Testing Cointegrating Relationships Using Irregular and Non-Contemporaneous Series with an Application to Paleoclimate Data

J. Isaac Miller*

This Draft: January 29, 2019

Abstract

Time series that are observed neither regularly nor contemporaneously pose problems for most multivariate analyses. Common and intuitive solutions to these problems include interpolation and other types of imputation to a higher, regular frequency. However, interpolation is known to cause serious problems with the size and power of statistical tests. Due to the difficulty in dating paleoclimate data such as CO₂ concentrations and surface temperatures, time series of such measurements are observed neither regularly nor contemporaneously. This paper presents large- and small-sample analyses of the size and power of cointegration tests of time series with these features and supports the robustness of cointegration of these two series found in the extant literature. Compared to linear or higher-order polynomial interpolation, step interpolation results in the least size distortion and is therefore recommended.

JEL Classification: C12, C22, C32, Q54
Key words and phrases: cointegration, irregularly spaced time series, non-contemporaneous time series, paleoclimate data

*Department of Economics, University of Missouri, Columbia, Missouri, USA. ORCID: 0000-0003-1756-5502. Contact: millerjisaac@missouri.edu.
1 Introduction

Multivariate time series analyses typically rely on data observed both regularly and at the same frequency. Time series observed at different frequencies, or mixed-frequency time series, the subject of this special issue, provide important examples in which not all observations are contemporaneously observed. Yet statistical approaches designed to accommodate different frequencies typically rely on some contemporaneously observed and all regularly spaced data.

The present analysis considers an even more extreme case, in which two or more time series are irregularly (or unevenly) spaced and very few or none of the observations are contemporaneously observed. In other words, the series are asynchronously observed or temporally misaligned. This case captures the spirit of mixed-frequency time series, both in the commonly used sense of referring to two or more series observed at different but regular frequencies (Miller, 2010; Ghysels and Miller, 2014, 2015; Miller and Wang, 2016; e.g.) and in the sense in which a single series may change in frequency of observation over its span (Busetti and Taylor, 2005). Methods that may be employed in this case are similar to those that may be utilized to analyze mixed-frequency time series in both of these senses.

A common example from finance is provided by daily equity prices traded on markets in different time zones (non-contemporaneous), not over the weekend (unevenly spaced), and especially in different countries with different holidays (irregular). Weekly interest rate data provide an example from macroeconomics. Specifically, the weekly federal funds rate data available from the St. Louis Fed are regular, but each week ends on Wednesday, in contrast to a week of Treasury bill rates, which ends on Friday (non-contemporaneous). A third example, from cliometrics, is provided by historical GDP data originally estimated by Angus Maddison, which extend back annually to 1950 for most countries, but irregularly as far back as 1 CE for a few countries.

Paleoclimate data provide a fourth and striking example. Of particular importance are the series of carbon dioxide (CO$_2$) concentrations in parts per million by volume (ppm or ppmv) and temperatures in degrees Celsius ($^\circ$C). The era of human development and proliferation has occurred during the Holocene, which is a relatively small part (1-2%) of the paleoclimate record. Yet understanding the relationship between CO$_2$ and both local and global temperatures over the whole record can inform climate scientists, statisticians, and social scientists about the possible effects of future anthropogenically emitted CO$_2$.

These data are acquired by drilling ice cores, primarily in Antarctica or Greenland. Although the cores can be sampled at regular depths, dating the samples results in irregular spacing. When more than one ice core is sampled, the irregularity also means that observations of the relevant measurements are not contemporaneous. The process of collecting data from ice cores is discussed in more detail in Section 3.

The principal statistical tool used to assess correlations of stochastically trending series over a long period – i.e., long-run comovement – is cointegration analysis. Cointegration analysis has been applied to paleoclimate data by Kaufmann and Juselius (2010a, 2010b, 2013, 2016), who find a cointegrating relationship between CO$_2$ concentrations and temperature. Katarina Juselius was one of the pioneers of likelihood-based analysis of cointegrated time series (Johansen and Juselius, 1990, 1992), while Robert Kaufmann was one of the
first to apply cointegration analysis to recent climate series (Stern and Kaufmann, 2000; Kaufmann and Stern, 2002). In order to overcome irregularity and non-contemporaneity, these researchers linearly interpolated the paleoclimate series.

The statistical literature on missing observations provides a wide range of alternatives to linear interpolation: both frequentist and Bayesian, and both continuous-time solutions that preserve the irregularity and discrete-time solutions aimed at evening out observations. Moreover, some of these methods have been employed explicitly in dealing with paleoclimate data. Nieto-Barajas and Sinha (2015) and Doan et al. (2015) provide good discussions of this literature.

However, to the best of the author’s knowledge, only linear interpolation has been used in cointegration analysis of these series, and all of the imputation methods – including linear interpolation – appear to be implemented univariately before joint inference is conducted. Linear interpolation is a natural solution, because it does not change the long-run relationship between the series (Miller, 2010). Yet, linear interpolation is not without consequence, because it can seriously affect tests about that relationship, as Pretis and Hendry (2013) argue and Ghysels and Miller (2014) show for series observed regularly but at different frequencies.

This paper addresses the size and power of standard cointegration tests when the data are irregularly and non-contemporaneously observed. Both linear interpolation and step interpolation – i.e., filling the data gaps using the last available observation – are considered. Step interpolation is proposed by Ryan and Giles (1999) in the context of univariate unit root testing with irregularly spaced data. Busetti and Taylor (2005) consider the related problem of univariate stationarity testing with irregularly spaced data, but their solution, involving a careful accounting of the spacing, does not work in the present context of non-contemporaneously observed series.

The present analysis reveals that linear or higher-order polynomial interpolation may have substantial consequences for the size of cointegration tests, causing over-rejection of the null of no cointegration when the series are not cointegrated. In other words, the cointegration finding of Kaufmann and Juselius (2013) comes into question. The negative effects on size are ameliorated by step interpolation, and they are further abated by increasing the order of the autoregression.

Taking into account the size distortions, the empirical evidence generally supports the cointegration of CO₂ concentrations and temperature found by Kaufmann and Juselius (2013) in the sense that the null of no cointegration is rejected. However, an additional rejection of the cointegration null against the stationary alternative either supports the stationarity finding of Davidson et al. (2016), who also use linear interpolation, or else suggests the possibility that the size distortion is not completely eliminated or that level shifts apparent in the data may be contributing additional distortion.

The rest of the paper is structured as follows. Section 2 lays out the main methodological framework for testing for cointegration and dealing with irregularly and non-contemporaneously observed series, and the main theoretical result on size distortion is presented therein. Section 3 contains the empirical application and a set of Monte Carlo simulations tailored to the application in order to demonstrate the potential magnitude of the size distortion. Section 4 briefly concludes and an appendix contains proofs of the main
and auxiliary theoretical results of Section 2.

