

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

# **Global Temperatures and Greenhouse Gases: A Common Features Approach**

Li Chen, Jiti Gao and Farshid Vahid

October 2019

Working Paper 23/19

# Global Temperatures and Greenhouse Gases: A Common Features Approach

Li Chen<sup>1,3</sup>, Jiti Gao<sup>2</sup> and Farshid Vahid<sup>2</sup>

<sup>1</sup> Wang Yanan Institute for Studies in Economics, Xiamen University

<sup>2</sup> Department of Econometrics and Business Statistics, Monash University

<sup>3</sup> Department of Finance, School of Economics, Xiamen University

September 2, 2019

## Abstract

We propose a common features approach and establish that global temperatures and greenhouse gases share a common trend without conditioning on the exact nature of their trends. We also find deterministic cycles in global temperature series with a period of 72 years whose amplitude is non-negligible compared to the warming effect caused by anthropogenic greenhouse gas emissions. We also explore the direction of causation in the co-trending relationship by establishing an error correction mechanism, which shows that greenhouse gas emissions do not respond to the equilibrium errors whereas the global temperature responds to equilibrium errors so as to maintain the co-trending relationship. Finally, we forecast future temperatures conditional on representative concentration pathways (RCP) of greenhouse gases considered by the intergovernmental panel on climate change (IPCC). Our forecasts show that the cyclical component could play a significant role in the next 30 years in such a way that would make the politics of effective emission control policy making more challenging.

**Keywords:** Climate Change; Cycle; Endogeneity; Trending Behaviour

**JEL Classification:** C22; C53; Q54

# 1 Introduction

No one understands trends, but every one sees them in the data.

*Laws and Limits of Econometrics*, Peter C. B. Phillips (2003)

There is no doubt that global temperatures have been on a warming trend since the mid-twentieth century. The intergovernmental panel on climate change report (IPCC 2014) concludes that anthropogenic greenhouse gas emissions are extremely likely to have been the dominant cause of such a warming phenomenon, which is a key feature of climate change we are faced with today. Scientists and international associations have been appealing for the control of global warming by reducing the emissions of greenhouse gases, which is also the subject of much research in the area of climate change. In this paper, we are interested in exploring the long-run relationship between global temperatures and greenhouse gases. Specifically, we ask the question do they share a common trend?

This is a simple but fundamental question in the research of climate change. While climate scientists primarily rely on climate models to study global warming as in Cubasch & Meehl (2001) and Mitchell & Karoly (2001), econometricians use time series models for this purpose, but they have conditioned their analysis on the nature of trend in these variables. For example, Kaufmann & Stern (2002) and Kaufmann et al. (2006) argue that global temperature anomalies and radiative forcing of greenhouse gases contain stochastic trends (unit roots) that form a *cointegration relationship*. Mills (2009) confirms the robustness of such a cointegration relationship by examining the sensitivity of temperature to radiative forcing of greenhouse gases. Kaufmann et al. (2010) conduct an in-sample forecasting experiment, which justifies the rationality of viewing temperature series as an integrated process. Kaufmann et al. (2013) argue that when physical mechanisms are taken into account, statistical results are consistent with the assumption of stochastic trends. Recently, Chang et al. (2016) applied a new unit root test to suggest the existence of unit root in the temperature data.

However, there are other econometricians who believe temperatures are trend-stationary time series that contain non-linear deterministic trends. For example, Gao & Hawthorne (2006) and Gay et al. (2009) argue that temperature series are better characterized by a trend-stationary process with smooth nonlinear time trend or linear trends with breaks rather than unit root time series. Estrada et al. (2010) find that

econometric methods are sometimes incorrectly used in the papers that view temperature as a unit root process with stochastic trends. As a result, the cointegration conclusion is not reliable as it is based on the assumption that the historical climate data are integrated time series. Estrada et al. (2013) use a nonlinear co-trending test proposed by Bierens (1997) to show that temperature and radiative forcing in fact share a common nonlinear deterministic trend.

After careful consideration of the above analyses with different conditions and results, we conclude that in practice, it is extremely difficult to establish the nature of trend in a time series with a finite sample. In other words, it is difficult to differentiate an integrated time series (for example, a random walk process) from a nonlinear trend-stationary process (for example, a stationary process around a nonlinear deterministic trend). This is in accordance with the messages in Phillips (2001, 2003, 2005, 2010), in which the author explained *the mysteries of trend* and *the challenges of trending time series on empirical analysis*. This problematic issue has also been studied in the econometrics literature (see Nelson & Plosser (1982), Kwiatkowski et al. (1992), DeJong et al. (1992), Bierens (1997), etc), in particular with temperature data in Breusch & Vahid (2011).

We acknowledge the difficulty in determining the exact nature of trend in the non-stationary trending time series, and, in particular, the debate on the real characteristic of trend in the global temperature series. Therefore, the main contribution of this paper is to study the co-trending relationship without conditioning on a particular type of trend for the empirical time series. Moreover, we consider the possibility of endogeneity in the co-trending regression equation and propose a method to address this issue as well.

The problem of endogeneity can arise when the regressand and regressors in a regression model are simultaneously determined. As pointed out in Kaufmann et al. (2006), the warming effect of greenhouse gas emissions leads to changes in global temperatures, which may also affect the concentration of greenhouse gases in the atmosphere. Therefore, endogeneity problem may appear so that the estimation of the co-trending relationship could be affected when we regress the time series of temperature on that of greenhouse gases. In particular, the ordinary least squares (OLS) estimator of the co-trending relationship is likely to be biased and possibly inconsistent. Hence conclusions of statistical inferences are likely to be misleading (see Phillips

& Hansen (1990) and Chen (2017) for a detailed discussion).

In this paper, we approach both issues of trend nature ambiguity and the problem of endogeneity from a common features perspective (Engle & Kozicki 1993). A *feature* is a dominant statistical property of a variable that cannot be eliminated by linearly transforming that variable<sup>1</sup>. For example, a trend, whether deterministic or stochastic, is a feature. Also, nonlinearity can be regarded as a feature as in Anderson & Vahid (1998). Two series have a *common feature* if each one of them has a certain feature, but a linear combination of them does not have that feature. For example, cointegrated series have a common stochastic trend, and variables that move concurrently with the business cycle have a common cycle (see Vahid & Engle 1993, for a formal definition). When time series share a common trend as a common feature, no matter whether it is stochastic or deterministic, we call the combination that does not have a trend a *co-trending relationship*. It indicates a stable long-run relationship between the time series.

The rationale of the econometric method proposed in this paper is based on Phillips (1998), in which the author presents an interesting explanation of “spurious regressions”. It is worth noting that Phillips (1998) provides a theoretical linkage between stochastic trends and deterministic trends. Specifically, a unit root process could be approximated by a set of orthogonal deterministic basis functions that

$$y_t \sim \sum_{k=1}^K b_k \phi_k \left( \frac{t}{n} \right), \quad (1)$$

where  $y_t$  is an I(1) process,  $\phi_k(u)$  is a set of orthogonal basis functions defined on  $[0,1]$  for  $k = 1, 2, \dots, K$ . In this regression, the coefficient  $(b_k/\sqrt{n})$  becomes random. Meanwhile, a deterministic trend could also be fitted by orthogonalized Brownian motions that

$$f(r) \sim \sum_{i=1}^N d_i W_i(r), \quad (2)$$

for  $r \in [0, 1]$ , where  $W_i(r)$  is a set of independent standard Brownian motions for

---

<sup>1</sup>A “feature” is a dominant statistical property that satisfies:

- if  $x_t$  has (does not have) the feature, then  $cx_t$  where  $c \neq 0$  will also have (not have) it,
- if  $x_t$  and  $y_t$  do not have the feature,  $x_t + y_t$  will not have it,
- if  $x_t$  has the feature and  $y_t$  does not have it, then  $x_t + y_t$  will have the feature.

$i = 1, 2, \dots, N$ . Phillips (1998) has two major implications for this paper. First, as deterministic and stochastic trends can approximate each other quite well, it is difficult to distinguish between a stochastic trend and a nonlinear deterministic trend. Second, the trending characteristics in the nonstationary time series could be projected onto and extracted by a set of orthogonal basis functions of time, regardless of whether they are stochastic or deterministic. At the same time, they are uncorrelated with the regression errors as they are purely deterministic functions. Therefore, for the trending regression model we analyse in this paper, the orthogonal basis functions could be taken as instrumental variables to estimate the co-trending parameter.

By using the deterministic basis functions of time as instrumental variables, we are able to estimate and test for the co-trending relationship without the need to know the nature of trends in the trending time series. We call this a common features approach as in this method, trend is taken as a feature and extracted by a set of basis functions of time. In other words, given that trend is the dominant feature of both global temperatures and the radiative forcing of well-mixed greenhouse gases, our objective is to see if there is a linear combination of these two time series that does not have a trend. Therefore, we will first estimate the co-trending relationship without conditioning on the nature of their trends, and then examine the co-trending relationship by a statistical test.

The remaining part of this paper is organized as follows. Section 2 introduces the empirical data and motivates this paper from the practical point of view. Then we establish the econometric model for detecting common trends, and present the asymptotic results of the proposed estimators as well as technical assumptions needed to deliver them. In Section 3, we show the empirical findings of the co-trending relationship between global temperatures and greenhouse gases with further discussions on the regression residuals and forecasts. Numerical simulation results are reported in Section 4. Section 5 concludes the paper. Justifications of the assumptions and mathematical proofs of the main theorems are provided in the Appendix.

