000	0.009	-1.625	-0.014	1.575	-0.162	-0.001	0.157	-0.463	1.032	0.013	0.528
800	0	-1.000	0.003	-4.351	0.081	4.445	-0.153	0.002	0.157	-0.448	0.895	-0.015	0.052
10000	1	-1.000	9.6e-04	-15.476	-0.086	15.988	-0.154	-8.6e-04	0.159	-0.598	0.801	-0.005	0.013

Notes: Model (1) with δ_t deterministic; *trace*: See Table 1

Table 3: Estimation of the Bivariate ECM - Periodic Stochastic Cycles

	trace	β_2		$T(\beta_2+1)$			$\sqrt{T}(\beta_2+1)$			α_1		α_2	
T	mean	mean	sd	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	$\tau=0.05$	$\tau=0.5$	$\tau=0.95$	mean	sd	mean	sd
$\gamma=0;\delta_0=7;\lambda=16;\sigma_\xi=1$													
50	1	-0.999	0.031	-2.560	0.014	2.643	-0.362	0.002	0.373	-0.641	1.040	-0.076	0.678
100	1	-0.999	0.020	-3.332	-0.006	3.532	-0.333	-6.8e-04	0.353	-0.557	0.879	-0.053	0.282
800	1	-0.999	0.007	-9.189	0.130	9.372	-0.324	0.004	0.331	-0.649	0.764	-0.018	0.048
10000	1	-1.000	0.002	-33.230	0.192	33.205	-0.332	0.001	0.332	-0.925	0.378	-0.001	0.014
$\gamma=0.01;\delta_0=7;\lambda=16;\sigma_\xi=1$													
50	1	-0.999	0.030	-2.465	0.015	2.547	-0.348	0.002	0.360	-0.611	1.070	-0.083	0.696
100	1	-0.999	0.019	-3.091	-0.004	3.294	-0.309	-4.4e-04	0.329	-0.556	0.884	-0.044	0.298
800	0.999	-0.999	0.004	-5.605	0.079	5.784	-0.198	0.002	0.204	-0.498	0.870	-0.018	0.050
10000	0	-1.000	2.2e-04	-3.636	-0.006	3.701	-0.036	-6.5e-05	0.037	-0.270	0.962	-0.004	0.013
$\phi_0=7;\phi_1=0;\lambda=16;\sigma_\xi=1$													
50	1	-0.999	0.031	-2.497	0.015	2.618	-0.353	0.002	0.370	-0.632	1.049	-0.085	0.681
100	1	-0.999	0.021	-3.389	0.002	3.526	-0.338	2.5e-04	0.352	-0.563	0.883	-0.064	0.300
800	1	-0.999	0.007	-9.172	0.197	9.343	-0.324	0.007	0.330	-0.662	0.788	-0.023	0.059
10000	1	-1.000	0.002	-33.441	0.084	33.153	-0.334	8.4e-04	0.331	-0.871	0.490	-0.003	0.017
$\phi_0=7;\phi_1=0.5;\lambda=16;\sigma_\xi=1$													
50	1	-0.999	0.015	-1.236	0.008	1.284	-0.174	0.001	0.181	-0.589	1.484	-0.021	1.229
100	1	-0.999	0.010	-1.663	-0.002	1.739	-0.166	-2.0e-04	0.174	-0.451	1.026	-0.021	0.509
800	0.999	-0.999	0.003	-4.526	0.045	4.663	-0.160	0.001	0.164	-0.524	0.856	-0.019	0.059
10000	1	-1.000	9.9e-04	-16.558	-0.013	16.350	-0.165	-1.3e-04	0.163	-0.623	0.782	-0.005	0.015
$\phi_0=7;\phi_1=\left(1-\frac{1}{T}\right);\lambda=16;\sigma_\xi=1$													
50	1	-1.000	0.001	-0.120	0.002	0.120	-0.017	3.1e-04	0.017	-1.032	12.229	-0.082	12.224
100	0.899	-1.000	4.9e-04	-0.080	-5.1e-04	0.083	-0.008	-5.1e-05	0.008	-0.869	9.833	0.078	9.831
800	0	-1.000	2.1e-05	-0.027	1.0e-04	0.028	-9.7e-04	3.8e-06	0.001	-0.981	3.607	-0.017	3.597
10000	1	-1.000	4.9e-07	-0.008	-3.9e-05	0.008	-8.1e-05	-3.9e-07	8.2e-05	-1.007	1.029	-0.018	1.016

Notes: Model (1) with δ_t stochastic; *trace*: See Table 1

Table 4: Estimation of the Bivariate ECM - Alternative Models with Stochastic Cycles

	trace	β_2		$T(\beta_2+1)$			$\sqrt{T}(\beta_2+1)$			α_1		α_2	
T	mean	mean	sd	q05	q50	q95	q05	q50	q95	mean	sd	mean	sd
Persistent cycle-near unit root decomposition													
50	0.040	-1.000	0.051	-4.006	-0.005	3.991	-0.566	-8.2e-04	0.564	-0.602	0.835	-0.034	0.147
100	0.055	-0.999	0.028	-4.406	-0.025	4.543	-0.440	-0.002	0.454	-0.634	0.791	-0.026	0.101
800	1	-0.999	0.008	-11.188	0.171	11.391	-0.395	0.006	0.402	-0.819	0.573	-0.006	0.035
10000	1	-1.000	0.002	-38.917	-0.228	39.253	-0.389	-0.002	0.392	-0.999	0.032	-3.1e-06	0.010
AR(2) process with complex unit roots: $\phi_1 = 1.9960534; \phi_2 = -1$													
800	6.0e-04	-1.000	0.005	-7.158	0.022	6.969	-0.253	7.7e-04	0.246	-0.920	0.394	7.7e-04	0.010
10000	0.914	-1.000	0.001	-16.406	-0.389	16.825	-0.164	-0.003	0.168	-0.957	0.287	1.8e-04	0.006
AR(2) process with complex unit roots: $\phi_1 = 1.6355457; \phi_2 = -0.7$													
800	0	-0.999	0.112	-147.065	0.098	148.402	-5.199	0.003	5.246	-1.000	0.044	6.1e-06	5.9e-04
10000	1	-1.000	0.031	-525.754	-6.460	534.869	-5.257	-0.064	5.348	-0.999	0.012	6.8e-07	1.6e-04
AR(2) process with complex unit roots: $\phi_1 = 1.411423; \phi_2 = -0.5$													
800	0	-0.999	0.156	-203.144	0.339	204.569	-7.182	0.012	7.232	-1.000	0.044	4.8e-06	4.7e-04
10000	1	-1.000	0.043	-732.138	-11.932	737.315	-7.321	-0.119	7.373	-0.999	0.012	3.4e-07	1.3e-04
Fractional sinusoidal waveform process: Temperatures													
50	0	-0.999	0.048	-3.983	0.056	4.026	-0.563	0.008	0.569	-0.667	0.768	-0.001	0.017
100	0	-0.999	0.034	-5.598	0.027	5.709	-0.559	0.002	0.571	-0.854	0.536	-2.5e-04	0.012
800	0	-1.000	0.012	-16.210	-0.093	15.951	-0.573	-0.003	0.564	-1.000	0.043	-9.2e-05	0.004
10000	1	-1.000	0.003	-55.627	0.496	56.793	-0.556	0.005	0.567	-0.999	0.012	3.5e-05	0.001
Fractional sinusoidal waveform process: CO2 emissions													
50	0.981	-0.942	7.321	-580.065	2.905	583.686	-82.033	0.410	82.545	-0.997	0.173	-6.1e-06	0.004
100	1	-0.940	4.503	-728.347	2.376	740.218	-72.834	0.237	74.021	-0.999	0.123	2.2e-05	0.002
800	1	-0.978	1.231	-1.6e+03	2.604	1.6e+03	-56.618	0.092	58.207	-0.999	0.042	-1.7e-05	9.9e-04
10000	1	-1.003	0.297	-4.8e+03	-30.469	4.8e+03	-48.940	-0.304	48.541	-0.999	0.012	7.4e-06	2.7e-04

Notes: Models (5), AR(2) with complex UR, and Fractional sinusoidal; *trace*: See Table 1

$u_t = 0.2u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{WN}(0,1)$, and we allow (ϕ_1, ϕ_2) to take three different combinations to capture distinct cycle lenghts: $(1.9960534, -1), (1.411423, -0.5)$, and $(1.6355457, -0.7)$.[5] The two fractional sinusoidal waveform processes are generated according to the model and the estimated coefficients for paleoclimatic temperatures and CO2 emissions (we refer the reader to Proietti & Maddanu, 2022 for details). Once again, the main conclusions obtained earlier hold true for these data generating processes. Noticeably, from the trace test there is a switch from cointegration to stationary data as T gets large.

3.2 A Theoretical Result

The Monte Carlo simulations suggest that the estimation of the standard VECM when the data has persistent cycles provides unbiased estimators, a different rate of convergence in distribution and, consequently, erroneous inference, as well as cointegration tests that identify variables as

[5]The results for $T = 50, 100$ have been left out because in these cases the cycles are not noticeable in the time series.

being stationary. We now show this result analytically for a simple model.

Consider as the true data generating process the following bivariate vector time series process:

Assumption 1: Let $z_t = (y_t, x_t)'$, where $y_t = x_t + u_{yt}$ and $x_t = \delta \cos\left(\lambda\pi \frac{t}{T}\right) + u_{xt}$, with $u_t = (u_{yt}, u_{xt})$ following a bivariate standard normal distribution and u_{yt} and u_{xt} are mutually independent with variances σ_y^2 and σ_x^2, respectively.