2 Framework for Testing

Consider a bivariate time series \((z_t^*)\)\(_{t=1}^T\), decomposed as \(z_t^* = (y_t^*, x_t^*)'\). The bivariate assumption is merely to ease subsequent notation and exposition, and it is appropriate for the empirical application, but it is not intended to be a methodological limitation. Each component has a stochastic trend (unit root) and the series are related by

\[ z_t^*\alpha = \varepsilon_t^* \]  

where \(\alpha\) is normalized as \(\alpha = (1, \beta)'\).

In estimating a long-run relationship, there is no meaningful distinction between the labels \(x\) and \(y\) or their ordering in \(z_t^*\), and the normalization of the first element of \(\alpha\) is arbitrary. A constant in the cointegrating relationship will be considered subsequently in the application to paleoclimate data, but its presence affects the limiting distributions discussed below in a predictable way and is not relevant to the sampling issues of primary concern.

Although the relationship in (1) is discrete, it can be motivated by sampling from a continuous-time relationship along the lines of Phillips (1991), Corradi (1997), Chambers (2003, 2011), inter alia. In the application considered in this paper, all series are stock-sampled, so there is no size distortion resulting from different sampling schemes noted by Miller and Wang (2016), and no loss of generality in the ensuing discussion of size and power.

The series \((\varepsilon_t^*)\)\(_{t=1}^T\) has a stochastic trend if the stochastic trends of each series \((y_t^*)\) and \((x_t^*)\) are distinct – i.e., if they are not cointegrated. On the other hand, \((\varepsilon_t^*)\) has no stochastic trend if the stochastic trends of \((y_t^*)\) and \((x_t^*)\) are common – i.e., if they are cointegrated.

Assume an invariance principle of the form \(T^{-1/2} \sum_{t=1}^{[Tr]} \Delta z_t^* \to_d B(\tau)\) as \(T \to \infty\). \(B\) is a vector Brownian motion such that \(B = \Sigma^{1/2}W\), where \(W\) is a vector of independent standard Brownian motions and \(\Sigma = \Sigma^{1/2}\Sigma^{1/2}\) is the lower triangular Cholesky decomposition. For expositional simplicity and to best illustrate the size distortions, assume that the increments \(\Delta z_t^*\) are serially uncorrelated, so that \(\Sigma\) represents both the long-run and contemporaneous variances of the increments.

2.1 Cointegration Tests

Consider three common cointegration tests: the trace test and the residual-based Dickey-Fuller coefficient and \(t\)-tests. Define \(r_{0t}^* = \Delta z_t^*\) and \(r_{1t}^* = z_t^*\) and four sample moments \(R_{jk}^* = T^{-1} \sum_t r_{jt}^* r_{kt}^*\) for \(j, k = 0, 1\). Up to a negligible term, the simplest version of the trace test may be written as

\[ \hat{\psi}_T^* = \text{tr}(T^{-1}R_{11}^{-1}R_{10}^*(R_{00}^*)^{-1}R_{01}^*); \]  

(2)
which has a well-known limiting distribution given by

$$
\hat{\psi}_T^* \rightarrow_d \text{tr}\left\{ \left( \int WdW' \right)' \left( \int WW' \right)^{-1} \int WdW' \right\} = \psi,
$$

under the null of no cointegrating vectors.

In order to identify the size distortion from interpolation as clearly as possible, we follow Ghysels and Miller (2014, 2015) in ignoring reduced rank nulls, in which case \( W \) is univariate rather than bivariate, deterministic trends, in which case some of the \( W \)'s in the distribution are replaced by known functions of \( W \), and serial correlation, which requires conditioning out lags of \( \triangle z_t^* \) before estimation but does not change the limiting distributions. The reader is referred to Johansen and Juselius (1990) and Johansen (1995) for more details. Considering reduced rank nulls or allowing for deterministic components should not change the fundamental size problem. However, allowing for serial correlation ameliorates size distortion, as discussed in the subsequent empirical application.

The residual-based cointegration test statistics are based on a regression of \( y_t^* \) onto \( x_t^* \), keeping in mind that the choice of regressor and regressand is arbitrary. The relationship in (1) is rewritten as

$$
y_t^* = x_t^* \beta + \varepsilon_t^* \tag{3}
$$
to formalize the regression. Defining \( \hat{\beta} \) to be the least squares estimator of \( \beta \) and \( \hat{\alpha} = (1, -\hat{\beta})' \). The residuals from the regression are \( \hat{\varepsilon}_t^* = z_t^* \hat{\alpha} \).

The simplest versions of the residual-based tests are conducted by regressing \( \triangle \varepsilon_t^* = \triangle z_t^* \hat{\alpha} \) onto \( \varepsilon_{t-1}^* = z_{t-1}^* \hat{\alpha} \). The least squares estimator of the regression coefficient is \( \hat{\alpha}'R_{11}^*\hat{\alpha}^{-1}\hat{\alpha}'R_{10}^*\hat{\alpha} \), so that the well-known coefficient and \( t \)-test statistics may be expressed as

$$
\hat{\rho}_T^* = (\hat{\alpha}'T^{-1}R_{11}^*\hat{\alpha})^{-1}\hat{\alpha}'R_{10}^*\hat{\alpha} \quad \text{and} \quad \hat{\tau}_T^* = (\hat{\alpha}'R_{00}^*\hat{\alpha}T^{-1}R_{11}^*\hat{\alpha})^{-1/2}\hat{\alpha}'R_{10}^*\hat{\alpha} \tag{4}
$$

using this notation.

Following Phillips and Ouliaris (1990), define \( \kappa = (1, \kappa_2)' \) with \( \kappa_2 = -\left( \int W_2W_2' \right)^{-1}\int W_2W_1 \) and \( Q(r) = W_1(r) - \int W_1W_2^2(r)W_2(r) = \kappa'W(r) \) with \( W = (W_1, W_2)' \). The distributions of the test statistics are given by

$$
\hat{\rho}_T^* \rightarrow_d \left( \int Q^2 \right)^{-1} \int QdQ = \rho \quad \text{and} \quad \hat{\tau}_T^* \rightarrow_d \left( \kappa' \kappa \int Q^2 \right)^{-1/2} \int QdQ = \tau, \tag{5}
$$

under the null of no cointegration.

Size distortion results when the limiting distributions of the respective test statistics using actual data differ from \( \psi, \rho, \) and \( \tau \), rendering invalid the standard critical values based on these distributions.