## 2 Data and Methodology

### 2.1 Global temperatures and greenhouse gases

We present in Figure 1 the annual data of global temperature anomalies (GTA) from 1850 to 2018. It is one of the commonly used measures of global temperatures compiled by the Climatic Research Institute<sup>2</sup>. As shown in the graph, an upward warming trend is evident in GTA, in particular after the 1970s. While more specifically, we see fast global warming periods from 1910 to 1940, and then from 1970 to 2018. We also see slightly cooling periods, for example, from 1880 to 1910, and 1940 to 1970. Breusch & Vahid (2011) provide a discussion of whether the trend in this data series is deterministic or stochastic, and the difficulty of determining that from the observed data. In particular, they explain that it will be even harder to determine if this series has a stochastic trend against the alternative that it has a broken (piecewise linear) deterministic trend.

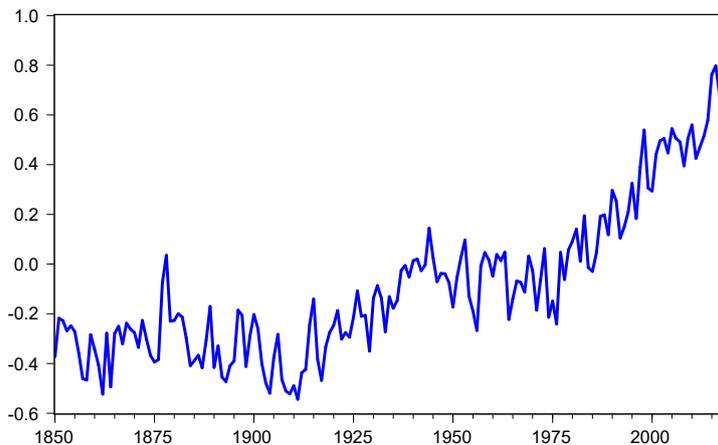


Figure 1: Global temperature anomalies (GTA), 1850-2018

Figure 2 shows the logarithms of  $CO_2$  equivalent of total greenhouse gases considered in the Kyoto protocol, and it is taken as a measure for the total emissions of greenhouse gases (GHG). Similar to GTA, GHG also exhibits an upward trend. A striking feature is that the slope of the curve is larger after 1950 than that before 1950. This indicates a growing amount of emissions of greenhouse gases with the rapid industrialization of many countries. Meanwhile, the graph of greenhouse gases in Figure

---

<sup>2</sup>They are combined land and ocean temperatures, expressed as deviations from the 1960-1990 average. This series is downloadable from <https://crudata.uea.ac.uk/cru/data/temperature/>. The version used in this paper was downloaded on 12 June 2019.

2 is much smoother than the graph of global temperatures and one may expect the series to be dominated by a trend with a greater degree of integration, for exaple, an I(2) process.

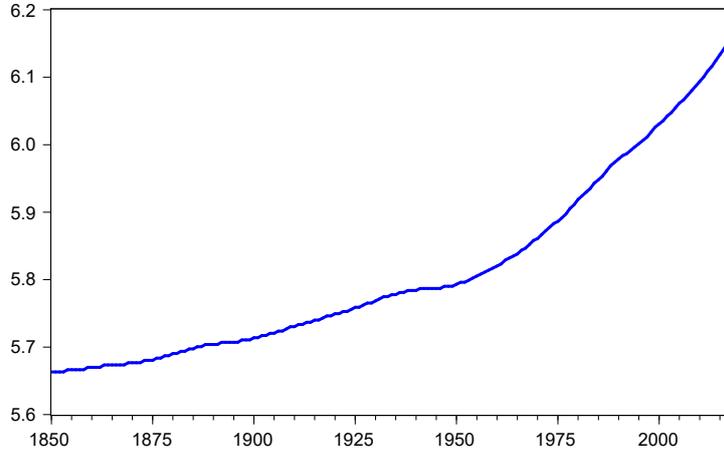


Figure 2: Logarithm of  $CO_2$  equivalent greenhouse gas emissions (GHG), 1850-2018

## 2.2 Uncertainty of trend characteristics in the data

As we can see from the graphs that trends are dominant components in both GTA and GHG. In this subsection, we present some of the empirical facts in pursuing the real nature of trends in both time series. These findings motivates us to propose the methodologies in this paper.

When GTA and GHG are taken as integrated time series that exhibit stochastic trends, a common practice is to use statistical tools to test for the orders of integration in the time series. In this paper, we apply the *Augmented Dickey-Fuller (ADF) unit root test* and the *KPSS unit root test* at the same time. As the *statement of unit root* is taken as the null hypothesis in the ADF test and as the alternative hypothesis in the KPSS test, we can control for the probability of making a false inferential conclusion<sup>3</sup>.

Table 1 shows the  $p$ -values of the unit root tests for GTA and GHG. We can conclude that GTA is an I(1) process, while GHG is an I(2) process at the 5% significance

---

<sup>3</sup>ADF tests are found to have low power against certain alternative hypothesis in DeJong et al. (1992). Therefore, it is very likely that we fail to reject the null hypothesis even when the null is false. This will lead to a high probability of making the Type-II error. Fortunately, the risk of making such a wrong conclusion in the ADF test could be controlled by applying the KPSS test, in which the unit root hypothesis is in the alternative hypothesis. Thus the probability of incorrectly accepting the unit root hypothesis (making the type I error in the KPSS test) is controlled by the significance level we choose in the KPSS test.

level. Therefore, regressing GTA on GHG leads to an unbalanced regression as the orders of integration are different on both sides of the regression model.

Table 1:  $p$ -values of unit root tests

	ADF test	KPSS test
$GTA$	0.59	< 0.01
$\Delta GTA$	< 0.01	> 0.10
$GHG$	0.99	< 0.01
$\Delta GHG$	0.55	< 0.01
$\Delta^2 GHG$	< 0.01	> 0.10

Meanwhile, we note that in the ADF and KPSS tests, the deterministic trend component is always specified as a linear function of time. Bierens (1997) finds that the linear deterministic trends could be relaxed to nonlinear ones, but the consequence is that the existence of unit root becomes uncertain. In other words, it is also reasonable to believe that both trending time series of GTA and GHG are stationary around a nonlinear deterministic trend. In this subsection, we test for stationarity on the residuals of GTA and GHG by removing a fitted linear, quadratic, and cubic time trends respectively. Table 2 presents our findings with respect to different specifications of deterministic trends.

Table 2: Specifications of deterministic trends and unit root test results

	deterministic trend forms	unit root test on residuals *
$GTA$	$\alpha_0 + \alpha_1 t$	I(1)
	$\alpha_0 + \alpha_1 t + \alpha_2 t^2$	I(0)
	$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$	I(0)
$GHG$	$\alpha_0 + \alpha_1 t$	I(2)
	$\alpha_0 + \alpha_1 t + \alpha_2 t^2$	I(2)
	$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3$	I(0)

\* At the 5% significance level.

Table 2 shows that GTA could be viewed as an I(1) process combined with a linear time trend, and GHG could be viewed as an I(2) process with a linear or even quadratic time trend. However, when we specify more nonlinearity on the deterministic trend component, the characteristic of integration “disappears” for both time series of GTA and GHG. In other words, they could also be regarded as a nonlinear deterministic

trend with stationary errors. This is similar to the debate on the nature of trends in macroeconomic variables that ensued from the publication of Nelson & Plosser (1982).

From our point of view, we see these contradictory results as an identification problem of attributing nonstationarity to components of either deterministic trends or stochastic trends. Consequently, trends in GTA and GHG can hardly be correctly specified as the exact nature of them are unknown (possibly unknowable) (see Phillips (2010)). The research question of this paper is that with all this uncertainty in the nature of the trend in these series, how can we establish if they have a long-run relationship, and if so, how to estimate this relationship and determine a measure of confidence in the estimates.

### 2.3 Model and estimation method

Suppose that  $y_t$  and  $x_t$  have the same feature  $F$ . Under the null that  $F$  is a common feature between  $y_t$  and  $x_t$ , there is one and only one  $\beta$  such that  $e_t = y_t - x_t\beta$  does not have feature  $F$ . Specifically,  $F$  is taken as ‘trend’ in this paper and we consider trending time series of the form

$$y_t = f_t + u_t, \tag{3}$$

$$x_t = g_t + v_t, \tag{4}$$

for  $t = 1, 2, \dots, n$ , where  $f_t$  and  $g_t$  are trending components,  $u_t$  and  $v_t$  are ergodic stationary disturbances. As it is difficult to determine the exact nature of trends in nonstationary trending time series  $x_t$  and  $y_t$ ,  $f_t$  and  $g_t$  are not specified to certain forms, nor are they estimable from observed values of  $y_t$  and  $x_t$ . Therefore, it can be linear or finite order polynomials of time, it can be an integrated stochastic process of order 1 or larger, it can also be a piecewise linear trend, or even a combination of these forms. It is important to be clear that the possibility that these series may have a common stochastic trend but different deterministic trends is of no interest to us. If the trends in  $y_t$  and  $x_t$  are common, then there exists one and only one  $\beta_0$  such that  $f_t = \beta_0'g_t$  and there is no trend in  $y_t - \beta_0'x_t$ . Our objective in this paper is to estimate  $\beta_0$  and establish its confidence intervals.