That is, in first differences the models are

$$\triangle y_t = \delta\left[\cos\left(\lambda\pi\frac{t}{T}\right) - \cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right] + (\triangle u_{xt} + \triangle u_{yt}) \tag{12}$$

$$\triangle x_t = \delta\left[\cos\left(\lambda\pi\frac{t}{T}\right) - \cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right] + \triangle u_{xt} \tag{13}$$

and, equivalently, as the bivariate model

$$\triangle z_t = \mu_t \iota + \alpha\beta' z_{t-1} + \widetilde{u}_t, t = 2, ..., T \tag{14}$$

where

$$\mu_t \equiv \mu_t(\delta, \lambda) = \delta\left[\cos\left(\lambda\pi\frac{t}{T}\right) - \cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right]; \tag{15}$$

$$\iota' = (1,1); \alpha' = (-1,0); \beta' = (1,-1); \widetilde{u}_t = (\triangle u_{xt} + u_{yt}, \triangle u_{xt})'. \tag{16}$$

The misspecified model is the standard bivariate cointegrated ECM with $y_t = x_t + u_{yt}$ and $x_t = x_{t-1} + u_{xt}$, that is,

$$\begin{pmatrix} \triangle y_t \\ \triangle x_t \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}(1,-1)\begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix}, \quad t = 2, ..., T, \tag{17}$$

which ignores the time-varying intercept (cyclical component) $\mu_t(\delta, \lambda)$ and the errors are misspecified. From the standard Johansen (1988, 1991, 1995) approach, the model is

$$\triangle z_t = \alpha\beta' z_{t-1} + u_t, t = 2, ..., T, \tag{18}$$

where $u_t \sim i.i.d.N_2[0, \Omega]$, with Ω nonsingular, the log-likelihood for $r = 1$ (cointegration) is

$$\widehat{l}_T = -0.5T\ln\left(1 - \widehat{\lambda}\right) - 0.5T\ln\left(\det\left(S_{00,T}\right)\right) \tag{19}$$

plus a constant, where $\widehat{\theta}$ is the largest solution of the generalized eigenvalue problem

$$\det\left[\theta S_{11,T} - S_{10,T}S_{00,T}^{-1}S_{01,T}\right] = 0 \tag{20}$$

with

$$S_{00,T} = \frac{1}{T}\sum_{t=1}^{T}\triangle z_t \triangle z_t'; \quad S_{11,T} = \frac{1}{T}\sum_{t=1}^{T} z_{t-1}z_{t-1}'; \quad S_{01,T} = \frac{1}{T}\sum_{t=1}^{T}\triangle z_t z_{t-1}', \tag{21}$$

$S_{10,T} = S_{01,T}'$, and the ML estimator for β is the eigenvector associated with the largest eigenvalue $\widehat{\theta}$:

$$S_{10,T}S_{00,T}^{-1}S_{01,T}\widehat{\beta} = \widehat{\theta}S_{11,T}\widehat{\beta}. \tag{22}$$

The next Lemma provides the limiting result for the matrices $S_{ij,T}, i, j = 0, 1$ when the data follows model (1) for a constant δ.

Lemma 1: Under Assumption 1, as $T \to \infty$,

$$S_{00,T} \xrightarrow{p} 2\Upsilon; S_{11,T} \xrightarrow{p} \frac{1}{2}\delta^2 \left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota + \Upsilon; S_{01,T}, S_{10,T} \xrightarrow{p} -\Upsilon, \tag{23}$$

where $\Upsilon = \begin{bmatrix} \left(\sigma_x^2 + \sigma_y^2\right) & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 \end{bmatrix}, \iota = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. For $\lambda > 2$ integer, $S_{11,T} \xrightarrow{p} \frac{1}{2}\delta^2\iota + \Upsilon$.

Proof. See details in the Appendix.

From the previous Lemma, and following the Lemma 2 in Andersson et al. (1983), the ordered solutions $\widehat{\theta}_1 \geq \widehat{\theta}_2$ of the generalized eigenvalue problem in Equation (20) converge in probability to the ordered solutions $\theta_1 \geq \theta_2$ of $\det\left[\theta\left(\frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota + \Upsilon\right) - \Upsilon'\left(2\Upsilon\right)^{-1}\Upsilon\right] = 0$. These are provided in the next Theorem. Contrary to the standard cointegration approach, there are no zero solutions and, in particular, the largest $\widehat{\theta}$ converges in probability to $\frac{1}{2}$ for all admissible model parameters. Moreover, unlike the standard cointegration approach, since the smallest θ_2 is not zero, $T\widehat{\theta}_2$ will not converge in distribution to a random variable, but rather grows to infinity and, therefore, the trace and the maximum eigenvalue tests will also diverge as $T \to \infty$. The implication of this is that the cointegration tests will provide evidence of $rank = 2$, i.e. suggesting that the data is stationary and therefore there is no need to take first differences, as it is assumed in the case of the VECM model.

Theorem 1: Under Assumption 1, as $T \to \infty$, the ordered solutions $\widehat{\theta}_1 \geq \widehat{\theta}_2$ of the generalized eigenvalue problem in Equation (20) converge in probability to $\theta_1 = \frac{1}{2}$ and $\theta_2 = \frac{1}{\frac{\delta^2}{\sigma_x^2}\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+2} < \frac{1}{2}$, respectively. For $\lambda > 2$ integer, $\theta_2 = \frac{1}{\frac{\delta^2}{\sigma_x^2}+2}$. Moreover, $T\widehat{\theta}_2 \to \infty$, as $T \to \infty$.

Proof. The details are omitted, as the result is straightforward to obtain.

As a consequence of Theorem 1, the next Theorem shows that the MLE of $\widehat{\alpha}$ is consistent and $\widehat{\beta}$ is still consistent, but it converges at a rate-$\sqrt{T}$ and not at a rate-T, as in the standard cointegration approach. The main reason is because $S_{11,T}$ is $O_p(1)$, something distinct from the I(1) case where $\frac{1}{T}S_{11,T} = O_p(1)$.

Theorem 2: Under Assumption 1, as $T \to \infty$, $\widehat{\beta} \xrightarrow{p} b$, where the normalized b is given by $b = (1, -1)'$, and $\widehat{\alpha} \xrightarrow{p} \alpha$, where $\alpha' = (-1, 0)$. Moreover, as $T \to \infty$, $\sqrt{T}\widehat{\beta} = O_p(1)$ and $T\widehat{\beta} \to \infty$.

Proof. See the details in the Appendix.

Moreover, while not covered in these, the Monte Carlo simulations suggest that both for δ_t not constant, and for other more general models, the MLE of $\widehat{\beta}$ remains consistent, but the rate of convergence is $\sqrt{T}$, slower than T, the standard one.

4 Long-Run Covariability and Low-Frequency Inference

As discussed in the previous section, not accounting for the statistical properties of persistent cycles may lead to erroneous inferences within the cointegration framework, popularized in climate econometrics by Kaufmann *et al.* (2013) and Kaufmann & Juselius (2013). To overcome this issue, and with the objective of obtaining sensible (and testable) baseline estimates for climate sensitivity, we propose to use the low-frequency "filtering" framework of MW.

Indeed, low-frequency variation can be extracted by using relatively small number q of weighted averages, where the weights are deterministic (and known) low-frequency trigonometric series. Consider the simplest case of a single time series x_t observed over $t = 1, ..., T$, letting $\Psi_j(s) = \sqrt{2}cos(js\pi)$, so that $\Psi_j(t/T)$ has period $2T/j$, with $\Psi(s) = (\Psi_1(s), \Psi_2(s), ..., \Psi_q(s))'$ a $\Re^q$ valued function, with $\Psi(T) = (\Psi((1-1/2)/T), \Psi((2-1/2)/T), ..., \Psi((T-1/2)/T))'$ a $T \times q$ matrix obtained by evaluating the function at $s = (t-1/2)/T$ for $t = 1, ..., T$. Then, we can obtain low-frequency projections by obtaining the fitted values from running OLS regressions of x_t on $\Psi(T)$, such that $\hat{x}_t = \bar{x} + \Psi((t-1/2)/T)'X_T$, where $\bar{x}$ and X_T are OLS coefficients, the latter having the simple form $X_T = T^{-1}\sum_{t=1}^{T}\Psi((t-1/2)/T)'x_t$, such that the j^{th} regression coefficient X_{jt} is the j^{th} cosine transform of $(x_1, x_2, ..., x_T)'$.

Müller & Watson (2018) show that this can be extended to a multivariate setting, in a way that long-run correlation (which we will denote by ρ_L) or regression coefficients can be obtained in a standard way based on the q cosine projections. Considering the scarcity of low-frequency information in the data, it is natural that only a small number of projection coefficients are employed to capture low-frequency variability, in turn leading to a typical "small-sample" problem. An advantage of these procedures is that statistical inference is straightforward and is applicable to both weakly and highly persistent time series, with little requirements regarding error and model assumptions.