### 2.2 Irregularity and Non-Contemporaneity

Now, suppose that \( (x_t^*) \) and \( (y_t^*) \) are observed only at times \( T_p \) and \( S_q \) respectively. Equivalently, suppose that only \( (x_{T_p}) \) and \( (y_{S_q}) \) are observed, where \( x_{T_p} = x_{T_p}^* \) and \( y_{S_q} = y_{S_q}^* \). There is no guarantee that \( S_q = T_p \) for any \( q \) or \( p \), so clearly the three tests statistics cannot be computed, unless the data are modified in some way.
In the related context of univariate unit root testing, Ryan and Giles (1999) consider three strategies for dealing with irregularly spaced data. Their preferred method is to ignore the spaces and just use \((x_{T_p})\) for testing, and the method of Busetti and Taylor (2005) may be viewed as an improvement. However, these methods clearly fail when trying to relate \((x_{T_p})\) to \((y_{S_q})\) when \(S_q \neq T_p\). They are simply not designed for this case.

Ryan and Giles (1999) also consider linear interpolation and step interpolation, both of which impute the observable series to a frequency of the greatest common denominator of all time increments. If the time increments are measured in years, for example, these methods may be used to impute data to an annual frequency. A step-interpolated series of all time increments. If the time increments are measured in years, for example, these series is given as

\[
\hat{x}_t^S = x_{T_p} \text{ for } t \in [T_p, T_{p+1}), \quad \text{while a linearly interpolated series } \hat{x}_t^L \\
\text{may be defined as } \hat{x}_t^L = x_{T_p} + \frac{t-T_p}{T_{p+1}-T_p}(x_{T_{p+1}} - x_{T_p}) \text{ for } t \in [T_p, T_{p+1}].
\]

Suppose for the moment that the data are non-contemporaneous but regularly spaced \(m\) units apart, so that \(q = p\) and \(m = T_{p+1} - T_p = S_{p+1} - S_p\). To fix ideas, suppose that \(T_p\) is Thursday of week \(p\) and \(S_q\) is Saturday of the same week. In that case, the series \((x_{T_p})\) is interpolated as

\[
\hat{x}_t^S = \hat{x}_{T_p+i}^S = x_{T_p} \text{ for } i = 0, \ldots, m-1 \quad \text{or} \\
\hat{x}_t^L = \hat{x}_{T_p+i}^L = x_{T_p} + \frac{i}{m}(x_{T_{p+1}} - x_{T_p}) \text{ for } i = 0, \ldots, m.
\]

using step interpolation and linear interpolation respectively. The first difference of each of these series is given as

\[
\Delta \hat{x}_{T_p+i}^S = \begin{cases} x_{T_p} - x_{T_{p-1}} \text{ for } i = 0 \\ 0 \text{ for } i = 1, \ldots, m-1 \end{cases} \quad \text{or} \\
\Delta \hat{x}_{T_p+i}^L = \frac{1}{m}(x_{T_{p+1}} - x_{T_p}) \text{ for } i = 1, \ldots, m
\]

respectively. The series \((y_{S_q})\) is similarly interpolated using these methods.

**Why Not Linear Interpolation?** Broadly speaking, issues that affect the size of these tests under the null, which may be as simple as serial correlation or much more complicated, typically manifest in the limit of \(R_{10}^*\). This limit generally contains nuisance parameters so that the distributions \(\psi, \rho\), and \(\tau\), from which critical values are drawn, are not obtained.

In the univariate unit root case of Ryan and Giles (1999), a feasible version of the key sample moment \(R_{10}^*\) is \(T^{-1} \sum_T \hat{x}_{T_p+i}^S \Delta \hat{x}_{T_p+i}^S\), which may be written as

\[
T^{-1} \sum_p x_{T_{p-1}}(x_{T_p} - x_{T_{p-1}})
\]

using step interpolation and equations (6) and (8). Similarly, for this example with regular spacing, \(T^{-1} \sum_T \hat{x}_{T_p+i}^L \Delta \hat{x}_{T_p+i}^L\) may be written as

\[
T^{-1} \sum_p x_{T_p}(x_{T_{p+1}} - x_{T_p}) + \frac{m-1}{2m} \left[ T^{-1} \sum_p (x_{T_{p+1}} - x_{T_p})^2 \right]
\]

using linear interpolation and equations (7) and (9).
The expression in (10) and the first term of that in (11) have the expected asymptotic distribution required to obtain the standard distribution of unit root tests under the unit root null. The critical difference lies in the second term of (11), which is non-negligible and therefore distorts the size of the tests. Consequently, step interpolation, which is not burdened by this term, is preferred over linear interpolation for unit root testing.

**Why Not Higher-Order Polynomial Interpretation?** Higher-order alternatives to linear interpolation, such as cubic splines, may improve the fit of the interpolated data to the latent data. However, it is unlikely that such methods would alleviate size distortion in unit root and cointegration tests, because they suffer from the same problem as linear interpolation: the sample moment in (10) is augmented by size-distorting terms similarly to that in (11). Size distortion is effectively ameliorated by eliminating this term, which is accomplished by step interpolation.

**Why Not Step Interpolation?** Step interpolation is effective in a univariate context. Is it effective in a multivariate context? Like unit root tests, the cointegration tests discussed above rely on the sample moment $R_{10}^*$. However, this sample moment is a matrix rather than a scalar, so that the limiting distribution of $T^{-1} \sum_t y^*_t \Delta x^*_t$ is important.

Consider again the example in which observations are regular. The step-interpolated analog of this moment is $T^{-1} \sum_t y^*_t \Delta x^*_t$, which may be rewritten as

$$T^{-1} \sum_p y^*_{T_p-1} (x_{T_p} - x_{T_p-1}) + T^{-1} \sum_p (y_{S_p-1} - y^*_{T_p-1}) (x_{T_p} - x_{T_p-1})$$

similarly to (10) but with an additional term, because $(y^*_t)$ is not observed at $T_{p-1}$. As in (11), the first term has the required limiting distribution, while the second is distortive. (See the proof of Lemma A2[c] in the appendix.)

The second term of (12) creates a problem. It reflects the non-contemporaneity of the two series, even though they are regularly observed in this example. Conceptually, this term embodies the correlation between the change in $y_t$ from Thursday of last week (if it could be observed) to last Saturday and the change in $x_t$ from Thursday of last week to Thursday of this week. This correlation is not generally zero, which results in size distortion of cointegration tests using step interpolation, even though step interpolation does not distort the size of unit root tests.

2.3 Large-Sample Results

The aim of this section is to present the formal asymptotic distributions of the test statistics under the null of no cointegration using step-interpolated series when observations are neither contemporaneous nor regularly spaced.