To this end, we then consider the pseudo-structural model

$$\begin{aligned} y_t &= \alpha_0 + \beta_0' x_t + e_t \\ x_t &= g_t + v_t \end{aligned} \tag{5}$$

where  $g_t$  is a trend process. We call this a pseudo-structural system because  $e_t$  and  $v_t$  are correlated<sup>4</sup>, and hence in the first equation the right-hand side variable is endogenous. We first focus on consistent estimation of  $\beta_0$  in this pseudo-structural system.

Our first argument on the estimation method is that the powers of  $(t/n)$  are legitimate instruments that can produce an instrumental variable (IV) estimator of  $\beta_0$  in equation (5), *regardless of whether the trend ( $g_t$ ) is deterministic or stochastic*. That powers of  $(t/n)$  can be used as instruments in estimating cointegrating relationships is evident in Park & Phillips (1988), and the performance of this IV estimator in the case of I(1) variables has been studied in Phillips & Hansen (1990). Here we propose that this applies more generally to all forms of common trends. The rationale is based on Phillips (1998) as discussed in the introduction part.

The instruments could be selected as the first  $(p + 1)$  elements<sup>5</sup> from a set of orthogonal basis functions. For example, the *Legendre polynomials* defined on  $[0,1]$  of the form

$$\phi_j(\tau) = \sqrt{2j+1} \sum_{s=0}^j (-1)^{j+s} C_j^s C_{j+s}^s \tau^s, \quad j = 0, 1, 2, \dots \tag{6}$$

The instrumental variables satisfy  $\int_0^1 \phi_i(\tau) \phi_j(\tau) d\tau = \delta_{ij}$ , where  $\delta_{ij} = 1$  for  $i = j$  and  $\delta_{ij} = 0$  for  $i \neq j$ ,  $i, j = 0, 1, \dots, p$ . According to Phillips (1998), trending components in  $x_t$  could be viewed as ‘*spuriously correlated*’ with the basis functions  $\phi_j(\tau_t)$ , where  $\tau_t = t/n$ ,  $t = 1, 2, \dots, n$ , regardless of whether the trend in  $x_t$  is deterministic or stochastic. Meanwhile, under the null hypothesis of common trend,  $e_t = y_t - \alpha_0 - x_t' \beta_0$  is merely a zero-mean stationary process without any trend. Therefore,  $e_t$  is “*uncorrelated*” with the basis functions so that  $(\phi_0(\tau), \phi_1(\tau), \dots, \phi_p(\tau))$  could be used as instruments to

---

<sup>4</sup>In fact, since  $f_t = \beta_0' g_t$ , we have  $y_t = \beta_0' g_t + u_t = \beta_0' x_t - \beta_0' v_t + u_t$ , where the regression error  $e_t = -\beta_0' v_t + u_t$ . Therefore,  $cov(e_t, v_t) = -\beta_0' \Sigma_v + cov(u_t, v_t)$  which is non-zero unless  $cov(u_t, v_t) = \beta_0' \Sigma_v$ , where  $\Sigma_v$  is the variance-covariance matrix of  $v_t$ .

<sup>5</sup>We include a constant term in the instruments by letting  $\phi_0(\tau) \equiv 1$ .

estimate  $(\alpha_0, \beta'_0)$  based on the “*orthogonality conditions*” as

$$\frac{1}{n} \sum_{t=1}^n \phi_j(\tau_t) e_t \xrightarrow{p} 0, \quad \text{as } n \rightarrow \infty, \quad (7)$$

for  $j = 0, 1, 2, \dots, p$ , where  $e_t = y_t - \alpha_0 - x'_t \beta_0$ .

In matrix notation, let  $y = (y_1, y_2, \dots, y_n)'$ ,  $e = (e_1, e_2, \dots, e_n)'$ ,  $\gamma_0 = (\alpha_0, \beta'_0)'$  and  $X = (1_n, \tilde{X})$ ,  $\Phi = (1_n, \tilde{\Phi})$ , where  $1_n$  is a  $n$ -dimensional vector of 1s,

$$\tilde{X} = X_{n \times k} = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{k1} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}, \quad (8)$$

for  $t = 1, 2, \dots, n$  and

$$\tilde{\Phi} = \tilde{\Phi}_{n \times p} = \begin{pmatrix} \phi_1(\frac{1}{n}) & \phi_2(\frac{1}{n}) & \dots & \phi_p(\frac{1}{n}) \\ \dots & \dots & \dots & \dots \\ \phi_1(\frac{n}{n}) & \phi_2(\frac{n}{n}) & \dots & \phi_p(\frac{n}{n}) \end{pmatrix}. \quad (9)$$

Therefore, the regression equation could be written as

$$y = X\gamma + e. \quad (10)$$

When orthogonal basis functions as columns in  $\Phi$  are taken as instrumental variables, we define a *two-stage least squares estimator* as

$$\hat{\gamma}_{2sls} = (\hat{\alpha}_{2sls}, \hat{\beta}'_{2sls})' = (X'\Phi(\Phi'\Phi)^{-1}\Phi'X)^{-1}X'\Phi(\Phi'\Phi)^{-1}\Phi'y. \quad (11)$$

In the following subsection, we will show the asymptotic properties of the 2SLS estimator. We will also show that the Sargan’s test, originally proposed to test for over-identifying restrictions, could be used to examine whether or not  $y_t$  and  $x_t$  share a common trend.

## 2.4 Assumptions and asymptotic properties

We first give some assumptions before establishing the asymptotic results.

**Assumption 1:** Let  $\{y_t, x_t\}_{t=1}^n$  be nonstationary trending time series. There exists a  $(k+1)$ -dimensional vector  $\gamma_0 = (\alpha_0, \beta_0)'$  such that  $e_t = y_t - \alpha_0 - x_t' \beta_0$  is strictly stationary with  $E[e_1] = 0$  and  $\sigma_e^2 = E[e_1^2] < \infty$ .

**Assumption 2:** Let  $\Phi(\tau) = (\phi_0(\tau), \phi_1(\tau), \phi_2(\tau), \dots, \phi_p(\tau))'$  be a  $(p+1) \times n$  matrix for  $\tau \in [0, 1]$  and  $p \geq k$  such that  $\phi_0(\tau) \equiv 1$  and

(i)  $\int_0^1 \phi_i(\tau) \phi_j(\tau)' d\tau = \delta_{ij}$ , where  $\delta_{ij} = 1$  for  $i = j$  and  $\delta_{ij} = 0$  for  $i \neq j$ . In addition  $\int_0^1 \phi_i(\tau)^4 d\tau < \infty$  for  $i = 0, 1, 2, \dots, p$ ;

(ii)  $E[U'\varepsilon] = 0$ , where  $U = W^{1/2}\Phi'X$  and  $\varepsilon = W^{1/2}\Phi'e$ , in which  $W = (\Phi'\Phi)^{-1}$ ;

(iii) Suppose that there exists a  $(k+1)$ -dimensional positive definite deterministic matrix  $D_n$  with  $\|D_n\|_2^2 \rightarrow \infty$  as  $n \rightarrow \infty$ , such that  $(D_n^{-1}U'UD_n^{-1}, D_n^{-1}U'\varepsilon) \rightarrow_{\mathcal{D}} (\xi, \eta)$ , where  $\xi$  is a matrix of positive random variables and  $\eta$  is a  $k+1$ -dimensional mixture Normal random variable  $N(0, \xi \lambda_e^2)$  with  $E\|\xi\|_2 < \infty$ , in which  $\lambda_e^2 = E[e_1^2] + 2\sum_{j=1}^{\infty} E[e_1 e_{1+j}]$  is the long-run variance and  $\|A\|_2 = \sqrt{\sum_{i=1}^d \sum_{j=1}^d a_{ij}^2}$  for matrix  $A = \{a_{ij} : 1 \leq i, j \leq d\}$ .

Before the justifications of Assumptions 1 and 2 are given in Appendix A.1, we establish the main results of this paper.

**Theorem 1:** Let **Assumptions 1** and **2** hold. As  $n \rightarrow \infty$ , the two-stage least squares estimator satisfies:

$$(U'U)^{1/2} (\hat{\gamma}_{2sls} - \gamma) \xrightarrow{D} N(0, \lambda_e^2 I_{(k+1)}), \quad (12)$$

where  $U = W^{1/2}\Phi'X$  and  $\lambda_e^2$  is the long-run variance of  $e_t$ .

In practice,  $\lambda_e^2$  can be consistently estimated by  $\hat{\lambda}_e^2$  that  $\hat{\lambda}_e^2 \rightarrow_p \lambda_e^2$  as  $n \rightarrow \infty$ . For example, in Newey & West (1987, 1994) and Andrews (1991), they propose

$$\hat{\lambda}_e^2 = \sum_{j=-p_n}^{p_n} k(j/p_n) \hat{\gamma}(j). \quad (13)$$

**Remarks:** In the long-run variance estimator, the autocovariances of  $\gamma(j)$  could be estimated by  $\hat{\gamma}(j) = n^{-1} \sum_{t=1+j}^n \hat{e}_t \hat{e}_{t-j}$  in which  $\hat{e}_t = y_t - \hat{\alpha}_{2sls} - x_t' \hat{\beta}_{2sls}$ ; the truncation parameter  $p_n$  satisfies  $p_n \rightarrow \infty$ ,  $p_n/n \rightarrow 0$  as  $n \rightarrow \infty$ ;  $k(u)$  is a kernel function defined on  $[-1, 1]$ , for example, a commonly used kernel function is the Bartlett kernel  $k(u) = (1 - |u|)\mathbf{1}(|u| < 1)$  in which  $\mathbf{1}(\cdot)$  is the indicator function.