4.1 Long-Run Covariability

We examine the long-run covariability of temperatures and CO_2 emissions by computing the low-frequency correlation ρ_L and the bivariate regression coefficient β_L measuring the long-run relationship between temperatures and CO_2. From visual inspection of both series, we will assume that our time span of interest is smaller than 150 (150,000 years) and greater than 75 (75,000 years) datapoints. This then corresponds to a range of about $q = 10$ to $q = 21$ cosine transforms, respectively. The maximum span is by default 1,596,000 years. Thus, we take an agnostic view as to what the "long run" is by not imposing a specific value for q, not least because, beyond Milankovitch cyclicality, there is no consensus in climate science on what this should be. Hence, we present results obtained for $q = 10, ..., 21$ in Table 5.

From these results, we observe that there is strong evidence that CO_2 and temperature are

Table 5: Long-Run (LR) Covariation Measures for CO_2 and Temperature

q	LR correlation (ρ_L)	LR regression coefficient (β_L)	LR regression st. error (σ_L)
10	0.865 [0.450, 0.960]	0.068 [0.025, 0.090]	0.395 [0.279, 0.809]
11	0.828 [0.500, 0.947]	0.068 [0.026, 0.093]	0.447 [0.313, 0.811]
12	0.841 [0.450, 0.952]	0.069 [0.028, 0.094]	0.447 [0.317, 0.864]
13	0.841 [0.450, 0.947]	0.072 [0.028, 0.096]	0.464 [0.332, 0.866]
14	0.865 [0.550, 0.953]	0.077 [0.040, 0.100]	0.498 [0.364, 0.825]
15	0.877 [0.637, 0.959]	0.082 [0.048, 0.102]	0.516 [0.382, 0.874]
16	**0.945 [0.900, 0.980]**	**0.090 [0.069, 0.100]**	**0.533 [0.386, 0.921]**
17	0.937 [0.850, 0.974]	0.092 [0.073, 0.104]	0.621 [0.459, 1.013]
18	0.940 [0.850, 0.974]	0.093 [0.072, 0.105]	0.627 [0.455, 1.005]
19	0.943 [0.850, 0.974]	0.094 [0.072, 0.105]	0.628 [0.463, 1.029]
20	0.943 [0.850, 0.974]	0.095 [0.075, 0.106]	0.643 [0.475, 1.033]
21	0.937 [0.850, 0.974]	0.095 [0.074, 0.107]	0.663 [0.496, 1.123]

Notes: x is CO_2 and y is temperature; results based on the Posterior Median and a Coverage Probability of 0.90 in square brackets; ρ_L is the long-run correlation coefficient, β_L is the long-run slope coefficient and σ_L the standard error.

highly correlated over the long run. Indeed, the in-sample long-run correlation $\widehat{\rho}_L$ of the two series is positive and very large – always above 0.83, achieving a maximum of 0.95 with $q = 16$ and remaining high thereafter, with the confidence intervals noticeably narrowing for $q > 15$.[6] The estimated long-run regression coefficient $\widehat{\beta}_L$ is, as expected, positive, thus confirming that an increase in CO_2 emissions of 1ppm is associated with an estimated increase in temperatures in the long-run. Note, however, that the slope coefficient is very sensitive to the choice of q, monotonically increasing from $0.068°C$ for $q = 10$ to $0.095°C$ with $q = 21$.

Even though the above results give us a reasonable range of values, for the sake of simplicity, we define a criterion that allows us to select the most reasonable long-run measures: we pick q such that the estimated long-run correlation is the largest and the range for the 90% confidence set is the narrowest. For the pair (CO_2, Temperature) this corresponds to $q = 16$ (in bold, Table 5), implying a periodicity smaller than 99.75 (i.e. of about 100,000 years). Although admittedly ad-hoc, the basic idea is to pick the q for which the data seems to be most informative. Note that for $q < 16$, the 90% confidence sets for the correlation and the regression coefficient are substantially wider, therefore suggesting that this strategy is a sensible one.

[6]To put it in perspective, it is worth mentioning that in their empirical applications with macro variables, Müller & Watson (2018) found for fixed q only a few number of pairs of series with such a large degree of covariability. Despite the differences in terms of the observed data, it seems that climate data has a more significant long-term covariability compared to macroeconomic data.

Thus, when $q = 16$, the long-run correlation equals 0.945 with a corresponding 90% confidence set of $[0.900, 0.980]$, while the regression coefficient is 0.090, with a standard error 0.533. This value is somewhat large, especially when compared to the ones found in the literature – c.f. Castle & Hendry, 2020, who estimated an impact of $0.060°C$. The lower bound of the 90% confidence set [0.069, 0.100] is close to the Castle & Hendry (2020) estimate, but does not include it.

For the sake of completeness, we also computed the long-run measures for the bivariate relationships of ice volume with temperatures or CO_2 emissions, in Table 6. As in the previous case, $q < 16$ delivers wide confidence sets for the pair {Ice, Temperature}. The in-sample long-term correlation for $q = 16$ equals -0.871 with a confidence set of $[-0.946, -0.638]$ (similar results are obtained for $q > 16$) and an estimated long-run regression coefficient of -4.991. For {Ice, CO_2} the correlation is also large (-0.918 with an interval of $[-0.945, -0.800]$) and the regression coefficient is -55.343, highlighting the long-run association between higher CO_2 emissions and reduced ice volumes.

Table 6: Long-Run Covariation Measures for Ice-Temperature and Ice-CO_2

q	LR correlation (ρ_L)	LR regression coefficient (β_L)	LR regression st. error (σ_L)
Ice and Temperature			
16	-0.871 [-0.946, -0.638]	-4.991 [-6.190, -3.745]	0.814 [0.608, 1.196]
Ice and CO_2			
22	-0.918 [-0.945, -0.800]	-55.343 [-66.609, -44.553]	9.891 [7.080, 14.917]

Notes: In the first row x is Ice and y is temperature, in the second x is Ice and y is CO_2; results based on the Posterior Median and a Coverage Probability of 0.90 in square brackets

The value of 0.090 for the long run coefficent relating temperatures to CO_2 and the corresponding 90% confidence set are probably overestimated. The most likely explanation is that the standard bivariate Müller & Watson (2018) approach is not taking into account the effect of exogenous variables on temperatures, namely orbital forcing and non-linear functions thereof. Therefore, in order to be able to assess the long-run covariability between temperature and CO_2 emissions without omitting orbital forcing, we extend the Müller & Watson (2018) approach so that these effects are partialled out – i.e. we first regress CO_2 on additional forcings, then temperatures on the residuals for the first-stage partialling out.[7] We do so because of the "small sample" problem induced by using only a small number of trigonometric projections. To avoid compounding this problem by adding additional regressors, and given that in our application we can safely treat orbital forcing as exogenous, we suggest this approach as a convenient way of addressing this issue.

[7]It is straightforward to show that a a variant of the Frisch-Waugh-Lovell also holds in the case of the MW framework (results available upon request).

The choice of the set of exogenous regressors z_t is not clear-cut, therefore we follow the existing literature and try different options to check the robustness and sensitivity of the results to the choice made. As in Castle and Hendry (2020), we assume that CO_2 depends, at most, on lagged temperatures and ice volume and the current and lagged non-linear impacts of the exogenous orbital variables, i.e., the largest set for z_t is

$$z_t = (Temp_{t-1}, Ice_{t-1}, O_t, O_{t-1}), \text{ where} \tag{24}$$

$$O = (Ec, Ob, Pr, EcOb, EcPr, PrOb, Ec^2, Ob^2, Pr^2). \tag{25}$$

For comparison, we also consider smaller sets by dropping those variables that were not found statistically significant in explaining CO_2 levels in previous studies.[8]

In Table 7, we present the long-run covariability measures for temperatures and (distinct) CO_2 residuals for the "optimal" q, in the sense that the correlation is the largest and the range for the confidence set is the narrowest.[9] First, we note that the selected q is always 16, the same value as in the simple bivariate case. Then, the results differ depending on whether lagged temperatures and ice volume as determinants for CO_2 emissions are included or not. If included, we only find a significant long-run relationship between temperatures and partialled out CO_2 for a confidence set of 67%. Even in that case, the long-run correlation is not large (0.317) and the long-run regression coefficient equals 0.108. Once we include only the exogenous orbital variables, the results become very consistent. The long-run correlation is large (around 0.8) and the long-run regression coefficient is about 0.076 with a 90% confidence set of [0.048, 0.103].

The main takeaways are that 1) regardless of whether or not orbital effects are partialled out, the long-run correlation between temperatures and CO_2 emissions is very large (above 0.7); 2) with respect to the long-run impact of CO_2 emissions on temperature levels, we obtain larger point estimates than those reported in Castle & Hendry (2020), which can be rather interpreted as a lower bound. Castle & Hendry (2020) also obtain a solved long-run coefficient of $0.066°C$ increase in temperatures for a 1 ppm increase in the CO_2 emissions, whereas our long-run estimated impact ranges from about $0.070°C$ to $0.100°C$, depending on the exact specification.

Moreover, as a complementary alternative to partialling out, we also consider long-run co-variability measures between temperature and a common factor f_t^w capturing the main dynamics of all (or part of) the other variables in the model, i.e. CO_2, ice volume and orbital variables. Following Castle and Hendry (2020), we assume that temperatures in period t are explained by the set of variables $w_t = (Temp_{t-1}, CO_{2t}, Ec_t, EcOb_t, EcOb_{t-1}, EcPr_t)$ and formulate a factor model in which low-frequency measures of $Temp_t$ are related with the main common factor of w_t. In Table 8, we present the results for $\widehat{f}_t^w$ extracted from w_t, also considering different subsets

[8] Interestingly enough, we did not find important differences in the residuals from $CO_{2t} = \theta' z_t + v_t$ compared to the observed CO_2 emissions.