Large-sample results for linear and higher-order polynomial interpolation are not considered for two reasons. First, just as substituting $y_{S_p-1}$ into (10) in place of the first $x_{T_p}$ results in a second, size-distorting term in equation (12), so would substituting $y_{S_p-1}$ into (11), which already contains a size-distorting term. Second, in the related case with mixed-frequency data Ghysels and Miller (2014) have already shown that these tests may suffer
from extreme size distortion due to linear interpolation. In a more general context, Pretis and Hendry (2013) argued along similar lines against using linear interpolation for unit root and cointegration tests. In other words, linear and higher-order interpolation distorts size relative to step interpolation, and the present analysis focuses on the least distortive method.

In order to illustrate the problem clearly, assume that \((x^*_t)\) and \((y^*_t)\) are martingales with stationary increments. Also, assume that the number of times that \((x^*_t)\) and \((y^*_t)\) are observed contemporaneously is finite or zero as the sample size increases. Define \(\pi_{yx}^p = E(y_{S_q} - y_{S_q-1} - x_{T_p} - x_{T_p-1})\) and \(\pi_{xy}^q = E(x_{T_p} - x_{T_p-1})(y_{S_q} - y_{S_q-1})\) and assume that \(\pi_{yx}^p, \pi_{xy}^q < \infty\) exist such that \(T^{-1} \sum_p \pi_{yx}^p \rightarrow \pi_{yx}\) and \(T^{-1} \sum_q \pi_{xy}^q \rightarrow \pi_{xy}\). Further, define \(G_T = \sum_p \pi_{yx}^p G + \sum_q \pi_{xy}^q\) and \(G = \begin{bmatrix} 0 & \pi_{yx} \\ \pi_{xy} & 0 \end{bmatrix}\), so that \(G_T \rightarrow G\), and define \(F = \begin{bmatrix} \sigma_{yy} & 0 \\ 0 & \sigma_{xx} \end{bmatrix}\) to be a diagonal matrix with diagonals equal to that of \(\Sigma\).

As long as the number of observations increases at the same rate as \(T\), \(G\) is non-negligible. It is non-negligible even if \(T\) is much larger than the number of observations of each series, because the summands in \(G_T\) generally increase as the ratios of \(T\) to these observations increase. In other words, the number of high-frequency periods in the overlap between the gaps in observations of the two series increases as the gaps widen. If there is no overlap and the observations are contemporaneous, then \(G = 0\). If the number of times that \((x^*_t)\) and \((y^*_t)\) are observed contemporaneously increases proportionally with the sample size, then the zeros in \(F\) are replaced with a function of \(\sigma_{xy}\), and \(F = \Sigma\) if the proportion is one.

Define the lower diagonal matrix \(L\) such that \(\Sigma = L'L\), which is not the usual Cholesky decomposition. (See Phillips and Ouliaris, 1990.) Define \(F^+ = L^{-1/2}FL^{-1}\) and \(G^+ = L^{-1/2}GL^{-1}\) and note that \(F^+\) is an identity matrix if \(F = \Sigma\) while \(G^+\) is zero if \(G\) is also zero.

Under the preceding assumptions and notations, the following proposition holds.

**Proposition.** The three test statistics \(\hat{\psi}_M\), \(\hat{\rho}_M\), and \(\hat{\tau}_M\) for tests conducted on step-interpolated series have limiting null distributions given by

\[
[a] \quad \hat{\psi}_M \rightarrow_d \text{tr}\{(\int WdW' + G^+)(\int WdW')^{-1}(\int WdW' + G^+)(F^+)^{-1}\},
\]

\[
[b] \quad \hat{\rho}_M \rightarrow_d (\int Q^2)^{-1}(\int QdQ + \kappa'G^+\kappa), \text{ and}
\]

\[
[c] \quad \hat{\tau}_M \rightarrow_d (\kappa'F^+\kappa\int Q^2)^{-1/2}(\int QdQ + \kappa'G^+\kappa)
\]

as \(T \rightarrow \infty\).

The limits in the proposition appear similar to those in Theorem 1 of Ghysels and Miller (2014), simply because the distorting terms enter into the null distributions in the same
way. However, the matrices $F^+$ and $G^+$ are quite distinct from and more complicated than the matrices $C^+_m$ and $D^+_m$ of those authors. Using linear interpolation instead of step interpolation, the distributions would suffer not from $F^+$ and $G^+$, but from $F^+ + C^+_m$ and $G^+ + D^+_m$, although $C^+_m$ and $D^+_m$ would need to be redefined to accommodate irregular spacing.

When $F = \Sigma$ and $G = 0$, the expected limiting null distributions in (3) and (5) are obtained. Otherwise, deviations of $F$ from $\Sigma$ and $G$ from zero cause size distortions. As implied above by the discussion of $F$, such deviations are driven primarily by non-contemporaneity. Irregularity seems to be a milder problem.

The extent of the size distortion in a finite – but relatively large – sample is explored below using Monte Carlo simulations designed to mimic the paleoclimate data.

3 Application to Paleoclimate Data

Kaufmann and Juselius (2013) analyze a system of fourteen paleoclimate series gathered from various sources and reflecting various states of the planet over the past 391 thousand years known as the so-called Vostok period. Chief among these series are the temperature series of Jouzel et al. (2007)\(^1\) and the CO\(_2\) concentration series of Lüthi et al. (2008),\(^2\) dated using the EDC3 chronology of Parrenin et al. (2007).

There are a few reasons for considering only these two series. One reason is that the scientific community views CO\(_2\) concentrations as the main anthropogenic driver of temperature changes in recent years and – given its relatively strong persistence in the atmosphere compared to other greenhouse gases – likely also in future years. Understanding the relationship between these two series is important.\(^3\)

Another reason is that Kaufmann and Juselius (2013) identified one of ten cointegrating relationships exclusively between these two series. If their results are robust, similar results should hold in a system with just these two series. A third reason is that the data for these two series are available over a much longer time span by way of drilling at the Dome C location rather than the Vostok location.

Global temperature and atmospheric CO\(_2\) are inferred from measurements made from ice cores sampled at regular depths. Atmospheric CO\(_2\) concentrations are inferred directly from glaciochemical analysis. However, the procedures for estimating temperatures and dating the ice cores are more complicated.

To measure temperatures, $^{18}O$ (Oxygen-18) content and its deviation $\delta^{18}O$ from that of ocean water is determined by stable isotope analysis of sections of an ice core. Local temperature is inferred from $\delta^{18}O$ using a constant or nearly constant scaling, as illustrated by Figure 5.8 of Bradley (1999). Global temperatures may be further inferred from local temperatures by a scalar multiple (Mason-Delmotte et al., 2010).