It is noted that the rate of convergence of the estimator depends on matrix  $D_n$

ensuring  $D_n^{-1}U'UD_n^{-1} \rightarrow_D \xi$ . The conventional rates of convergence for Cases (I)–(III) in Appendix A.1 may all be achievable.

In most of the applications, we only consider the case of  $k = 1$  to analyse co-trending relationship between two trending time series  $x_t$  and  $y_t$ . Also, since the basis functions  $\phi_j(\tau)$  are uncorrelated with  $y_t - x_t\beta_0$  under the common trend hypothesis for  $j = 1, 2, \dots, p$ , the moment conditions could be modified to

$$\frac{1}{n} \sum_{t=1}^n \phi_j(\tau_t)(y_t - x_t'\beta) \xrightarrow{p} 0, \quad (14)$$

for  $j = 1, 2, \dots, p$  when we only include  $\beta_0$  in the estimation. Then, the two-stage least squares estimator could be simplified as

$$\tilde{\beta}_{2sls} = (\tilde{X}'\tilde{\Phi}(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'\tilde{X})^{-1}\tilde{X}'\tilde{\Phi}(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'y, \quad (15)$$

and the following Corollary could be applied.

**Corollary 1:** Let **Assumptions 1** and **2** hold. We have for the case of  $k = 1$ , as  $n \rightarrow \infty$ ,

$$(\tilde{X}'\tilde{\Phi}(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'\tilde{X})^{1/2}(\tilde{\beta}_{2sls} - \beta_0) \xrightarrow{D} N(0, \lambda_e^2). \quad (16)$$

Once we have obtained an estimate for  $\gamma_0$ , we need to test whether it is a true co-trending parameter. In other words, based on the estimated value of  $\gamma_0$ , we are interested in testing the following hypotheses:

$$\mathbb{H}_0 : x_t \text{ and } y_t \text{ share a common trend}; \quad (17)$$

$$\mathbb{H}_1 : x_t \text{ and } y_t \text{ do not share a common trend}. \quad (18)$$

The hypotheses state that no trend shall be detected in  $e_t$  under  $\mathbb{H}_0$ . We borrow the idea of over-identification test in the conventional IV/GMM estimation theory that examines the validity of instrumental variables. That is, the IVs should be uncorrelated with the regression errors. In our paper, we consider an auxiliary regression

$$\hat{e}_t = \theta_0 + \theta_1\phi_1(\tau_t) + \theta_2\phi_2(\tau_t) + \dots + \theta_p\phi_p(\tau_t) + \zeta_t, \quad (19)$$

for  $\tau_t = t/n$  and  $t = 1, 2, \dots, n$ , where  $\zeta_t$  is the regression error. Hence, based on the

above regression model, “the IVs are uncorrelated with the residuals  $\widehat{e}_t$ ” is equivalent to “there is no trend in  $\widehat{e}_t$ ”. In other words, under  $\mathbb{H}_0$ , none of  $\theta_1, \theta_2, \dots, \theta_p$  in equation (19) should be significant. Following Sargan (1958), the test statistic is defined as

$$J_n(\widehat{\beta}_{2sls}) = \frac{\widehat{e}'\Phi(\Phi'\Phi)^{-1}\Phi'\widehat{e}}{\widehat{\lambda}_e^2}, \quad (20)$$

in which  $\widehat{\lambda}_e^2$  is a consistent estimator of  $\lambda_e^2$ , for example, as equation (13).

In order to establish Theorem 2 below, we will need to strengthen Assumption 1 as follows.

**Assumption 3:** (i) Let Assumption 1 hold.

(ii) Suppose that  $e_t$  is  $\alpha$ -mixing with coefficient  $\alpha(j)$  satisfying  $\sum_{j=1}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(j) < \infty$  for some  $\delta > 0$  such that  $E[|e_1|^{2+\delta}] < \infty$ .

(iii) Let Assumption 2(i) hold.

(iv) There is a consistent estimator  $\widehat{\lambda}_e^2$  of  $\lambda_e^2$  such that  $\frac{\widehat{\lambda}_e^2}{\lambda_e^2} \xrightarrow{P} 1$ .

**Theorem 2:** (Sargan’s test for common trend) Let Assumption 3 hold. Then as  $n \rightarrow \infty$ , we have

$$J_n(\widehat{\beta}_{2sls}) \xrightarrow{D} \chi_{p-k}^2. \quad (21)$$

The proofs of Theorems 1 and 2 are given in Appendix A.2 below.

## 3 Empirical Results

### 3.1 Estimating the co-trending relationship

In this section, we study the long-run relationship between global temperature anomalies (GTA) and the logarithm of  $CO_2$  equivalent total greenhouse gases (GHG), and test whether or not they share a common trend. We use *Legendre polynomials* to form a coordinate system and take them as instruments to estimate the common trend coefficient  $\beta_0$ .

The fitted model is reported as follows with estimated standard errors in the parentheses.

$$GTA_t = -11.815 + 2.018 \text{ GHG}_t + \widehat{e}_t, \quad (22)$$

(s.e.)            (0.5569)\*\*\*            (0.0953)\*\*\*

and the Sargan test for overidentifying restrictions is  $J = 4.124$  which is less than the

5% critical value of  $\chi_2^2$ . Therefore, we conclude that GTA and GHG have a common trend. Unit root tests also indicate that  $\hat{e}_t$  is stationary, thus indicating that GTA and GHG are co-trending. Based on equation (22), we could plot the fitted values of GTA and residuals  $\hat{e}_t$  in Figure 3 and 4 below.

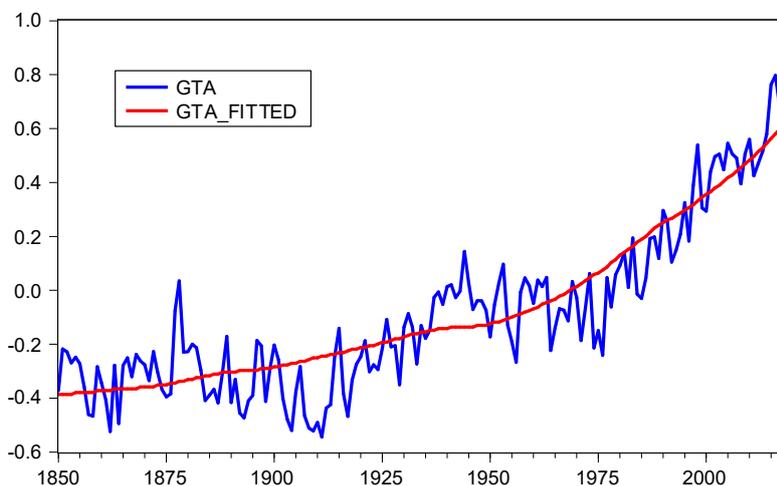


Figure 3: Fitted values of GTA by GHG.

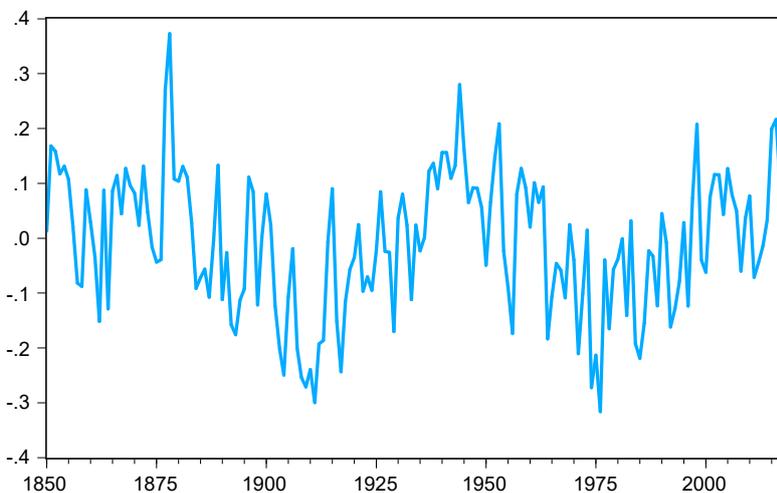


Figure 4: Time series plot of residuals  $\hat{e}_t$

### 3.2 Error correction mechanism

Engle and Granger (1987) established that  $I(1)$  variables that are cointegrated have an error correction representation. While in co-trending relationship, it is still reasonable

to believe that at least one of  $\Delta GTA_t$  or  $\Delta GHG_t$ , or both, must respond to previous equilibrium error  $GTA_t - \beta GHG_t$ . Therefore, we reveal the error correction mechanism by estimating the two equations below.

$$\widehat{\Delta GTA}_t = \text{const} - \frac{0.310}{(0.075)^{***}} \widehat{e}_{t-1} + \text{controls} \quad (23)$$

$$\widehat{\Delta GHG}_t = \text{const} - \frac{0.001}{(0.001)} \widehat{e}_{t-1} + \text{controls} \quad (24)$$

where  $\widehat{e}_{t-1} = GTA_{t-1} - 2.018GHG_{t-1}$ , and *controls* are the lagged values of  $\Delta GTA_t$  and  $\Delta GHG_t$ . The non-response of  $GHG$  to the equilibrium error is a necessary condition for exogeneity of  $GHG$ . In other words, emissions of greenhouse gases cause increase in global temperatures, while such an increase does not lead to much concentrations of greenhouse gases.