[9] Results for other values of q produce qualitatively and quantitatively similar results, and are available upon request.

Table 7: Long-Run Covariation Measures for v_t and $Temperature$

q	LR correlation (ρ_L)	LR regression coefficient (β_L)	LR regression st. error (σ_L)
$z_t = (Temp_{t-1}, Ice_{t-1}, O_t, O_{t-1})$			
16	0.317 [-0.129, 0.539]	0.108 [-0.014, 0.254]	1.680 [1.265, 2.440]
	0.317 (0.131, 0.443)	0.108 (0.036, 0.188)	1.680 (1.413, 2.065)
$z_t = (O_t, O_{t-1})$			
16	0.744 [0.502, 0.917]	0.077 [0.048, 0.105]	1.009 [0.762, 1.479]
$z_t = O_t$			
16	0.744 [0.511, 0.917]	0.078 [0.048, 0.106]	1.012 [0.847, 1.245]
$z_t = \left(Ec, Ob, EcOb, EcPr, Ob^2\right)_t$			
16	0.804 [0.564, 0.918]	0.076 [0.048, 0.103]	0.980 [0.733, 1.425]
$z_t = (Ec_t, Ec_{t-1}, Ob_{t-1}, EcOb_{t-1}, Ob_t^2)$			
16	0.798 [0.512, 0.917]	0.075 [0.046, 0.102]	1.002 [0.752, 1.463]

Notes: Results based on the Posterior Median and a Coverage Probability of 0.90 in square brackets or 0.67 in parentheses.

of its elements, using $q = 16$ as before (results are similar for any $q > 15$).

First, we note that the percentage of the variance (PTV) explained by the principal component is almost 100%, even for the largest of sets w_t. On the other hand, analysing the corresponding factor loadings (displayed in $\widehat{\lambda}$), the loading associated with CO_{2t} is essentially one, while the other loadings are quite small and close to zero. These two results together mean that the principal component $\widehat{f}_t^w$ captures the dynamic features of CO_{2t} levels after extracting the information from the remaining variables. Thus, in essence, we accomplish the goal of obtaining low-frequency estimates of the relationship between temperatures and CO_2 levels in a much similar way to the partialling out procedure expounded above.

Thus, considering the results so far, our estimates point to long run temperature increases between $6°C$ and $10°C$. These are in the neighbourhood of, but larger than, the estimates of Castle & Hendry (2020), Kaufmann & Juselius (2013) and Knutti *et al.* (2017)).

4.2 Low-Frequency Factor Model and Long-Term Forecasts

In the previous section, we focused on the long-run relationship between temperatures and CO_2 concentration levels – first, a stripped down low-frequency bivariate analysis, followed by approaches that allow us to control for exogenous forcing. In this section we turn our attention to modelling the joint dynamics of the whole climate system by employing the low-frequency factor models developed by Müller & Watson (2021), which, as a useful by-product, allow us to compute

Table 8: Long-Run Covariation Measures for $\widehat{f}_t^w$ and $Temperature$

q	LR correlation (ρ_L)	LR regression coefficient (β_L)	LR regression st. error (σ_L)
$w_t = (Temp_{t-1}, CO_{2t}, Ec_t, EcOb_t, EcOb_{t-1}, EcPr_t)$			
$PTV = 97.688$; $\widehat{\lambda} = (0.102, 0.994, 0.000, 0.003, 0.003, 0.027)$			
16	0.945 [0.900, 0.980]	0.090 [0.069, 0.099]	0.528 [0.383, 0.912]
$w_t = (CO_{2t}, Ec_t, EcOb_t, EcPr_t)$			
$PTV = 97.914$; $\widehat{\lambda} = (0.999, 0.000, 0.003, 0.027)$			
16	0.945 [0.900, 0.980]	0.090 [0.069, 0.100]	0.532 [0.386, 0.920]
$w_t = (CO_{2t-1}, CO_{2t}, Ice_{t-1}, Ice_t)$			
$PTV = 98.712$; $\widehat{\lambda} = (0.706, 0.708, -0.010, -0.010)$			
16	0.943 [0.900, 0.974]	0.063 [0.048, 0.071]	0.558 [0.404, 0.962]
$w_t = (CO_{2t}, Ice_t)$			
$PTV = 99.999$; $\widehat{\lambda} = (0.999, -0.015)$			
16	0.945 [0.900, 0.980]	0.090 [0.069, 0.100]	0.533 [0.387, 0.922]

Notes: Results based on the Posterior Median and a Coverage Probability of 0.90 in square brackets. PTV is the percentage of the total variance explained by the principal component. $\widehat{\lambda}$ is the principal component coefficients (loadings)

long-range forecasts.[10] Indeed, given the very low-frequency nature of our data, as well as the mixture of endogenous and exogenous variables we study, we deem this forecasting model to be more appropriate than a standard dynamic factor model. Next, we present the main features of the model and analyze the estimated factor loadings, after which we focus on long-horizon predictive distributions for the variables of interest, under relevant scenarios for CO_2 levels.

Following Müller & Watson (2021) and Müller *et al.* (2020), the model for the set of observables $x_t \in \Re^n$ is

$$x_t = \mu + \lambda f_t^x + e_t, \tag{26}$$

where f_t^x denotes the (scalar) unobserved common factor(s), $\lambda = (\lambda_1, ..., \lambda_n)$ contains the factor loadings, e_t represents a vector of mutually independent errors that captures the residual variability in x_t, and $\mu = (\mu_1, ..., \mu_n)$ is the intercept. Moreover, f_t^x follows a local-level model, $e_{j,t}$, $j = 1, ..., n$ stationary $I(d_j)$ models and $\{f_t^x, e_{1,t}, ..., e_{n,t}\}$ are independent. The local-level model is the sum of uncorrelated I(0) and I(1) processes with common long-run variance σ^2. The scale of f_t^x and λ are not separately identified and the factor loading λ_1 is normalized to unity. The estimation of the model is carried out by Bayesian methods, assuming the same priors for the parameters as in Müller & Watson (2021) (see paper for details). Once the posterior distribution is recovered, an additional advantage of this approach is that we can analyse how the relationships

[10] See also Müller & Watson (2016) for a discussion on long-horizon prediction intervals and Müller *et al.* (2020) for an interesting application about the long-run path of GDP for a list of 113 countries.

differ across quantiles - we consider the median, the tails at 5% and 95% and two intermediate quantiles, 17% and 83%.

We consider three models: *i*) a simple bivariate one with the main variables $x_t = (Temp_t, CO_{2t})'$ with $n = 2$; *ii*) a model that adds ice volume, i.e. $x_t = (Temp_t, CO_{2t}, Ice_t)'$, $n = 3$; and *iii*) following Castle & Hendry (2020), we consider a model with exogenous orbital variables which includes significant non-linear interactions between the drivers, such that $x_t = (Temp_t, CO_{2t}, Ice_t, Ec_t, Ob_t, EcPr_t, Ob_t^2)'$, $n = 7$.

The results are displayed in Table 9. We can see that CO_2 is a strong driver of low-frequency comovements with temperatures, particularly at the highest quantiles. Interestingly, the salience of CO_2 remains strong even when orbital variables are included. Indeed, as expected and in accordance with the Milankovich hypothesis, orbital variables capture a significant proportion of the joint variation. In particular, note that the interaction $EcPr_t$ has a loading of 1.68 – the amount of solar energy is proportional (up to a phase) to $Ec \sin Pr$ and, to first order, $\sin x \approx x$ which explains the loading. Nevertheless, CO_2 levels still have a non-negligible impact for quantiles 83 and 95. As anticipated, variation in ice volume goes in an opposite direction, while the joint effects of nonlinear interactions amongst the orbital variables appear to be the most significant contributors to low-frequency movements in the respective set of variables in x_t.

Table 9: Factor Loadings for the Low-Frequency Factor Model

Variables	**PostMean**	**PostQ0.05**	**PostQ0.17**	**PostQ0.50**	**PostQ0.83**	**PostQ0.95**
$x_t = (Temp_t, CO_{2t})'$; $n = 2$						
$Temp$	1.00	1.00	1.00	1.00	1.00	1.00
CO_2	9.29	7.98	8.66	9.37	9.96	10.37
$x_t = (Temp_t, CO_{2t}, Ice_t)'$; $n = 3$						
$Temp$	1.00	1.00	1.00	1.00	1.00	1.00
CO_2	9.30	8.00	8.66	9.37	9.97	10.39
Ice	-0.15	-0.19	-0.17	-0.15	-0.12	-0.11
$x_t = (Temp_t, CO_{2t}, Ice_t, Ec_t, Ob_t, EcPr_t, Ob_t^2)'$; $n = 7$						
$Temp$	1.00	1.00	1.00	1.00	1.00	1.00
CO_2	1.50	-1.78	-0.40	1.50	3.42	4.79
Ice	1.36	-1.92	-0.56	1.33	3.35	4.71
Ec	1.35	-1.99	-0.66	1.37	3.33	4.65
Ob	0.78	-1.64	-0.62	0.56	2.32	3.76
$EcPr$	1.68	-1.70	-0.34	1.59	3.63	5.43
Ob^2	1.44	-1.85	-0.49	1.43	3.36	4.74

Notes: PostQ stands for "posterior quantile".