Dating ice cores is less precise. Bradley (1999) surveys various methods available for

---

\(^1\)Downloaded from doi.pangaea.de/10.1594/PANGAEA.683655 on May 30, 2018.

\(^2\)Downloaded from www.nature.com/articles/nature06949 on May 30, 2018.

\(^3\)As with correlation, cointegration does not imply causality. In fact, changes in temperature more often lead changes in CO\(_2\) concentrations over the paleoclimate record than the other way around.
this purpose. As examples, Figure 3 of Lorius et al. (1985) and Table 1 of Parrenin et al. (2007) map drilling depth to age, measured in years before present or thousands of year before present (kyr). Although the relationships are monotonic, their nonlinearity results in irregular spacing of the series over time. Different analyses on different cores results in non-contemporaneity for the same reason.

Temperatures are expressed as deviations in °C from the average over the most recent 1 kyr. 5,788 observations are available between 801.662 kyr (801,662 years before present, where “present” is set to 1950) and 0.038 kyr. Following Mason-Delmotte et al. (2010) these deviations are muted by a factor of 1/2 in order to control for polar amplification of climate fluctuations, because the data reflect surface temperatures of the Antarctic location at which they were collected. This constant factor makes no difference in the subsequent tests. Over this time span, the ranges of Dome C and global temperature deviations are [−10.58, 5.46]°C and [−5.29, 2.73]°C respectively, suggesting a fluctuation in global temperatures of about 8°C.

CO₂ concentrations are expressed in ppmv, and because CO₂ is nearly uniform in concentration around the globe (“well-mixed”), no modification is made to the data to reflect local conditions at the drilling site. 1,096 observations are available between 798.512 kyr and 0.137 kyr. CO₂ concentrations fluctuate by 127 ppmv over the range [171.6, 298.6] ppmv over this time span. To put this range into perspective, concentrations have increased from about 285 ppmv in the 1850’s to about 406 ppmv in 2017. This increase of 121 ppmv is roughly similar to the entire range of CO₂ concentrations over the entire paleoclimate record, but it occurred over a time span about two ten thousands as long, or five thousand times shorter.

In order to test for cointegration and estimate the system, Kaufmann and Juselius (2013) use linear interpolation to deal with the irregularity and non-contemporaneity of the data. They report that the time step used is 1 kyr, which suggests that after interpolating they keep just one observation for every 1,000 interpolated – i.e., they utilize skip sampling. Ghysels and Miller (2015) show that skip sampling in this way does not affect the size of cointegration tests, but Ghysels and Miller (2014) show that linear interpolation can drastically affect the size.

In this analysis, as much data as possible is used for both step and linear interpolation. However, for purposes of comparison, the time span is limited to the shorter period over which CO₂ is observed, effectively making $T = 798,380$ years. No skip sampling to the nearest kyr is employed.

The data are standardized after interpolation by subtracting the mean of each series and dividing by the respective standard deviations, as Kaufmann and Juselius (2013) do. Figure 1 shows the resulting standardized series from step interpolation over the full span, over which the two interpolation methods give visually similar sample paths, and from both types of interpolation over a much short span since 1,000 years before present. Note that the standardization affects which asymptotic critical values should be used: Table A3 (under hypothesis $H^*_{a}$) of Johansen and Juselius (1990) for the trace test and Tables Ib and IIb (demeaned) of Phillips and Ouliaris (1990) for the residual-based tests.

---

Figure 1: **Standardized Temperature and CO₂ Concentrations.** Vertical axis: standardized units. Horizontal axis: years before present (1950). Top panel: Standardized step-interpolated series over the full span, since 798,512 years before present. Bottom panel: Standardized step-interpolated and standardized linearly interpolated series since 1,000 years before present.
3.1 Small-Sample Results

Before examining the empirical results, it is useful to further evaluate the theoretical results by way of Monte Carlo simulations tailored specifically to match some of the time series characteristics of the paleoclimate data.

First, \( (x_t^*) \) is simulated such that the variance of its increments matches the sample variance of \( (x_{T_p} - x_{T_{p-1}})/\sqrt{T_p - T_{p-1}} \) for CO2 concentrations. Next, \( (y_t^*) \) is simulated using \( y_t^* = \hat{\gamma} + \hat{\beta} x_t^* + \varepsilon_t^* \) with \( \hat{\gamma} = -13.6346 \) and \( \hat{\beta} = 0.0490 \) estimated using step interpolation.

Under data generating process (DGP) #1, temperature and CO2 concentration are not cointegrated, so \( \varepsilon_t^* \) is generated as a random walk with a variance of 0.0035, selected so that the variance of increments of \( y_t^* \) matches the sample variance of \( (y_{S_q} - y_{S_{q-1}})/\sqrt{S_q - S_{q-1}} \).

Under DGP #2 the two series are cointegrated, so \( \varepsilon_t^* = (1 - \rho_T^2)u_t + \sigma_v v_t \), where \( u_t = \rho_T u_{t-1} + w_t \) with \( \rho_T = (1 - 15/T) = 0.999981 \), \( \sigma_v^2 = 0.001761 \) and standard normal \( (w_t, v_t)' \). The parameter \( \sigma_v^2 \) is chosen match the sample variance as described above, while \( \rho_T \) is chosen to match the estimated autoregressive parameter of the highly persistent fitted residuals estimated using step interpolation. For a sample size as large as 798,380, this value of \( \rho_T \) is not as close, in a statistical sense, to one as it appears.

The simulated annual series \( (x_t^*) \) and \( (y_t^*) \) are then “masked” so that they are observed when CO2 concentrations and temperature are observed respectively – i.e., irregularly and non-contemporaneously. The masked series are step- and linear interpolated, standardized, and then the tests are performed with zero or one lagged first difference allowed in the testing procedure to control for serial correlation. If the annual data were observed, no lagged first difference would be necessary. However, Miller and Wang (2016) show that including a lag is useful to control for size distortion in the related context of testing with temporally aggregated data.

Table 1 shows the results. The distinction between \( \hat{\psi}_T(1) \) and \( \hat{\psi}_T(2) \) lies in the null: \( \hat{\psi}_T(2) \) has a null of no cointegration, considered in the previous section, while cointegration is possible under the alternative; whereas \( \hat{\psi}_T(1) \) has a null of cointegration, which is impossible under the alternative. For this reason the rejection rates should be considered to be the empirical size of the tests using \( \hat{\psi}_T(2) \), \( \hat{\rho}_T \), and \( \hat{\tau}_T \) under DGP #1 but of that using \( \hat{\psi}_T(1) \) under DGP #2. In contrast, DGP #2 shows the power of tests using \( \hat{\psi}_T(2) \), \( \hat{\rho}_T \), and \( \hat{\tau}_T \).