### 3.3 Residual diagnostics

The plot of the residuals in figure 4 shows a very regular pattern that seems to have a long cycle. Therefore, we use a deterministic cyclical component to further approximate the residual sequence by

$$\widehat{e}_t = \frac{0.104}{(0.016)} \sin \left( \frac{-0.346}{(0.308)} + \frac{4.652\pi t/n}{(0.172)} \right) + \widehat{u}_t, \quad (25)$$

$$\widehat{u}_t = \frac{0.351}{(0.073)} \widehat{u}_{t-1} + \widehat{\omega}_t, \quad \widehat{\sigma}_\omega = 0.096, \quad (26)$$

where  $\omega_t$  is a white noise. The estimated equation shows that the sine function corresponds to a period of 72 years with an amplitude of  $0.104^\circ\text{C}$ .

**Remark 4** While this is arguably too long to believe that the 72 years cycle can be reliably estimated from such a relatively short sample, it is within the 60-80 year band as the range of the Atlantic Multidecadal Oscillation (AMO) in temperatures that several climate scientists have documented using 8000 years of climate proxy records (corals, tree rings, ice cores) (See, for example, Knudsen et al. (2011)).

While the long-term effect of greenhouse gases on temperature is determined by equation (22), the sine cycle has non-negligible implications for the dynamics of global temperatures. From its peak to its trough (which takes 36 years), the estimated cycle dampens the change in the temperature by  $0.2^\circ\text{C}$ . Conversely, from its trough to its

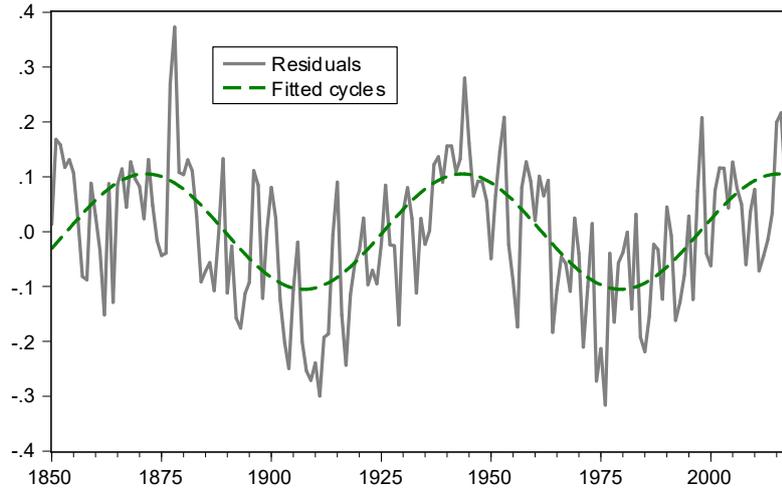


Figure 5: Residuals and the fitted sinusoid

peak (in 36 years), it adds  $0.2^{\circ}\text{C}$  to change in the temperature. While these effects in two consecutive 36-year periods cancel each other out, they are large enough to pose significant challenges for policy makers during the downward phase. In order to highlight this, Table 3 decomposes the predicted change in temperature to a component due to change in GHG and a component due to the cycle in four different periods. From this table we can infer that while predicted temperature rose by  $0.716^{\circ}\text{C}$  from 1944 to 2016, in the first 36 years no warming was felt because the increase in trend was almost completely offset by the downward phase of the cycle. The bulk of the increase in this period was realised in the last 36 years from 1980 to 2016. Assuming that it corresponds to a real phenomenon, this estimated cycle in the global temperature can make consensus building for an effective carbon policy very challenging during its downward phase. We discuss this further in the next section.

Table 3: Estimated warming effects caused by GHG and deterministic cycles

Time range	Change in predicted GTA		
	due to change in GHG	due to the cycle	total
1872-1908	0.101	-0.208	-0.107
1908-1944	0.121	0.208	0.329
1944-1980	0.266	-0.208	0.058
1980-2016	0.450	0.208	0.658

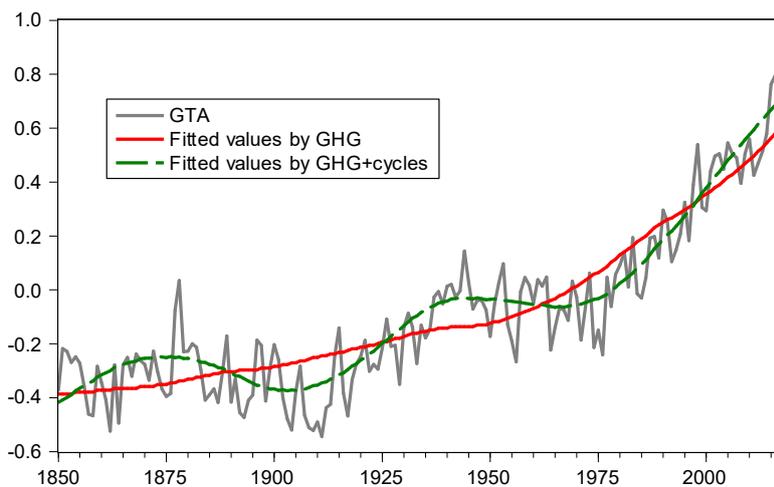


Figure 6: Fitted values with and without the cycle

### 3.4 Conditional forecasts of the future temperature

We have already shown that global temperatures and greenhouse gases share a common trend in the long-run. In addition, we have detected a deterministic cyclical component with 72 years' cycles in global temperatures. In the subsection, we forecast future temperature paths based on the estimated model in this paper under different scenarios of greenhouse gas emissions used in IPCC projections. The scenarios are the representative concentration pathways (RCPs) generated using integrated assessment models (IAMs) by the IPCC (See Meinshausen et al. 2011). We forecast the future temperature under two scenarios as plotted in Figure 7, the catastrophic path RCP8.5 and the medium-level path RCP6.0, respectively. The forecast paths of future temperatures are presented in Figure 8. Note that as the cyclical effect is taken into account in this paper, the forecasts exhibits different styles compared with previous ones, for example, Knutti & Sedlacek (2013). The detailed statistics are summarized in Table 4.

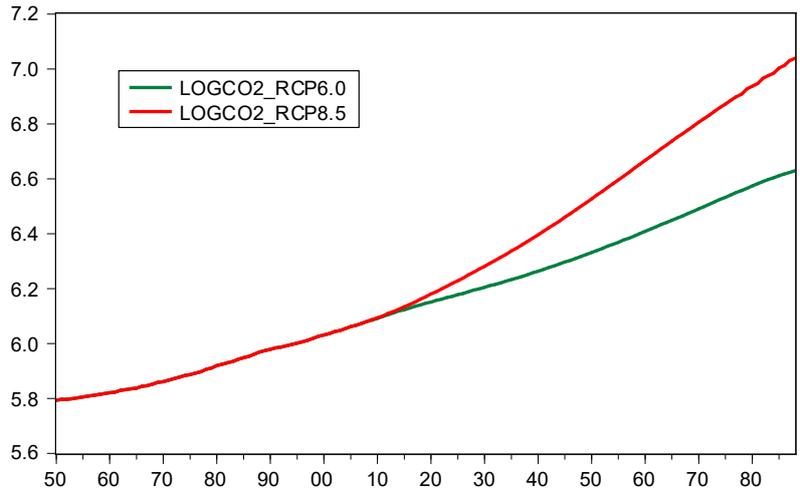


Figure 7: Representative concentration pathways, 1950-2089

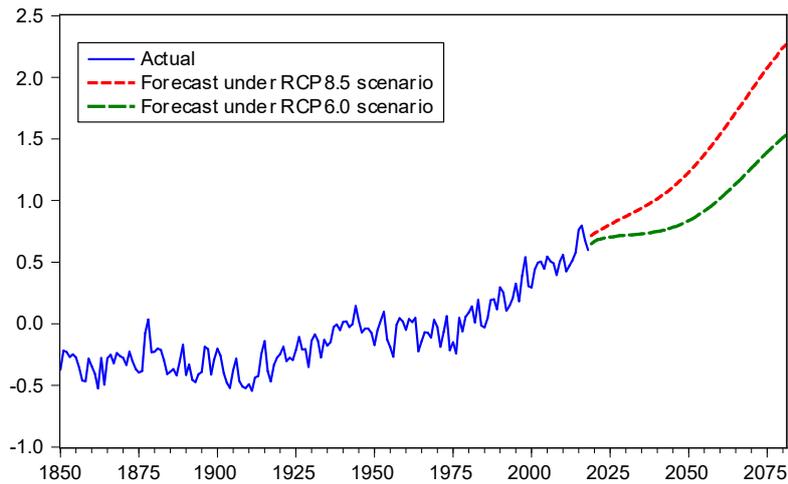


Figure 8: Forecast values of temperature based on two RCPs

If greenhouse gases follow the RCP6.0 scenario, the global temperature is forecast to increase by nearly  $1^{\circ}\text{C}$  from 2019 to 2089. However, the bulk of this increase will occur after 2053, and the increase from 2019 to 2052 will only be a modest  $0.2^{\circ}\text{C}$ . The reason is that the estimated sinusoidal cycle reached its peak in 2016 and will be in its downward phase until it reaches its trough in 2052. In the catastrophic greenhouse gases scenario RCP8.5, while the cyclical component will reduce the warming effect of greenhouse gases by 23% in the period 2019-2052, its effect will not be large enough to disguise the warming trend in the temperature. Moreover, the forecast change in

the 36 year period from 2053 to 2089 is an unprecedented 1.175°C, which is likely to have catastrophic consequences.