Having documented and quantified the low-frequency relationship between temperatures and

CO_2 levels, and given the recent developments in the climate debate, it is important to understand how temperatures can evolve in the long-term for specific scenarios of CO_2 emissions. Here, we follow Castle & Hendry (2020) by conditioning on the relatively stable path of orbital variables and on a stable level of 385 ppm for CO_2 (consistent with current anthropogenically induced levels) to obtain long-term conditional forecasts for temperatures and ice volume using the low-frequency factor models estimated in the previous section. We consider the horizon of 50 observations (50,000 years, half of the lowest period associated with the optimal q as explained in section 4.1, which is a relatively long horizon period, given such extremely low frequency type of data.

In Figure 2 we have the conditional forecasts for temperatures, while Figure 3 displays projections for ice volumes, both in the case of $n = 7$. The projections for temperatures are close to the sample mean, but, crucially, predicted to increase. Turning to the long term of implications for ice volumes, we confirm the prediction that ice sheets will tend to reduce for high levels of CO_2 (naturally in tandem with the higher temperatures predicted above). See Diebold & Rudebusch (2021) for a discussion, albeit at shorter horizons.

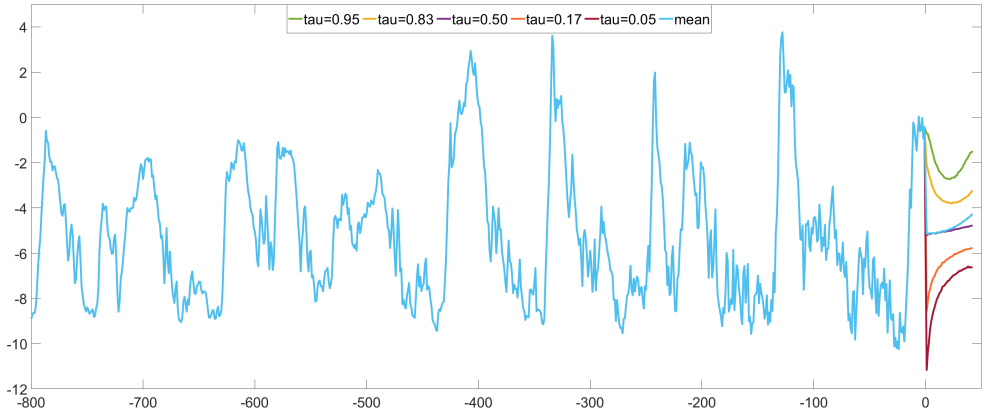


Figure 2: Conditional Forecasts of Temperatures

5 Concluding Remarks

This paper puts the spotlight on the cyclicality of paleoclimate data. On the one hand, we show that, despite point estimation remaining unaffected, ignoring persistent cycles leads to erroneous cointegration inference, a result akin to the issue of near-integrated regressors well known in the cointegration literature. On the other hand, if one is interested in equilibrium climate sensitivity long-range estimates, we suggest using the low-frequency filtering procedures of Müller & Watson (2018) and Müller & Watson (2021), which circumvent some of the pitfalls regarding the presence of persistent cycles, spurred mainly by orbital forcings, and the uncertainty around the correct order of integration of the data.

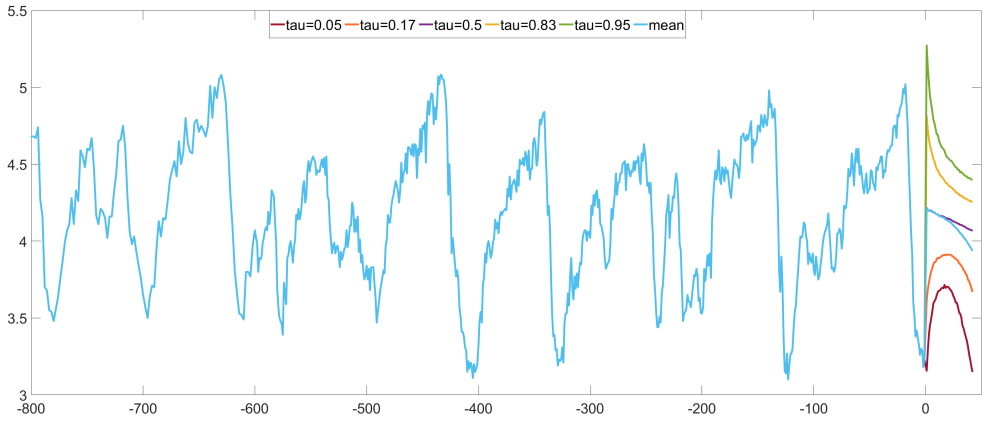


Figure 3: Conditional Forecasts of Ice Volumes

Our results confirm the strong relationship between temperatures and CO_2 for a periodicity of around 100,000 years. The long-run coefficient thus obtained validates, to a large extent, previous equilibrium climate sensitivity exercises conducted by several authors. If anything, our findings suggest that the recent simulations of Castle & Hendry (2020) are a lower bound, indicating that warming due to the increase of CO_2 concentration levels can be even more substantial. This holds even when orbital forcings are taken into account, either by partialling out or by means of an extracted common factor. Joint modelling of all time series allows us to consider a long-term scenario exercise, in which we show the extent of acceleration in temperature increases, as well as ice-sheet recession.

In this setup, the choice of q is usually left at the researchers' discretion. We proposed a "data-driven" approach, whereby we picked q such that the target long-run correlation is maximized. An interesting issue to consider is the choice of q when two or more time series display different cyclicality - it is not clear whether different q's should be selected for each series, or whether a common q would be preferable. We leave this for future research.

We should acknowledge the shortcomings of this approach. Indeed, by focusing narrowly on the long-run relationship between temperatures and CO_2, we make no attempt to model the mechanisms that link orbital forcing to temperatures, which is obviously important from a theoretical and an empirical point of view. Our first set of results suggest that orbital variables play a role, albeit minor compared to CO_2, in driving temperatures, at least in a linear setting, despite the inclusion of non-linear orbital terms. Reassuringly, the multivariate factor model of Section 4 reveals a stronger association between temperature and orbital variables, with the posterior mean estimates confirming that certain orbital variables help explain temperature variation, especially the interaction between eccentricity and precession.

Naturally, there are alternative approaches that the applied researcher may wish to consider. For instance, work by Friedrich *et al.* (2023), studying the relationship between radiative forcings

and global temperatures, employs a high-dimensional VAR that, much like in our empirical analysis, allows for the order of integration to be ignored. Still in a (structural) VAR setting, Duffy & Simons (2023) characterize cointegration through impulse response functions, thus allowing long-run relationships to be identified even if the data does not contain exact unit roots - it would certainly be interesting to analyze how paleoclimate data fares with this method. Another work that is in spirit very similar to the approach we propose here is the paper by Blazsek & Escribano (2022), who use score-driven ice-age models, which control for omitted exogenous variables and extreme events and report results very similar to Castle & Hendry (2020) (and not too distinct from our own).

Moreover, while we highlight the issue of persistent cyclicality for cointegration inference, the cyclical models we study are fairly simple and a great deal more work is required to embed them into a VECM framework. One approach that certainly deserves further attention is the recent work by Kang & Marmer (2023), which considers highly persistent stochastic cycles, potentially spanning substantial fractions of the sampling period, and therefore very close to the characteristics exhibited by paleoclimate data.

References

Baillie, R, T., & Chung, S.-K. 2002. Modeling and forecasting from trend-stationary long memory models with applications to climatology. *International Journal of Forecasting*, **18**(2), 215–226.

Bierens, H. J. 2001. Complex unit roots and business cycles: Are they real? *Econometric Theory*, **17**(5), 962–983.

Blazsek, S., & Escribano, A. 2022. Robust Estimation and Forecasting of Climate Change Using Score-Driven Ice-Age Models. *Econometrics*, **10**(6), 9.

Bloomfield, P. 1992. Trends in global temperature. *Climatic Change*, **21**(1), 1–16.

Bloomfield, P., & Nychka, D. 1992. Climate spectra and detecting climate change. *Climatic Change*, **21**(3), 275–287.

Castle, J. L., & Hendry, D. F. 2020. Climate econometrics: An overview. *Foundations and Trends in Econometrics*, **10**(3-4), 145–322.

Covey, C., Sloan, L. C., & Hoffert, M. I. 1996. Paleoclimate data constraints on climate sensitivity: The paleocalibration method. *Climatic Change*, **32**, 165–184.

Davidson, J. E. H., Stephenson, D. B., & Turasie, A. A. 2016. Time series modeling of paleoclimate data: Paleoclimate data. *Environmetrics*, **27**(1), 55–65.

Diebold, F. X., & Rudebusch, G. D. 2021. Probability assessments of an ice-free Arctic: Comparing statistical and climate model projections. *Journal of Econometrics*, Jan., S0304407620304012.

Ditlevsen, P., Mitsui, T., & Crucifix, M. 2020. Crossover and peaks in the Pleistocene climate spectrum; understanding from simple ice age models. *Climate Dynamics*, **54**, 1801–1818.

Dou, L., & Müller, U. 2021. Generalized Local-to-Unity Models. *Econometrica*, **89**, 1825–1854.

Duffy, J. A., & Simons, J. A. 2023. Cointegration without Unit Roots. *working paper*.

Elliott, G. 1998. On the Robustness of Cointegration Methods When Regressors Almost Have Unit Roots. *Econometrica*, **66**, 149–158.

Fomby, T. B., & Vogelsang, T. J. 2002. The application of size-robust trend statistics to global-warming temperature series. *Journal of Climate*, **15**(1), 117–123.