Compared to a nominal size of 0.05, the empirical size of the tests using linear interpolation and no lagged differences is quite distorted: 0.212, 0.995, 0.000, and 0.247. Under-rejection by \( \hat{\rho}_T \) is not necessarily a problem if the power is close to one, but it is also 0.000. In fact, the rejection rates of \( \hat{\psi}_T(2) \), \( \hat{\rho}_T \), and \( \hat{\tau}_T \) are lower under the alternative than under the null, suggesting that the tests are biased.

Including a lagged difference helps the trace test enormously, showing an acceptable size of 0.035 for \( \hat{\psi}_T(1) \) and acceptable size and power of 0.021 and 0.967 respectively for \( \hat{\psi}_T(2) \). The results show similar improvement for \( \hat{\tau}_T \), but the rejection rates for \( \hat{\rho}_T \) are now much too high rather than much too low.

Now, looking at step interpolation with no lagged differences, the sizes are all acceptable: 0.029, 0.047, 0.068, and 0.063 for \( \hat{\psi}_T(1) \), \( \hat{\psi}_T(2) \), \( \hat{\rho}_T \), and \( \hat{\tau}_T \) and the powers of the latter three are all optimal. The numbers are virtually unchanged if a lagged difference is included.
Table 1: Rejection Rates from 1,000 Monte Carlo Simulations. \( \hat{\psi}_T(j) \) denotes a trace test of the null of \( j \) stochastic trends \((2 - j \) cointegrating relationships) against the alternative of \(< j \) stochastic trends. \( p \) denotes number of lagged differences included. Rows labeled with \( \hat{z}^L_t \) and \( \hat{z}^S_t \) report results from linear interpolation and step interpolation respectively. CV denotes asymptotic critical values for a nominal size of 0.05 from Johansen and Juselius (1990) and Phillips and Ouliaris (1990).

|          | DGP #1 |               |               |               |               |               |               |
|----------|--------|---------------|---------------|---------------|---------------|---------------|
|          |        | \( p \) | \( \hat{\psi}_T(1) \) | \( \hat{\psi}_T(2) \) | \( \hat{\rho}_T \) | \( \hat{\tau}_T \) |
| \( \hat{z}_L^L \) |        | 0              | 0.658         | 0.995         | 0.000         | 0.247         |
|          |        | 1              | 0.004         | 0.021         | 0.949         | 0.018         |
| \( \hat{z}_L^S \) |        | 0              | 0.004         | 0.047         | 0.068         | 0.063         |
|          |        | 1              | 0.004         | 0.047         | 0.069         | 0.064         |
|          |        | \( \hat{z}_S^L \) |               |               |               |               |
|          |        | \( \hat{z}_S^S \) |               |               |               |               |
|          |        | CV             | 9.094         | 20.168        | 20.494        | 3.365         |

3.2 Empirical Results and Discussion

What can we glean from the simulation results? Echoing the advice of Ryan and Giles (1999) for unit root testing, Ghysels and Miller (2014) for cointegration testing, and Pretis and Hendry (2013), linear interpolation should be avoided in cointegration tests. Although step interpolation generates asymptotic size distortions for the tests, too, Monte Carlo simulations tailored to the paleoclimate data reveal these distortions to be small and that tests using step-interpolated data are as powerful as can be expected for both \( p = 0 \) and \( p = 1 \).

Given that they use linear interpolation, can we expect the results of Kaufmann and Juselius (2013) to be robust? Fortunately, the answer is in the affirmative, because they use the trace test with \( p = 1 \). (It is common practice to select the lag order using an information criterion, in which case it is usually chosen to be larger than zero.) From Table 1, this is precisely the case in which tests using linear interpolation have good size and power – albeit with the qualification that their data are aggregated to one-thousand-year increments rather than one-year increments as in the table. In the bivariate system presently considered, their result corresponds to a rejection of the null of no cointegration by \( \hat{\psi}_T(2) \), \( \hat{\rho}_T \), and \( \hat{\tau}_T \) and a failure to reject by \( \hat{\psi}_T(1) \).

Table 2 shows the results using the longer data set with fewer covariates. We include up to 50 lagged differences to check for robustness. Using linear interpolation with no lagged
Table 2: Test Statistics from the Bivariate Paleoclimate System. \( \hat{\psi}_T(j) \) denotes a trace test of the null of \( j \) stochastic trends (\( 2 - j \) cointegrating relationships) against the alternative of \( < j \) stochastic trends. \( p \) denotes number of lagged differences included. Rows labeled with \( \hat{z}_L^T \) and \( \hat{z}_S^T \) report results from linear interpolation and step interpolation respectively. CV denotes asymptotic critical values for a nominal size of 0.05 from Johansen and Juselius (1990) and Phillips and Ouliaris (1990).

differences, we reject both nulls with the trace statistics, suggesting that the series are cointegrated by any linear combination – i.e., that they do not have stochastic trends in the first place. This finding is not consistent with that of Kaufmann and Juselius (2013), but it is consistent with that of the univariate unit root tests of Davidson et al. (2016). In spite of the under-rejection of no cointegration by \( \hat{\rho}_T \) in the simulations, both residual-based tests support cointegration.

Increasing the number of lagged differences to one reverses the results of \( \hat{\psi}_T(1) \), so that all tests suggest a cointegrating relationship, consistent with the findings of Kaufmann and Juselius (2013). The sharp reversal of this conclusion and change in the values of the other test statistics clearly illustrate the change in the rejection rates noted in the simulations.

The results using step interpolation are robust to the number of lags included, as the simulations suggest they should be in the absence of serial correlation in the data. The results are conflicting, however, as all nulls are rejected. In other words, the trace tests suggest two cointegrating relationships among stationary series, while the residual-based tests suggest a single cointegrating relationship.

Because of the evident conflict between the conclusions using linear and step interpolation at a single lagged difference, lags up to 50 are considered. 50 is an unusually high

<table>
<thead>
<tr>
<th>( p )</th>
<th>( \hat{\psi}_T(1) )</th>
<th>( \hat{\psi}_T(2) )</th>
<th>( \hat{\rho}_T )</th>
<th>( \hat{\tau}_T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{z}_L^T )</td>
<td>0</td>
<td>35.6</td>
<td>16208.3</td>
<td>23.5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4.5</td>
<td>479.4</td>
<td>11789.1</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>5.8</td>
<td>290.2</td>
<td>4367.0</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>10</td>
<td>205.4</td>
<td>2303.7</td>
</tr>
<tr>
<td>( \hat{z}_S^T )</td>
<td>0</td>
<td>13.9</td>
<td>462.9</td>
<td>789.9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>13.9</td>
<td>463.5</td>
<td>791.3</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>10</td>
<td>459.5</td>
<td>789.7</td>
</tr>
<tr>
<td></td>
<td>( \vdots )</td>
<td>50</td>
<td>305.6</td>
<td>356.6</td>
</tr>
<tr>
<td>CV</td>
<td>9.09</td>
<td>20.17</td>
<td>20.49</td>
<td>3.37</td>
</tr>
</tbody>
</table>
number of lags in time series applications, but it is not unreasonable in this context given the large sample size and the distance between each observation. At that number of lags, there is agreement between the tests run on linearly and step-interpolated series.