The point that we would like to emphasise is that the 72 cycle that is estimated to have reached its peak in 2016 and is forecast to be in its downward phase in the next 30 years, will make it difficult for the general public to appreciate the full effect of greenhouse gas emissions on global temperatures. In the countries such as Australia with little or no emission control policies, this may make the politics of introducing carbon policies even more challenging.

Table 4: Changes in temperature forecast, 2019-2089

Time range	$\Delta$ GTA conditional on RCP6.0			$\Delta$ GTA conditional on RCP8.5		
	due to $\Delta$ GHG	due to the cycle	total	due to $\Delta$ GHG	due to the cycle	total
2019 - 2052	0.402	-0.204	0.198	0.897	-0.204	0.693
2053 - 2089	0.565	0.208	0.773	0.967	0.208	1.175

## 4 Verifications by Numerical Simulations

In this section, we examine the performances of the test in Theorem 2 for common trends. We consider the data generating process

$$y_t = (1 - \rho)g_{1t} + \rho g_{2t} + e_t, \quad (27)$$

$$x_t = g_{1t} + v_t, \quad (28)$$

where  $v_t = 0.1v_{t-1} + \xi_{1t}$  and  $e_t = 0.1e_{t-1} + \xi_{2t}$ , in which  $\xi_{it} \sim i.i.d.N(0, 1)$  for  $i = 1, 2$ . Two trend components  $g_{1t}$  and  $g_{2t}$  are involved in the data generating mechanism. When  $\rho = 0$ , it gives the scenario of the null hypothesis (common trend). When  $0 < \rho < 1$ , it represents the scenario of the alternative hypothesis, and determines the degree of discrepancy between trends in the two time series. Specifically, we consider two kinds of trends.

- Case I: Stochastic trend

$g_{1t} = w_{1t}$ ,  $g_{2t} = 2w_{2t}$ , where  $w_{it} = w_{i,t-1} + \varepsilon_{it}$  in which  $\varepsilon_{it} \stackrel{i.i.d.}{\sim} N(0, 0.25)$  for  $i = 1, 2$ .

- Case II: Deterministic trend

$g_{1t} = \Phi_p \theta_1$ ,  $g_{2t} = \Phi_p \theta_2$ , where  $\theta_1, \theta_2$  are  $p \times 1$  vectors whose elements are randomly

selected from a uniform distribution from -3 to 3.  $\Phi_p$  are deterministic basis functions defined as follows.

The basis functions we employ in the simulation are *normalized Legendre polynomials*  $\Phi_p(\tau_t) = (\phi_0(\tau_t), \phi_1(\tau_t), \dots, \phi_p(\tau_t))$  on  $[0,1]$ . Specifically,

$$\begin{aligned}\phi_0(\tau) &= 1, & \phi_1(\tau) &= \sqrt{3}(2\tau - 1), & \phi_2(\tau) &= \sqrt{5}(6\tau^2 - 6\tau + 1), \\ \phi_3(\tau) &= \sqrt{7}(20\tau^3 - 30\tau^2 + 12\tau - 1), & \phi_4(\tau) &= \sqrt{9}(70\tau^4 - 140\tau^3 + 90\tau^2 - 20\tau + 1), \\ \phi_5(\tau) &= \sqrt{11}(252\tau^5 - 630\tau^4 + 560\tau^3 - 210\tau^2 + 30\tau - 1), \\ \phi_6(\tau) &= \sqrt{13}(924\tau^6 - 2772\tau^5 + 3150\tau^4 - 1680\tau^3 + 420\tau^2 - 42\tau + 1).\end{aligned}$$

The simulation is repeated for 10,000 times for each of sample sizes 200, 400, and 800, and  $p = 4, 5$  and 6, respectively.

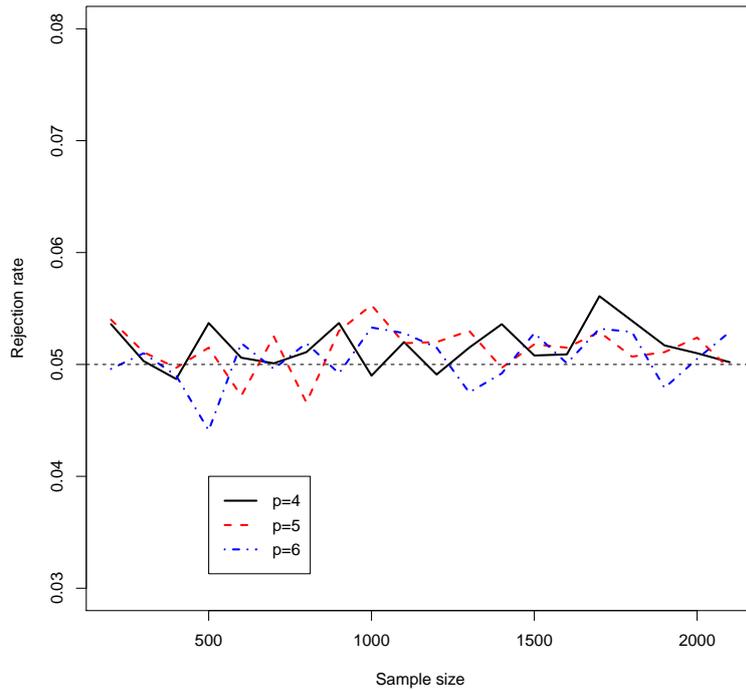


Figure 9: Sizes of the  $J$ -test in Case I under different sample sizes  $n$ .

In Figure 9 and Figure 10, we report the size of the  $J$ -test for both cases of stochastic and deterministic trends. We set  $\rho = 0$  to simulate scenario of the null hypothesis and plot the rejection rate of the  $J$ -test under 5% significance level against sample sizes from 200 to 2100. The two Figures show that the rejection rate is close to 5% and there is nearly no size distortion for the  $J$ -test.

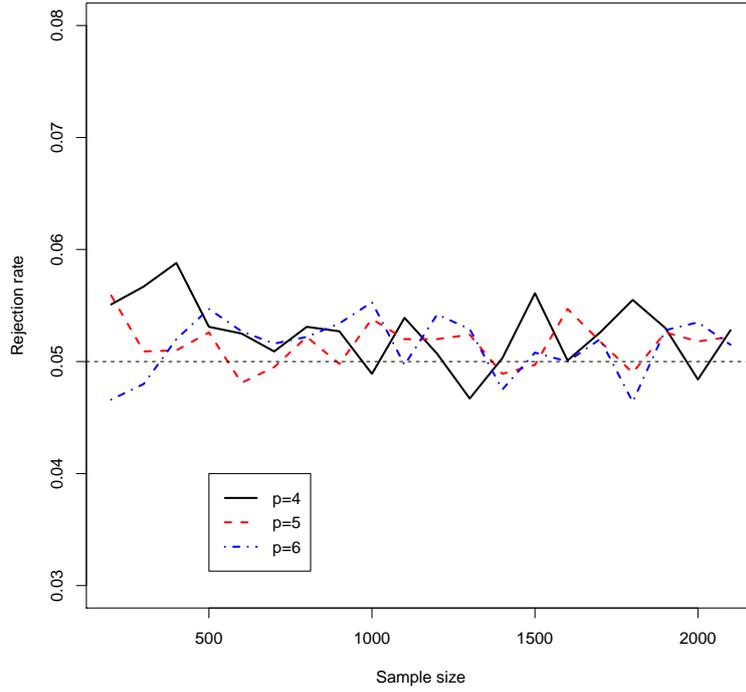


Figure 10: Sizes of the  $J$ -test in Case II under different sample sizes  $n$ .

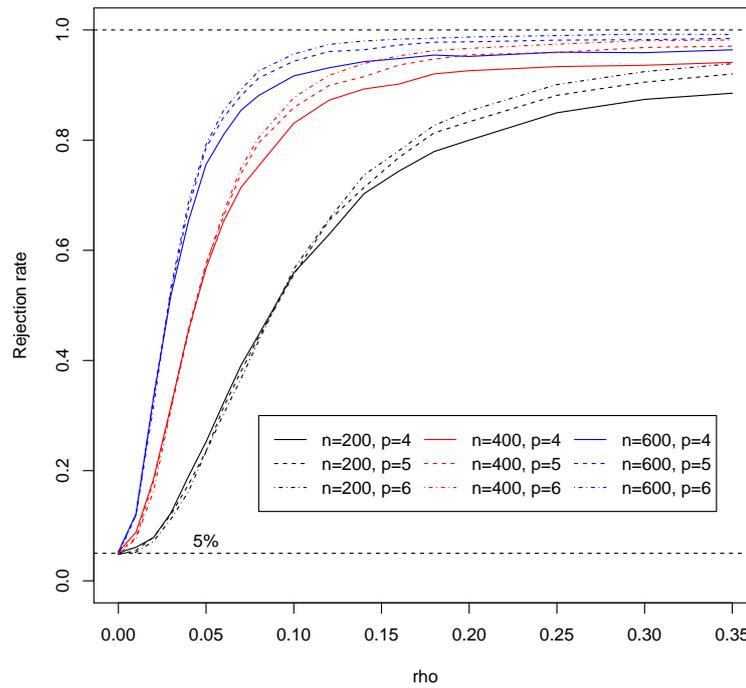


Figure 11: Rejection rates of the  $J$ -test in Case I under different values of  $\rho$ .