Friedrich, M., Margaritella, L., & Smeekes, S. 2023. High-Dimensional Causality for Climatic Attribution. *working paper*.

Gabriel, V. J., Martins, L. F., & Phella, A. 2023. Common persistent cycles. *manuscript*.

Gordon, A. H. 1991. Global warming as a manifestation of a random walk. *Journal of Climate*, **4**(6), 589–597.

Gordon, D C, Boudreau, P R, Mann, K H, Ong, J E, Silvert, W L, Smith, S V, Yattayakorn, G, Wulff, F, & Yanagi, T. 1996. LOICZ Biogeochemical modelling Guidelines. *LOICZ REPORTS & STUDIES NO. 5*, 104.

Imbrie, J., & Imbrie, J. Z. 1980. Modeling the climatic response to orbital variations. *Science*, **207**, 943–953.

Jaccard, S. L., Galbraith, E. D., Martínez-García, A., & Anderson, R. F. 2016. Covariation of deep Southern Ocean oxygenation and atmospheric CO2 through the last ice age. *Nature*, **530**(7589), 207–210.

Johansen, S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, **59**(6), 1551–1580.

Johansen, S. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Jouzel, J., Masson-Delmotte, V., Cattani, O., Dreyfus, G., Falourd, S., Hoffmann, G., Minster, B., Nouet, J., Barnola, J. M., Chappellaz, J., Fischer, H., Gallet, J. C., Johnsen, S., Leuenberger, M., Loulergue, L., Luethi, D., Oerter, H., Parrenin, F., Raisbeck, G., Raynaud, D., Schilt, A., Schwander, J., Selmo, E., Souchez, R., Spahni, R., Stauffer, B., Steffensen, J. P., Stenni, B., Stocker, T. F., Tison, J. L., Werner, M., & Wolff, E. W. 2007. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science*, **317**(5839), 793–796.

Kang, N., & Marmer, V. 2023. Modeling Long Cycles. *Journal of Econometrics*, **forthcoming**.

Kärner, O. 1996. Global temperature deviations as a random walk. *Journal of Climate*, **9**(3), 656–658. Publisher: American Meteorological Society.

Kaufmann, R. K., & Juselius, K. 2013. Testing hypotheses about glacial cycles against the observational record. *Paleoceanography*, **28**(1), 175–184.

Kaufmann, R. K., & Pretis, F. 2020. *Testing hypotheses about glacial dynamics and the stage 11 paradox using a statistical model of paleo-climate*. Tech. rept.

Kaufmann, R. K., & Pretis, F. 2021. Understanding glacial cycles: A multivariate disequilibrium approach. *Quaternary Science Reviews*, **251**(Jan.), 106694.

Kaufmann, R. K., Kauppi, H., Mann, M. L., & Stock, J. H. 2013. Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. *Climatic Change*, **118**(3-4), 729–743.

Kawamura, K., Parrenin, F., Lisiecki, L. E., Uemura, R., Vimeux, F., Severinghaus, J. P., Hutterli, M. A., Nakazawa, T., Aoki, S., & Jouzel, J. 2010. Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature*, **448**, 912–916.

Knutti, R., Rugenstein, M. A. A., & Hegerl, G. C. 2017. Beyond equilibrium climate sensitivity. *Nature Geoscience*, **10**(10), 727–736.

Lea, D. W. 2004. The 100 000-year cycle in tropical SST, greenhouse forcing, and climate sensitivity. *Journal of Climate*, **17**(11), 2170–2179.

Lisiecki, L. E. 2010. Links between eccentricity forcing and the 100,000-year glacial cycle. *Nature Geoscience*, **3**, 349–352.

Lisiecki, L. E., & Raymo, M. E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic 18O records. *Paleoceanography*, **20**.

Loulergue, L., Schilt, A., Spahni, R., Masson-Delmotte, V., Blunier, T., Lemieux, B., Barnola, J., Raynaud, D., Stocker, T. F., & Chappellaz, J. 2008. Orbital and millennial-scale features of atmospheric CH4 over the past 800,000 years. *Nature*, **453**(7193), 383–386.

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K., & Stocker, T. F. 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature*, **453**, 379–382.

Mann, M. E., & Lees, J. M. 2019. Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, **33**, 409–445.

Meyers, S.R., Sageman, B.B., & Pagani, M. 2008. Resolving Milankovitch: Consideration of signal and noise. *American Journal of Science*, **308**, 770–786.

Miller, J. I. 2019. Testing Cointegrating Relationships Using Irregular and Non-Contemporaneous Series with an Application to Paleoclimate Data. *Journal of Time Series Analysis*, **40**, 936–950.

Mitchell, J. M. 1976. An overview of climatic variability and its causal mechanisms. *Quaternary Research*, **6**, 481–493.

Muller, R. A., & MacDonald, G. J. 1997. Glacial cycles and astronomical forcing. *Science*, **277**, 215–218.

Müller, U. K., & Watson, M. W. 2016. Measuring uncertainty about long-run predictions. *The Review of Economic Studies*, **83**(4), 1711–1740.

Müller, U. K., & Watson, M. W. 2018. Long-run covariability. *Econometrica*, **86**(3), 775–804. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15047.

Müller, U. K., & Watson, M. W. 2021. Low-Frequency analysis of economic time series. *Handbook of Econometrics*, **7**, 117.

Müller, U. K., Stock, J. H., & Watson, M. W. 2020. An econometric model of international growth dynamics for long-horizon forecasting. *The Review of Economics and Statistics*, Oct., 1–47.

Pesaran, M. H., Shin, Y., & Smith, R. J. 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics*, **16**, 289–326.

Phillips, P. C. B. 2021. Estimation and Inference with Near Unit Roots. *Cowles Foundation Discussion Paper*, **10-22-2021**.

Proietti, T., & Maddanu, F. 2022. Modelling cycles in climate series: The fractional sinusoidal waveform process. *Journal of Econometrics*, **forthcoming**, 105299.

Rial, J. A. 1999. Pacemaking the ice ages by frequency modulation of Earth's orbital eccentricity. *Science*, **285**, 564–568.

Woodward, W. A., & Gray, H. L. 1993. Global warming and the problem of testing for trend in time series data. *Journal of Climate*, **6**(5), 953–962. Publisher: American Meteorological Society.

Woodward, W. A., & Gray, H. L. 1995. Selecting a Model for Detecting the Presence of a Trend. *Journal of Climate*, **8**(8), 1929–1937. Publisher: American Meteorological Society.

Wunsch, C. 2008. The spectral description of climate change including the 100 ky energy. *Climate Dynamics*, **20**, 353–363.

6 Appendix

6.1 Lemma A

Under Assumption 1, as $T \to \infty$,

$$\frac{1}{T}\sum_{t=1}^{T} \triangle y_t^2 \xrightarrow{p} 2\left(\sigma_x^2 + \sigma_y^2\right) \tag{27}$$

$$\frac{1}{T}\sum_{t=1}^{T} \triangle x_t^2 \xrightarrow{p} 2\sigma_x^2; \qquad \frac{1}{T}\sum_{t=1}^{T} \triangle y_t \triangle x_t \xrightarrow{p} 2\sigma_x^2 \tag{28}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) + \sigma_x^2 + \sigma_y^2 \tag{29}$$

$$\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 \xrightarrow{p} \frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) + \sigma_x^2 \tag{30}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}x_{t-1} \xrightarrow{p} \frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) + \sigma_x^2 \tag{31}$$

$$\frac{1}{T}\sum_{t=1}^{T} \triangle y_t y_{t-1} \xrightarrow{p} -\sigma_x^2 - \sigma_y^2 \tag{32}$$

$$\frac{1}{T}\sum_{t=1}^{T} \triangle x_t x_{t-1} \xrightarrow{p} -\sigma_x^2; \qquad \frac{1}{T}\sum_{t=1}^{T} \triangle x_t y_{t-1} \xrightarrow{p} -\sigma_x^2 \tag{33}$$

$$\frac{1}{T}\sum_{t=1}^{T} \triangle y_t x_{t-1} \xrightarrow{p} \frac{1}{2}\delta^2\left(1 + \frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right) - \frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) - \sigma_x^2 \tag{34}$$

6.2 Proof of Lemma A

Under Assumption 1, as $T \to \infty$, for $0 < r \leq 1$, we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} x_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} y_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_{xt} = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} u_{yt} = O_p(1)$. The results in Lemma A follow from

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} \triangle y_t^2 \\
=\ &\frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right) - \delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right) + \triangle u_{xt} + \triangle u_{yt}\right)^2 \\
=\ &\frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos^2\left(\lambda\pi\frac{t}{T}\right) + \delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right) - 2\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
&+2\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{xt} + 2\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{yt} - 2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt} - 2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{yt} \\
&\left.+ (\triangle u_{xt})^2 + (\triangle u_{yt})^2 + 2\triangle u_{xt}\triangle u_{yt}\right] \\
&\xrightarrow{p} \frac{1}{2}\delta^2\left(1 + \frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right) + \frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) - \delta^2\left(1 + \frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right) + 2\sigma_x^2 + 2\sigma_y^2 \\
=\ &\frac{1}{2}\delta^2\left(\frac{\sin(2\lambda\pi) - 2\cos(\lambda\pi)\sin(\lambda\pi)}{2\lambda\pi}\right) + 2\sigma_x^2 + 2\sigma_y^2 = 2\left(\sigma_x^2 + \sigma_y^2\right);
\end{aligned}$$