However, the tests themselves do not agree, because all nulls are rejected. That disagreement could be caused by medium-run serial correlations not elsewhere picked up in the data and masked by either interpolation technique. Or, it could be caused by a modeling deficiency of the individual series. For example, although trending behavior appears in the data, Davidson et al. (2016) do not find statistical evidence of unit root-type trends.

More likely, the disagreement is caused by the presence of additive outliers in the underlying series, which is known to cause over-rejection in trace tests (Franses and Haldrup, 1994; Nielsen, 2004). An additive outlier is a long-lasting switch from one mean to another, and could certainly result from transitions between marine isotropic stages – i.e., between cool and warm periods in the paleoclimate data.

4 Conclusion

The analysis described in this paper fills important holes in two literatures. In the time series sampling literature, which considers issues of mixed observation frequencies, aggregation, spacing, etc., the analysis examines size distortion of standard cointegration tests when the underlying data are observed irregularly and non-contemporaneously and interpolation is used to overcome this complication.

The problems of irregularly and non-contemporaneously are salient in the paleoclimate literature. In the context of unit root and cointegration testing, they have been bridged by linear interpolation (Davidson et al., 2016, Kaufmann and Juselius, 2010a, 2010b, 2013, 2016). The present analysis reveals size distortion in standard cointegration tests using linear interpolation, echoing and expanding upon the findings of Ghysels and Miller (2014) of size distortion of such tests when linear interpolation is used in the related context of mixed-frequency time series.

Extreme size distortion may be mitigated by the usual way in which serial correlation is accommodated. Fortunately, this is exactly what Kaufmann and Juselius (2010a, 2010b, 2013, 2016), who conducted cointegration analyses on linearly interpolated paleoclimate data, do.

Step interpolation, proposed by Ryan and Giles (1999) for univariate unit root testing, works even better in this context. In contrast to linear interpolation, step interpolation employs only past information no future information. It may be the case in other contexts that the lack of future information is a constraint. However, the asymptotics of the cointegration tests better approximate null distributions using processes adapted only to past information.

The Monte Carlo simulations set up to emulate the paleoclimate data show that the size distortions are acceptable and power is optimal using step interpolation. Moreover, the conclusion of Kaufmann and Juselius (2013) that paleoclimate temperatures and CO$_2$ concentrations are cointegrated is largely supported, albeit with the qualification that size distortion may remain due to residual causes.
Step interpolation or another method that does not rely on future information should be used when a researcher with the goal of testing for unit roots, cointegration, or any inference that relies on unit root asymptotics is confronted with irregularly spaced and non-contemporaneous time series. This method is not free from size distortion, which seems unavoidable in this context, but the size distortion appears to be acceptable and certainly more so than with linear interpolation or other methods that use future information.

The method for collecting the data from ice cores results not only in irregularity and non-contemporaneity but also in measurement uncertainty. An extension of the present analysis might consider such uncertainty. Lorius et al. (1985) give the precision of their $\delta^{18}O$ measurement as $\pm0.15$ parts per thousand relative to that of ocean water, and $\pm0.1$ is given by Bradley (1999) as a typical range. In light of a fluctuation of about 8 parts per thousand (Figure 1 of Lorius et al., 1985), these measurements are relatively precise.

Estimation of dates is less precise, and, naturally, precision deteriorates with drilling depth. Lorius et al. (1985) give the precision of their dates as 10-15 kyr at the oldest part of their 150 kyr sample, while the more recent chronology of Parrenin et al. (2007) is estimated to be accurate to within 4 kyr at 150 kyr before present and no more than 6 kyr throughout the 798.38 kyr used in the present analysis.

A fundamental problem is that uncertainty intervals tell us nothing about the temporal correlation of the measurement uncertainties over time. Fortunately, measurement error in cointegration analysis is a smaller problem than in correlation analysis, as long as the error is only weakly dependent and $p$ is sufficiently large. The effects of uncertainty of measurements and dates provide interesting avenues for future research.

**Data Availability Statement**

The data that support the findings of this study were derived from the following resources available in the public domain:

- Jouzel et al. (2007): doi.pangaea.de/10.1594/PANGAEA.683655
- Lithi et al. (2008): www.nature.com/articles/nature06949
- NOAA: www.esrl.noaa.gov/gmd/ccgg/trends/

**References**


Doan, T.K., J. Haslett, and A.C. Parnell, 2015, Joint inference of misaligned irregular time series with application to Greenland ice core data, Advances in Statistical Climatology, Meteorology and Oceanography 1, 15-27.


Technical Appendix

Lemma A1. For the series $x_{T_p}$ and $y_{S_q}$ with $|S_q - T_q| < \infty$ and for $[Tr]$ defined to be the greatest integer not exceeding $T_p$, the functional central limit theorems

$$T^{-1/2}x_{[Tr]} \to_d B_x(r)$$
$$T^{-1/2}y_{[Tr]} \to_d B_y(r)$$

hold as $T \to \infty$.

Proof of Lemma A1. The normalized series $x_{T_p}$ may be written as a sum of its observed increments $\sum_{t=1}^{T_p}(x_{T_t} - x_{T_{t-1}})$ or of its unobserved increments $\sum_{t=1}^{T_p}(x_t^* - x_{t-1}^*)$. Using the usual Skorokhod embedding on the unit interval, the stochastic process $T^{-1/2}\sum_{t=1}^{T_p}(x_t^* - x_{t-1}^*)$ maps to $T^{-1/2}\sum_{t=1}^{\lceil Tr \rceil}(x_t^* - x_{t-1}^*)$, which gives the limit $B_x(r)$ by the functional central limit theorem assumed in Section 2 for the latent process $(x_t^*)$.

The proof for $y_{S_q}$ is similar, but its increments may be written as

$$\sum_{t=1}^{T_p}(y_t^* - y_{t-1}^*) + \sum_{t=S_q+1}^{S_q}(y_t^* - y_{t-1}^*)$$

if $S_q > T_q$ or

$$\sum_{t=1}^{T_p}(y_t^* - y_{t-1}^*) - \sum_{t=S_q+1}^{T_p}(y_t^* - y_{t-1}^*)$$

if $S_q < T_q$. As long as the difference is finite, the second term of each of these expressions collapses to zero when normalized by $T^{-1/2}$, and the limit $B_y(r)$ follows from the limit of the first term.