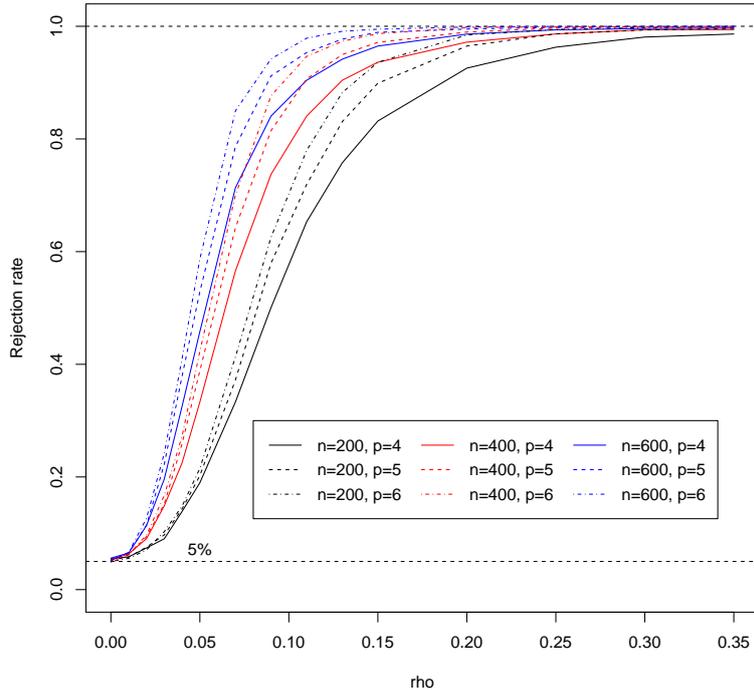


Figure 12: Rejection rates of the  $J$ -test in Case II under different values of  $\rho$ .

In Figure 11 and Figure 12, we report the rejection rates of our  $J$ -test under 5% significance level for Case I and Case II, respectively. Note that  $\rho$  varies from 0 to 0.35. Both Figure 11 and Figure 12 show that when the null hypothesis of common trend is true ( $\rho = 0$ ), the rejection rate is very close to 5% as the significance level we choose. While under the alternative hypothesis ( $\rho > 0$ ), we find that powers of the  $J$ -test significantly depends on  $\rho$ , the degree of no common trend and sample size  $n$ . For a given sample size, the power of the  $J$ -test grows with  $\rho$  as it controls the degree of no common trend. While for a give value of  $\rho$ , a larger sample size means more information to distinguish the alternative hypothesis from the null hypothesis, therefore, leading to a higher rejection rate.

## 5 Conclusion

In this paper, we proposed a common features approach to investigate the long-run relationship between global temperature anomalies and greenhouse gases without the need to determine the exact nature of trend in the data. We also showed that the test for over-identifying restrictions could be used to test for common trends.

We found that global temperatures and greenhouse gases share a common trend,

which indicates a stable long-run relationship between them. In addition, global temperature series exhibit a cyclical component with a period of approximately 72 years with a non-trivial amplitude. We also studied the error correction mechanism between the two trending time series and found that global temperatures respond to the equilibrium error of the common trend relationship, while greenhouse gas emissions do not.

With our estimated model that includes both the co-trending relationship and the cyclical characteristics, we forecast future temperatures based on the medium-level and the high-level emission pathways RCP6.0 and RCP8.5. While our long-term forecasts are in the range of those produced by the IPCC, we highlight the importance of the interaction between the trend and the cycle in global temperatures in the next 30 to 40 years. We estimate that the cycle in global temperatures has reached its peak in 2016 and will be in its downward phase in the next 30 years. This can potentially cause incorrect inference from data during this phase. Countries with some emission control policies may infer overly optimistic conclusions about the success of their policies in reducing the warming trend in temperatures. At the same time, countries with little or no emission control policies may conclude that their lack of action was warranted.

## References

- Anderson, H. M. & Vahid, F. (1998), ‘Testing multiple equation systems for common nonlinear components’, *Journal of Econometrics* **84**(1), 1–36.
- Andrews, D. (1991), ‘Heteroskedasticity and autocorrelation consistent co-variance matrix estimation’, *Econometrica* **59**(3), 817–854.
- Bierens, H. J. (1997), ‘Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate’, *Journal of Econometrics* **81**(1), 29–64.
- Billingsley, P. (1999), *Convergence of Probability Measures (2nd ed.)*, John Wiley & Sons.
- Breusch, T. & Vahid, F. (2011), ‘Global temperature trends’, *Department of Econometrics and Business Statistics Working Paper 4-11* .  
**URL:** <https://ideas.repec.org/p/msh/ebswps/2011-4.html>
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y. & Park, S. (2016), ‘Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate’.  
**URL:** [https://economics.missouri.edu/sites/default/files/wp-files/cc1\\_20161219.pdf](https://economics.missouri.edu/sites/default/files/wp-files/cc1_20161219.pdf)

- Chen, L. (2017), Trending time series models with endogeneity, PhD thesis, Department of Econometrics and Business Statistics, Monash University.
- Choi, I. (2017), ‘Efficient estimation of nonstationary factor models’, *Journal of Statistical Planning and Inference* **183**, 18–43.
- Cubasch, U. & Meehl, G. (2001), Projections of future climate change, in ‘Climate Change 2001: The Scientific Basis’, Cambridge University Press, pp. 525–582.
- DeJong, D. N., Nankervis, J. C., Savin, N. E. & Whiteman, C. H. (1992), ‘The power problems of unit root test in time series with autoregressive errors’, *Journal of Econometrics* **53**(1-3), 323–343.
- Dong, C. & Gao, J. (2019), ‘Expansion of lévy process functionals and its application in econometric estimation’, *Econometric Reviews* **39**, 125–150.
- Engle, R. F. & Kozicki, S. (1993), ‘Testing for common features’, *Journal of Business & Economic Statistics* **11**(4), 369–380.
- Estrada, F., Gay, C. & Sánchez, A. (2010), ‘A reply to does temperature contain a stochastic trend? evaluating conflicting statistical results by rk kaufmann et al’, *Climatic Change* **101**(3), 407–414.
- Estrada, F., Perron, P. & Martínez-López, B. (2013), ‘Statistically derived contributions of diverse human influences to twentieth-century temperature changes’, *Nature Geoscience* **6**(12), 1050–1055.
- Fan, J. & Yao, Q. (2003), *Nonlinear Time Series: Parametric and Nonparametric Methods*, Springer, New York.
- Gao, J. & Hawthorne, K. (2006), ‘Semiparametric estimation and testing of the trend of temperature series’, *The Econometrics Journal* **9**(2), 332–355.
- Gay, G. C., Estrada, F. & Sánchez, A. (2009), ‘Global and hemispheric temperatures revisited’, *Climatic Change* **94**(3-4), 333–349.
- 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.
- Kaufmann, R. K., Kauppi, H. & Stock, J. H. (2006), ‘Emissions, concentrations, & temperature: a time series analysis’, *Climatic Change* **77**(3-4), 249–278.
- Kaufmann, R. K., Kauppi, H. & Stock, J. H. (2010), ‘temperature contain a stochastic trend? evaluating conflicting statistical results’, *Climatic Change* **101**(3-4), 395–405.
- Kaufmann, R. K. & Stern, D. I. (2002), ‘Cointegration analysis of hemispheric temperature relations’, *Journal of Geophysical Research* **107**(D2), ACL 8–1 to 8–10.

- Knudsen, M. F., Seidenkrantz, M.-S., Jacobsen, B. H. & Kuijpers, A. (2011), ‘Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years’, *Nature Communications* **2**, 178.
- Knutti, R. & Sedlacek, J. (2013), ‘Robustness and uncertainties in the new CMIP5 climate model projections’, *Nature Climate Change* **3**, 369–373.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992), ‘Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?’, *Journal of Econometrics* **54**(1), 159–178.
- Meinshausen, M., Smith, S., Calvin, K., Daniel, J., Kainuma, M., Lamarque, J., Matsumoto, K., Montzka, S., Raper, S., Riahi, K., Thomson, A., Velders, G. & Vuuren, D. (2011), ‘The RCP greenhouse gas concentrations and their extensions from 1765 to 2300’, *Climatic Change* **109**, 213–241.
- Mills, T. C. (2009), ‘How robust is the long-run relationship between temperature and radiative forcing?’, *Climatic Change* **94**(3-4), 351.
- Mitchell, J. & Karoly, D. (2001), Detection of climate change and attribution of causes, in ‘Climate Change 2001: The Scientific Basis’, Cambridge University Press, pp. 427–431.
- Nelson, C. R. & Plosser, C. R. (1982), ‘Trends and random walks in macroeconomic time series: Some evidence and implications’, *Journal of Monetary Economics* **10**(2), 139–162.
- Newey, W. K. & West, K. D. (1987), ‘A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**(3), 703–708.
- Newey, W. K. & West, K. D. (1994), ‘Automatic lag selection in covariance matrix estimation’, *The Review of Economic Studies* **61**(4), 631–653.
- Park, J. & Phillips, P. C. B. (2001), ‘Nonlinear regression with integrated time series’, *Econometrica* **69**(1), 117–161.
- Park, J. Y. & Phillips, P. C. B. (1988), ‘Statistical inference in regressions with integrated processes: Part 1’, *Econometric Theory* **4**(3), 468–497.
- Phillips, P. C. B. (1998), ‘New tools for understanding spurious regressions’, *Econometrica* **66**(6), 1299–1325.
- Phillips, P. C. B. (2001), ‘Trending time series and macroeconomic activity: Some present and future challenges’, *Journal of Econometrics* **100**(1), 21–27.
- Phillips, P. C. B. (2003), ‘Laws and limits of econometrics’, *The Economic Journal* **113**, C26–C52.
- Phillips, P. C. B. (2005), ‘Challenges of trending time series econometrics’, *Mathematics and Computers in Simulation* **68**(5), 401–416.