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\triangle x_t^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+\triangle u_{xt}\right)^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos^2\left(\lambda\pi\frac{t}{T}\right)+\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)-2\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
& \left.+2\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{xt}-2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}+(\triangle u_{xt})^2\right] \\
& \xrightarrow{p}\frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)+\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)+2\sigma_x^2 \\
= \; & \frac{1}{2}\delta^2\left(\frac{\sin(2\lambda\pi)-2\cos(\lambda\pi)\sin(\lambda\pi)}{2\lambda\pi}\right)+2\sigma_x^2=2\sigma_x^2;
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\triangle y_t\triangle x_t \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+\triangle u_{xt}+\triangle u_{yt}\right)\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+\triangle u_{xt}\right) \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos^2\left(\lambda\pi\frac{t}{T}\right)+\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)-2\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
& +2\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{xt}+\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{yt}-2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{yt} \\
& \left.+(\triangle u_{xt})^2+\triangle u_{xt}\triangle u_{yt}\right] \\
& \xrightarrow{p}\frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)+\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)+2\sigma_x^2 \\
= \; & \frac{1}{2}\delta^2\left(\frac{\sin(2\lambda\pi)-2\cos(\lambda\pi)\sin(\lambda\pi)}{2\lambda\pi}\right)+2\sigma_x^2=2\sigma_x^2;
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}y_{t-1}^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(u_{xt-1}+u_{yt-1})\right)^2 \\
= \; & \frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)+2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{yt-1}\right. \\
& \left.+u_{xt-1}^2+u_{yt-1}^2+2u_{xt-1}u_{yt-1}\right] \\
& \xrightarrow{p}\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2+\sigma_y^2;
\end{aligned}
$$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+u_{xt-1}\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)+2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+u_{xt-1}^2 \\
&\xrightarrow{p} \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2;
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} y_{t-1}x_{t-1} \\
= &\frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(u_{xt-1}+u_{yt-1})\right)\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+u_{xt-1}\right) \\
= &\frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)+2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{yt-1}\right. \\
&\left.+u_{xt-1}^2+u_{xt-1}u_{yt-1}\right] \\
&\xrightarrow{p} \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2;
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\triangle y_t y_{t-1} \\
= &\frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(\triangle u_{xt}+\triangle u_{yt})\right)\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(u_{xt-1}+u_{yt-1})\right) \\
= &\frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)-\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
&+\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}+\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{yt} \\
&+\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{yt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{yt-1} \\
&\left.+\triangle u_{xt}u_{yt-1}+\triangle u_{yt}u_{yt-1}+\triangle u_{xt}u_{xt-1}+\triangle u_{yt}u_{xt-1}\right] \\
&\xrightarrow{p} \frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)-\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\sigma_x^2-\sigma_y^2=-\sigma_x^2-\sigma_y^2;
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\triangle x_t x_{t-1} \\
= &\frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+\triangle u_{xt}\right)\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+u_{xt-1}\right) \\
= &\frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)-\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
&\left.+\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}+\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+\triangle u_{xt}u_{xt-1}\right] \\
&\xrightarrow{p} \frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)-\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\sigma_x^2=-\sigma_x^2;
\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\triangle y_t x_{t-1} \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(\triangle u_{xt}+\triangle u_{yt})\right)\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+u_{xt-1}\right) \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)-\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
& +\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}+\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{yt} \\
& \left.\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}+\triangle u_{xt}u_{xt-1}+\triangle u_{yt}u_{xt-1}\right] \\
& \xrightarrow{p}\frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)-\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\sigma_x^2;
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\triangle x_t y_{t-1} \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left(\delta\cos\left(\lambda\pi\frac{t}{T}\right)-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+\triangle u_{xt}\right)\left(\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)+(u_{xt-1}+u_{yt-1})\right) \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left[\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)-\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)\right. \\
& +\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}+\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1} \\
& \left.+\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{yt-1}-\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{yt-1}+\triangle u_{xt}u_{yt-1}+\triangle u_{xt}u_{xt-1}\right] \\
& \xrightarrow{p}\frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)-\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)-\sigma_x^2=-\sigma_x^2.
\end{aligned}
$$

These results follow from simple convergence of moments. Namely, as $T\to\infty$,

$$
\frac{1}{T}\sum_{t=1}^{T}u_{xt-1}^2\xrightarrow{p}V(u_{xt-1})=\sigma_x^2;\ \frac{1}{T}\sum_{t=1}^{T}(\triangle u_{xt})^2\xrightarrow{p}V(\triangle u_{xt})=2\sigma_x^2
$$

$$
\frac{1}{T}\sum_{t=1}^{T}\triangle u_{xt}\triangle u_{yt}\xrightarrow{p}Cov(\triangle u_{xt},\triangle u_{yt})=0;\ \frac{1}{T}\sum_{t=1}^{T}u_{xt-1}u_{yt-1}\xrightarrow{p}Cov(u_{xt-1},u_{yt-1})=0
$$

$$
\frac{1}{T}\sum_{t=1}^{T}\triangle u_{xt}u_{xt-1}\xrightarrow{p}Cov(\triangle u_{xt},u_{xt-1})=-\sigma_x^2;\ \frac{1}{T}\sum_{t=1}^{T}\triangle u_{xt}u_{yt-1}\xrightarrow{p}Cov(\triangle u_{xt},u_{yt-1})=0.
$$

Those involving the trignometric functions are cumbersome but trivial. For example,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\delta^2\cos^2\left(\lambda\pi\frac{t}{T}\right) \\
= \;& \delta^2\frac{1}{T}\left(\frac{T}{2}+\frac{\cos\left((T+1)\frac{\lambda\pi}{T}\right)\sin\left(T\frac{\lambda\pi}{T}\right)}{2\sin\left(\frac{\lambda\pi}{T}\right)}\right)=\delta^2\left(\frac{1}{2}+\frac{\cos\left(\lambda\pi+\frac{\lambda\pi}{T}\right)\sin(\lambda\pi)}{2\sin\left(\frac{1}{T}\lambda\pi\right)/(1/T)}\right) \\
\to \;& \frac{1}{2}\delta^2\left(1+\frac{\cos(\lambda\pi)\sin(\lambda\pi)}{\lambda\pi}\right)
\end{aligned}
$$

because $\sin\left(\lambda\pi\frac{1}{T}\right)/(1/T) \to \lambda\pi$ as $T \to \infty$. Similar results are obtained for $\frac{1}{T}\sum_{t=1}^{T}\delta^2\cos^2\left(\lambda\pi\frac{(t-1)}{T}\right)$ and $\frac{1}{T}\sum_{t=1}^{T}2\delta^2\cos\left(\lambda\pi\frac{t}{T}\right)\cos\left(\lambda\pi\frac{(t-1)}{T}\right)$. Moreover,

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}2\delta\cos\left(\lambda\pi\frac{t}{T}\right)\triangle u_{xt} \\
= \; & 2\delta\left(\frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} - \frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{(t-1)}{T}+\lambda\pi\frac{1}{T}\right)u_{xt-1}\right) \\
= \; & 2\delta\left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} - \cos\left(\lambda\pi\frac{1}{T}\right)\frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1} \\ +\sin\left(\lambda\pi\frac{1}{T}\right)\frac{1}{T}\sum_{t=1}^{T}\sin\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}\end{array}\right) \\
= \; & 2\delta\left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} \\ -\cos\left(\lambda\pi\frac{1}{T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} + \frac{1}{T}\cos(0)u_{x0} - \frac{1}{T}\cos(\lambda\pi)u_{xT}\right) \\ +\sin\left(\lambda\pi\frac{1}{T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sin\left(\lambda\pi\frac{t}{T}\right)u_{xt} + \frac{1}{T}\sin(0)u_{x0} - \frac{1}{T}\sin(\lambda\pi)u_{xT}\right)\end{array}\right) \\
= \; & 2\delta\left(\begin{array}{c} \frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} \\ -\cos\left(\lambda\pi\frac{1}{T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt} + \frac{1}{T}u_{x0} - \frac{1}{T}\cos(\lambda\pi)u_{xT}\right) \\ +\sin\left(\lambda\pi\frac{1}{T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sin\left(\lambda\pi\frac{t}{T}\right)u_{xt} - \frac{1}{T}\sin(\lambda\pi)u_{xT}\right)\end{array}\right) \\
= \; & 2\delta\left(\begin{array}{c} \cos(\lambda\pi)\frac{1}{T}\sum_{t=1}^{T}u_{xt} + (\lambda\pi)\int_0^1\sin(\lambda\pi r)\left(\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}u_{xt}\right)dr \\ -\cos\left(\lambda\pi\frac{1}{T}\right)\left(\cos(\lambda\pi)\frac{1}{T}\sum_{t=1}^{T}u_{xt} + (\lambda\pi)\int_0^1\sin(\lambda\pi r)\left(\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}u_{xt}\right)dr + \frac{1}{T}u_{x0} - \frac{1}{T}\cos(\lambda\pi)u_{xT}\right) \\ +\sin\left(\lambda\pi\frac{1}{T}\right)\left(\sin(\lambda\pi)\frac{1}{T}\sum_{t=1}^{T}u_{xt} - (\lambda\pi)\int_0^1\cos(\lambda\pi r)\left(\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}u_{xt}\right)dr - \frac{1}{T}\sin(\lambda\pi)u_{xT}\right)\end{array}\right) \\
\to \; & 0.
\end{aligned}
$$

See Bierens (1994, Lem. 9.6.3, p. 200). A similar proof is done for $\frac{1}{T}\sum_{t=1}^{T}2\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)\triangle u_{xt}$, $\frac{1}{T}\sum_{t=1}^{T}\delta\cos\left(\lambda\pi\frac{t}{T}\right)u_{xt-1}$, and $\frac{1}{T}\sum_{t=1}^{T}\delta\cos\left(\lambda\pi\frac{(t-1)}{T}\right)u_{xt-1}$.