Lemma A2. Sample moments $R_{11}$, $R_{00}$, and $R_{10}$ using step-interpolated series have limits given by

[a] $T^{-1}R_{11} \to_d \int BB'$,

[b] $R_{00} \to_p F$, and

[c] $R_{10} \to_d \int BdB' + G$

as $T \to \infty$.

Proof of Lemma A2. For the proof of each part, consider four cases: (i) $t = S_q, T_p$, (ii) $t \in (S_q, S_{q+1})$, $t = T_p$, (iii) $t \in (T_p, T_{p+1})$, $t = S_q$, and (iv) $t \in (S_q, S_{q+1}) \cap (T_p, T_{p+1})$. A subtlety of this notation is that $S_q$ and $T_p$ are both functions of $t$ and of each other. In other words, only intervals “near” any given $t$ are considered.

For part [a], $\sum_t \hat{y}_{t-1}^S \hat{x}_{t-1}^S = \sum_t y_{S_{q+1} - T_{p+1}} x_{T_{p+1}}$ in case (i). In the remaining cases, $\sum_t \hat{y}_{t-1}^S \hat{x}_{t-1}^S = \sum_t y_{S_q - T_{p+1}} x_{T_{p+1}} + O_p(T)$, where the remainder term is $\sum_t (y_{S_q} - y_{S_{q+1}}) x_{T_{p+1}}$ in case (ii), $\sum_t y_{S_q - T_{p+1}} x_{T_{p+1}}$ in case (iii), and augmented by $\sum_t (y_{S_q} - y_{S_{q+1}}) (x_{T_{p+1}} - x_{T_{p+1}})$ in case
(iv). That these terms are $O_p(T)$ follows from the functional central limit theorem in Lemma A1 along with the usual covariance asymptotics in cases (ii) and (iii) and a law of large numbers in case (iv). To complete the proof of part (a), it only remains to invoke the continuous mapping theorem and the limit in Lemma A1 to get the limit of $T^{-2}\sum_t y_{S_{q-1}} x_{T_{p-1}}$. The limits of $T^{-2}\sum_t y_{S_{q-1}}^2$ and $T^{-2}\sum_t x_{T_{p-1}}^2$ follow similarly, and the limit of the matrix $T^{-1}R_{11}$ follows due to joint convergence.

The summations $\sum_t \triangle \hat{y}_t^S \triangle \hat{x}_t^S$ and $\sum_t \triangle \hat{x}_t^S \triangle \hat{y}_t^S$ in $R_{00}$ in part [b] equal zero in cases (ii)-(iv) and equal $\sum_p (y_{T_p} - y_{S_{q-1}})(x_{T_p} - x_{T_{p-1}})$ when $S_q = T_p$ in case (i). Because $S_q = T_p$ only finitely often, these summations are $O_p(T)$. The summation $\sum_t (\triangle \hat{x}_t^S)^2$ equals $\sum_p (x_{T_p} - x_{T_{p-1}})^2 = \sum_p (\sum_{t=T_{p-1}+1}^{T_p} \triangle x_t^S)^2$ which is $\sum_t (\triangle x_t^S)^2 + o_p(T)$ due to the uncorrelatedness of the increments of $(\triangle x_t^S)$, and the limit holds from a law of large numbers. The same argument holds for $\sum_t (\triangle \hat{y}_t^S)^2$ and the stated result holds because the convergences are joint.

For part [c], first consider $\sum_t \hat{y}_t^S \triangle \hat{x}_t^S$. In cases (iii) and (iv), $t > T_p$, so that $\triangle \hat{x}_t^S = x_{T_p} - x_{T_{p-1}} = 0$ and the whole summation is therefore zero. Write

$$\sum_t \hat{y}_t^S \triangle \hat{x}_t^S = \sum_p y_{T_{p-1}}^* (x_{T_p} - x_{T_{p-1}}) + \sum_p (y_{S_{q-1}} - y_{T_{p-1}}^*)(x_{T_p} - x_{T_{p-1}}) \quad (A.1)$$

and recall that $S_q$ is a function of $T_p$, so that the second term cannot be simplified. Looking at the first term of (A.1),

$$T^{-1} \sum_p y_{T_{p-1}}^* (x_{T_p} - x_{T_{p-1}}) = \sum_p (T^{-1/2} y_{T_{p-1}}^*) (T^{-1/2} x_{T_p} - T^{-1/2} x_{T_{p-1}}) \rightarrow_d \int B_y(r) dB_x(r)$$

follows using Lemma A1. The second term of (A.1) has a limit given by

$$T^{-1} \sum_p (y_{S_{q-1}} - y_{T_{p-1}}^*)(x_{T_p} - x_{T_{p-1}}) \rightarrow_p \pi_{yx}$$

by a law of large numbers, which gives the stated result for $\sum_t \hat{y}_t^S \triangle \hat{x}_t^S$. The same arguments hold for $T^{-1} \sum_t \hat{x}_t^S \triangle \hat{y}_t^S$, and the convergences of $T^{-1} \sum_t \hat{y}_t^S \triangle \hat{y}_t^S$ and $T^{-1} \sum_t \hat{x}_t^S \triangle \hat{x}_t^S$ are even more straightforward, because only case (i) and (iv) are possible and the second term of (A.1) is zero for those moments. The proof of part [c] is completed by noting that the convergences are joint.

**Proof of the Proposition.** The mechanism of the proof follows Ghysels and Miller (2014, 2015) closely. Define $\Sigma = L' L$ and generic matrices $\Xi_T$ and $\Xi$ such that $L' \Xi_T L$ denotes one of the three moments $T^{-1} R_{11}$, $R_{00}$, and $R_{10}$ in Lemma A2, and $L' \Xi L$ denotes their respective limits. In particular, the matrix $\Xi$ is $\int WW'$ for $T^{-1} R_{11}$, $\Xi = F^+$ for $R_{00}$, and $\int WdW' + G^+$ for $R_{10}$ from Lemma A2.

Ghysels and Miller (2014) show that $\alpha' L' \Xi_T L \alpha \rightarrow_d l_{11} \kappa' \Xi \kappa$, where $l_{11}$ is the element in the first row and column of $L$ and $\kappa = (1, \kappa_2)'$ with $\kappa_2 = -(\int W_2^2)^{-1} \int W_2 W_1$ as defined in the text. All that is needed to complete the proof in this case with irregularly and non-contemporaneously observed series is to substitute the respective matrices $\Xi_T$ into (2) and (4) and invoke the continuous mapping theorem with these limits. 

□