Phillips, P. C. B. (2010), ‘The mysteries of trend’, *Macroeconomic Review* **9**(2), 82–89.

Phillips, P. C. B. & Hansen, B. (1990), ‘Statistical inference in instrumental variables regression with I(1) processes’, *Review of Economic Studies* **57**(1), 99–125.

Sargan, J. D. (1958), ‘The estimation of economic relationships using instrumental variables’, *Econometrica* **26**(3), 393–415.

Vahid, F. & Engle, R. F. (1993), ‘Common trends and common cycles’, *Journal of Applied Econometrics* **8**(4), 341–360.

## 6 Appendix

### A.1: Justification of Assumptions 1–3

(a) **Assumption 1** establishes a co-trending relationship between trending time series  $y_t$  and  $x_t$ . Trends cancel out when they are combined by the coefficient (vector)  $\beta_0$ . (b) **Assumption 2**(i) defines a system of orthogonal basis function. For example, it could be the first  $p$  Legendre Polynomials normalized on  $[0,1]$ . (c) **Assumption 2**(ii) requires that the introduction of the  $\Phi$  matrix can ensure that  $U$  and  $\varepsilon$  are uncorrelated. (d) Assumption 2(iii) then imposes a high-level condition to ensure the joint convergence. This can be verified under some low-level conditions imposed on  $(x_t, e_t)$  (such as Assumption 2.1 of Park & Phillips (2001) for the integrated time series case—Case (II) below). Assumption 2(iii) covers many different types of nonstationarity, such as the deterministically nonstationary case (I):  $x_t = g(\tau_t) + v_t$ , the stochastically nonstationary case (II):  $x_t = x_{t-1} + v_t$ , and the mixture of stochastically and deterministically nonstationary case (III):  $x_t = x_{t-1} + g(\tau_t) + v_t$ , in which  $\tau_t = \frac{t}{n}$ .

In order to further justify the reasonableness of Assumption 2(ii) and (iii), we consider the case of  $k = 1$  and  $W = I_p$ . In this case, the elements of  $U'U$  are of the forms:  $(\sum_{t=1}^n \phi_1(\tau_t))^i (\sum_{t=1}^n \phi_1(\tau_t)x_{1t})^j$  for  $i, j = 0, 1, 2$ , and the elements of  $U'\varepsilon$  are of the forms:  $\sum_{t=1}^n \phi_1(\tau_t) \sum_{s=1}^n \phi_1(\tau_s)e_s$  and  $\sum_{t=1}^n \phi_1(\tau_t)x_{1t} \sum_{s=1}^n \phi_1(\tau_s)e_s$ . Choi (2017) imposes a set of similar conditions for the pure unit-root setting in Case (II). In the general forms considered above, existing results from Case (II) are not applicable to Cases (I) and (III). However, Lemmas B.1 and B.2 of Dong & Gao (2019) may be used to justify Assumption 2(iii).

In **Assumption 2**(iii), specifically, one may choose  $D_n$  simply as a diagonal matrix to reflect the trending strength of  $x_{it}$  for  $i = 1, 2, \dots, k$ . For the case of  $k = 1$ , for example, let  $v_t \sim i.i.d.N(0, 1)$ , (1) In Case (I):  $x_t = g(\tau_t) + v_t$ , we have  $n^{-1} \sum_{t=1}^n x_t \Phi(\tau_t) \xrightarrow{P} \int_0^1 g(r) \Phi(r) dr$ ; (2) In Case (II):  $x_t = \sum_{s=1}^t v_s$ , we have  $n^{-1} \sum_{t=1}^n x_t \Phi(\tau_t) \xrightarrow{D} \int_0^1 B(r) \Phi(r) dr$ ,

where  $B(\cdot)$  is a Brownian motion; (3) In Case (III):  $x_t = x_{t-1} + g(\tau_t) + v_t$ , we have  $n^{-1} \sum_{t=1}^n \frac{x_t}{n} \Phi(\tau_t) \xrightarrow{P} \int_0^1 \int_0^v g(u) \Phi(v) du dv$ .

Meanwhile, **Assumption 3(ii)** is quite standard. **Assumption (iv)** can be justifiable by using existing results, such as Andrews (1991).

## A.2: Proofs of the main theorems

### *Proof of Theorem 1*

Note that the two-stage least squares estimator is established as

$$\hat{\gamma}_{2sls} = (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1} X' \Phi(\Phi' \Phi)^{-1} \Phi' y.$$

Replacing  $y$  using  $y = X\beta + e$ , we have

$$\hat{\gamma}_{2sls} - \gamma_0 = (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1} X' \Phi(\Phi' \Phi)^{-1} \Phi' e, \quad (29)$$

which, along with Assumption 2(iii), implies

$$\begin{aligned} (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{1/2} (\hat{\gamma}_{2sls} - \gamma_0) &= (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1/2} X' \Phi(\Phi' \Phi)^{-1} \Phi' e \\ &\xrightarrow{D} N(0, \lambda_e^2 I_{(k+1)}), \end{aligned} \quad (30)$$

by using the continuous mapping theorem (see, for example, Billingsley (1999)). This completes the proof of Theorem 1.

### *Proof of Theorem 2*

Note that  $\hat{e} = y - X\hat{\gamma}_{2sls}$  and

$$\hat{\gamma}_{2sls} - \gamma_0 = (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1} X' \Phi(\Phi' \Phi)^{-1} \Phi' e,$$

we have

$$\begin{aligned} (\Phi' \Phi)^{-1/2} \Phi' \hat{e} &= (\Phi' \Phi)^{-1/2} \Phi' (y - X\hat{\gamma}_{2sls}) \\ &= (\Phi' \Phi)^{-1/2} \Phi' (y - X\gamma_0) - (\Phi' \Phi)^{-1/2} \Phi' X (\hat{\gamma}_{2sls} - \gamma_0) \\ &= \left[ I - (\Phi' \Phi)^{-1/2} \Phi' X (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1} X' \Phi(\Phi' \Phi)^{-1/2} \right] (\Phi' \Phi)^{-1/2} \Phi' e \\ &= \widehat{M}_1(n) \widehat{M}_2(n), \end{aligned}$$

where  $\widehat{M}_1(n) = I - (\Phi' \Phi)^{-1/2} \Phi' X (X' \Phi(\Phi' \Phi)^{-1} \Phi' X)^{-1} X' \Phi(\Phi' \Phi)^{-1/2}$  and  $\widehat{M}_2(n) = (\Phi' \Phi)^{-1/2} \Phi' e$ .

Under Assumption 3, similar to the proof of Theorem 2.21(i) of Fan & Yao (2003), as  $n \rightarrow \infty$ ,

we have

$$\widehat{M}_2(n) \xrightarrow{D} N(0, \lambda_e^2 I_{(p+1)}).$$

Meanwhile,  $\widehat{M}_1(n)$  is a idempotent  $(p+1) \times (p+1)$  symmetric matrix with rank  $p-k$  since

$$\begin{aligned} \text{tr}(M_1(n)) &= \text{tr}(I_{p+1}) - \text{tr}((\Phi'\Phi)^{-1/2}\Phi'X (X'\Phi(\Phi'\Phi)^{-1}\Phi'X)^{-1} X'\Phi(\Phi'\Phi)^{-1/2}) \\ &= \text{tr}(I_{p+1}) - \text{tr}(X'\Phi(\Phi'\Phi)^{-1/2}(\Phi'\Phi)^{-1/2}\Phi'X (X'\Phi(\Phi'\Phi)^{-1}\Phi'X)^{-1}) \\ &= \text{tr}(I_{p+1}) - \text{tr}(I_{k+1}) = p - k. \end{aligned}$$

Also, under Assumption 3,  $\widehat{\lambda}_e^2$  is a consistent estimator for  $\lambda_e^2$ . Therefore, based on the definition of  $\widehat{M}_1(n)$  and  $\widehat{M}_2(n)$ , we have

$$\begin{aligned} J_n(\widehat{\beta}_{2sls}) &= \widehat{e}'\Phi(\Phi'\Phi)^{-1}\Phi'\widehat{e}/\widehat{\lambda}_e^2 = \widehat{M}_2(n)'\widehat{M}_1(n)'\widehat{M}_1(n)\widehat{M}_2(n)/\widehat{\sigma}_u^2 \\ &= \widehat{M}_2(n)'\widehat{M}_1(n)\widehat{M}_2(n)/\widehat{\lambda}_e^2 \xrightarrow{d} \chi_{p-k}^2. \end{aligned}$$

The above result is obtained as  $\widehat{M}_2(n)$  converges to a  $(p+1)$ -dimensional Normal distribution, while  $\widehat{M}_1(n)$  converges in probability to a  $(k+1) \times (k+1)$  dimensional idempotent and symmetric matrix with rank  $p-k$ , and  $\widehat{\lambda}_e^2$  converges to  $\lambda_e^2$  in probability. ■