6.3 Proof of Lemma 1

From Lemma A,

$$
\begin{aligned}
S_{00,T} &= \frac{1}{T}\sum_{t=1}^{T}\triangle z_t\triangle z_t' = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\triangle y_t^2 & \triangle y_t\triangle x_t \\ \triangle x_t\triangle y_t & \triangle x_t^2\end{bmatrix} \\
&\xrightarrow{p} 2\begin{bmatrix}(\sigma_x^2+\sigma_y^2) & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2\end{bmatrix} = 2\Upsilon;
\end{aligned}
$$

$$
\begin{aligned}
S_{11,T} &= \frac{1}{T}\sum_{t=1}^{T}z_{t-1}z_{t-1}' = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}y_{t-1}^2 & y_{t-1}x_{t-1} \\ x_{t-1}y_{t-1} & x_{t-1}^2\end{bmatrix} \\
&\xrightarrow{p} \begin{bmatrix}\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2+\sigma_y^2 & \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2 \\ \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2 & \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\end{bmatrix} \\
&= \frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota+\Upsilon;
\end{aligned}
$$

and

$$\begin{aligned} S_{01,T} &= \frac{1}{T}\sum_{t=1}^{T}\triangle z_t z_{t-1}' = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \triangle y_t y_{t-1} & \triangle y_t x_{t-1} \\ \triangle x_t y_{t-1} & \triangle x_t x_{t-1} \end{bmatrix} \\ &\xrightarrow{p} \begin{bmatrix} -\sigma_x^2-\sigma_y^2 & -\sigma_x^2 \\ -\sigma_x^2 & -\sigma_x^2 \end{bmatrix} = -\Upsilon, \end{aligned}$$

where $\Upsilon = \begin{bmatrix} (\sigma_x^2+\sigma_y^2) & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 \end{bmatrix}$ and $\iota = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

6.4 Proof of Theorem 2

We start by sketching the proof of Theorem 1. Following the Lemma 2 in Andersson et al. (1983) and our Lemma 1, the ordered solutions $\widehat{\theta}_1 \geq \widehat{\theta}_2$ of the generalized eigenvalue problem

$$\det\left[\theta S_{11,T} - S_{10,T}S_{00,T}^{-1}S_{01,T}\right] = 0$$

converge in probability to the ordered solutions $\theta_1 \geq \theta_2$ of

$$\begin{aligned} \det\left[\theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota+\Upsilon\right)-\Upsilon'(2\Upsilon)^{-1}\Upsilon\right] &= 0 \Leftrightarrow \\ \det\begin{bmatrix} \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+(\sigma_x^2+\sigma_y^2)\right)-\frac{1}{2}(\sigma_x^2+\sigma_y^2) & \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 \\ \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 & \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 \end{bmatrix} &= 0 \Leftrightarrow \\ \det\begin{bmatrix} \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2+(\theta\sigma_y^2-\frac{1}{2}\sigma_y^2) & \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 \\ \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 & \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2 \end{bmatrix} &= 0 \Leftrightarrow \\ \det\begin{bmatrix} \Delta+(\theta\sigma_y^2-\frac{1}{2}\sigma_y^2) & \Delta \\ \Delta & \Delta \end{bmatrix} &= 0, \end{aligned}$$

where $\Delta = \theta\left(\frac{1}{2}\delta^2\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+\sigma_x^2\right)-\frac{1}{2}\sigma_x^2$. This is clearly very different to the standard Johansen's cointegration results. The determinant equals $\Delta^2+(\theta\sigma_y^2-\frac{1}{2}\sigma_y^2)\Delta-\Delta^2 = (\theta\sigma_y^2-\frac{1}{2}\sigma_y^2)\Delta$. The two roots do not depend on the parameter σ_y^2 and are given by

$$\theta_1 = \frac{1}{2} \text{ and } \theta_2 = \frac{1}{\frac{\delta^2}{\sigma_x^2}\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)+2} < \frac{1}{2} = \theta_1,$$

for all admissable $\sigma_x^2, \lambda, \delta$ because $\frac{\delta^2}{\sigma_x^2}\left(1+\frac{\sin(2\lambda\pi)}{2\lambda\pi}\right) > 0$. For any λ such that $\frac{\sin(2\lambda\pi)}{2\lambda\pi} = 0$ we have $\theta_2 = \frac{1}{\frac{\delta^2}{\sigma_x^2}+2}$. That is true for integer $\lambda > 2$. Again, contrary to the standard cointegration approach, there are no zero solutions and, in particular, the largest $\widehat{\theta}$ converge in probability to $\frac{1}{2}$. Moreover, since $\theta_2 \neq 0$, $T\widehat{\theta}_2$ will not converge in distribution to a random variable but rather tends to infinity and therefore the Johansen's tests will also diverge to infinity as $T \to \infty$.

Now, we can discuss the limiting properties of the estimator (Theorem 2). The ML estimator for β is the eigenvector associated with the largest eigenvalue $\widehat{\theta}$:

$$S_{10,T}S_{00,T}^{-1}S_{01,T}\widehat{\beta} = \widehat{\theta}S_{11,T}\widehat{\beta}.$$

The estimator $\widehat{\beta}$ converges in probability to the eigenvector associated with the largest eigenvalue $\theta_1 = \frac{1}{2}$ and therefore $p\lim_{T\to\infty} \widehat{\beta} = b$ where b is the solution to

$$p\lim_{T\to\infty} S_{10,T} S_{00,T}^{-1} S_{01,T} b = \theta p\lim_{T\to\infty} S_{11,T} b.$$

It can be easily shown that the normalized b is given by $b = (1, -1)'$, which is the true coefficient in the misspecified bivariate cointegrated ECM. To prove this, we find that the solution b for $\theta_1 = \frac{1}{2}$ satisfies

$$\frac{1}{2}\delta^2 \left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)(b_1 + b_2) = 0,$$

so that for the normalized $b_1 = 1$, we obtain $b_2 = -1$.

For the estimator $\widehat{\alpha}$ we have

$$\widetilde{\alpha} = S_{01,T}\widetilde{\beta}\left(\widetilde{\beta}' S_{11,T}\widetilde{\beta}\right)^{-1},$$

where $\widetilde{\beta}$ is the normalized MLE of $\widehat{\beta}$ (defined below), which converges to

$$-\Upsilon b\left(b'\left(\frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota + \Upsilon\right)b\right)^{-1} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \alpha.$$

For

$$\widehat{\Omega} = S_{00,T} - \widetilde{\alpha}\widetilde{\beta}' S_{10,T},$$

it converges to

$$2\Upsilon + \alpha b'\Upsilon = \begin{bmatrix} 2\sigma_x^2 + \sigma_y^2 & 2\sigma_x^2 \\ 2\sigma_x^2 & 2\sigma_x^2 \end{bmatrix} \neq \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_x^2 \end{bmatrix}.$$

In terms of the limiting distribution of $\widehat{\beta}$ notice first that the matrix $S_{11,T} = \frac{1}{T}\sum_{t=1}^{T} z_{t-1} z_{t-1}'$ is $O_p(1)$, something distinct from the I(1) case where $\frac{1}{T}S_{11,T} = O_p(1)$. In standard cointegration, the MLE $\widehat{\beta}$ of β is normalized as $\widetilde{\beta} = \widehat{\beta}\left(\beta'\widehat{\beta}\right)^{-1}\beta'\beta$ and one can write $\widetilde{\beta} - \beta = \beta_\perp U_T$, where $\beta_\perp'\beta = 0$ and $U_T = (\beta_\perp'\beta_\perp)^{-1}\left(\beta_\perp'\widehat{\beta}\right)\left(\beta'\widehat{\beta}\right)^{-1}(\beta'\beta)$, with an expansion

$$T.U_T = \left(T^{-1}\beta_\perp' S_{11,T}\beta_\perp\right)^{-1}\beta_\perp'\left(S_{10,T} - S_{11,T}\beta\alpha'\right)\Omega^{-1}\alpha\left(\alpha'\Omega^{-1}\alpha\right)^{-1} + o_p(1)$$

from which the distribution is derived. But in our setup, this can not be done this way because $S_{11,T}$ has a different order of integration. From the previous results, as $T \to \infty$,

$$\left(\widetilde{\beta}' S_{11,T}\widetilde{\beta}\right)^{-1} \xrightarrow{p} \left(b'\left(\frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota + \Upsilon\right)b\right)^{-1} = \frac{1}{\sigma_y^2}$$

and

$$S_{11,T}\widetilde{\beta} \xrightarrow{p} \left[\frac{1}{2}\delta^2\left(1 + \frac{\sin(2\lambda\pi)}{2\lambda\pi}\right)\iota + \Upsilon\right]\beta = -\sigma_y^2\alpha.$$

So $\widetilde{\beta}$ converges in distribution at a rate $\sqrt{T}$ instead of T. Deriving the expression of the limiting distribution for constant or time-varying δ_t is a very complicated exercise and, for that reason, we rest on the results we find in the Monte Carlo section.