MODELING LONG CYCLES

Vancouver School of Economics, University of British Columbia

NATASHA KANG* AND VADIM MARMER

ABSTRACT. Recurrent boom-and-bust cycles are a salient feature of economic and financial history. Cycles found in the data are stochastic, often highly persistent, and span substantial fractions of the sample size. We refer to such cycles as "long". In this paper, we develop a novel approach to modeling cyclical behavior specifically designed to capture long cycles. We show that existing inferential procedures may produce misleading results in the presence of long cycles, and propose a new econometric procedure for the inference on the cycle length. Our procedure is asymptotically valid regardless of the cycle length. We apply our methodology to a set of macroeconomic and financial variables for the U.S. We find evidence of long stochastic cycles in the standard business cycle variables, as well as in credit and house prices. However, we rule out the presence of stochastic cycles in asset market data. Moreover, according to our result, financial cycles as characterized by credit and house prices tend to be twice as long as business cycles.

KEY WORDS. Stochastic cycles, autoregressive processes, local-to-unity asymptotics, confidence sets, business cycle, financial cycle

JEL CLASSIFICATION: C12, C22, C5, E32, E44

1. Introduction

This paper develops an econometric framework for the inference on the cyclical properties of time series. We are particularly interested in stochastic cycles arising due to persistent low-frequency oscillatory impulse responses. The period of such cycles spans a substantial fraction of a sample, and the econometrician would be able to observe only a handful of peaks and troughs in data. We refer to such cycles as "long". It is also important to emphasize the stochastic nature of such cycles. While their peaks may appear to be regularly spaced, the timing of the peaks and even their number in a sample are determined by particular realizations of the shocks.

Long cycles are prevalent in macroeconomic and financial data. In a recent paper, Beaudry, Galizia, and Portier (2020) estimate that many variables have cycles of length of approximately

Date: April 26, 2021.

*Corresponding author. Email:natk@student.ubc.ca.

We thank Paul Beaudry for numerous discussions and comments, and Kevin Song for his contributions and discussions during early stages of the project. We also thank Jasmine Hao and Hiro Kasahara for helpful comments and suggestions. The financial support of the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. This research was enabled in part by support provided by Compute Canada (www.computecanada.ca).

32–40 quarters, which can correspond to 15% and even 20% of their observed samples. Using data from 1960 to 2011, Drehmann, Borio, and Tsatsaronis (2012) estimate that the length of the cycle for credit is 18 years or approximately 35% of their sample. The first contribution of our paper is to show that existing inferential procedures for the cycle length can be distorted in such cases. We find that substantial distortions occur when the cycle length exceeds 25% of the sample size.

In our second contribution, we propose a new econometric procedure for the inference on the periodicity of cycles. The novel aspect of our methodology is that it is specifically designed to take into account the possibility of long persistent stochastic cycles. Our procedure produces a confidence intervals for the cycle length that has the following property: its asymptotic coverage probability is correct regardless of the cycle length. Thus, the confidence interval is asymptotically valid both when the period is small relative to the sample size, and when it spans a substantial fraction of observed data. No other procedure in the existing literature has this property. When the data generating process (DGP) is acyclical, in large samples our procedure is expected to produce empty confidence intervals for the cycle length. Hence, the procedure can be used to rule out the cyclical behavior.

Stochastic cycles naturally arise in AR(2) models with complex roots. For example, Sargent (1987) describes the region for the values of the autoregressive coefficients that produce a spike in the spectrum in the interior of the $[0, \pi]$ range.[1] The period of such cycles is determined by the values of the autoregressive coefficients and has a fixed length in time units. According to this model, the cycle length would amount to a negligible fraction of the sample size in large samples. Thus, the long-cycle characteristics of data are not preserved asymptotically: while in a finite sample the cycle length can represent a substantial fraction of the sample size, it would be negligible in the asymptotic approximation. As a result, conventional asymptotic approximations to the finite sample distributions of estimators and statistics would be inaccurate.

The problem can be better understood from the perspective of the literature concerned with inference on the largest autoregressive root (Stock, 1991; Andrews, 1993; Hansen, 1999; Elliott and Stock, 2001; Mikusheva, 2007, 2012).[2] Long cycles correspond to low-frequency fluctuations, which can be indicative of autoregressive roots near unity. However, it has been shown in the literature that when autoregressive roots are close to unity, the conventional asymptotic theory does not provide an accurate approximation to the finite sample distributions of estimators and statistics. More accurate approximations can be obtained using the so-called local-to-unity asymptotics developed in Phillips (1987, 1988). Moreover, the local-to-unity asymptotics nests the conventional asymptotics as a limiting case. This is achieved by modeling the autoregressive coefficients as drifting with the sample size toward unity at the appropriate rate. Such specifications are consistent with stationary parameter values for finite-sample DGPs and can produce any desired level of persistence. Note also that investigating the asymptotic properties of statistics

[1] The region with an interior spike in the spectrum is a subset of the region with complex roots.

[2] Equivalently, the sum of the autoregressive coefficients.

under specifications with drifting coefficients is required to verify the uniform size of inferential procedures (Andrews, Cheng, and Guggenberger, 2020).

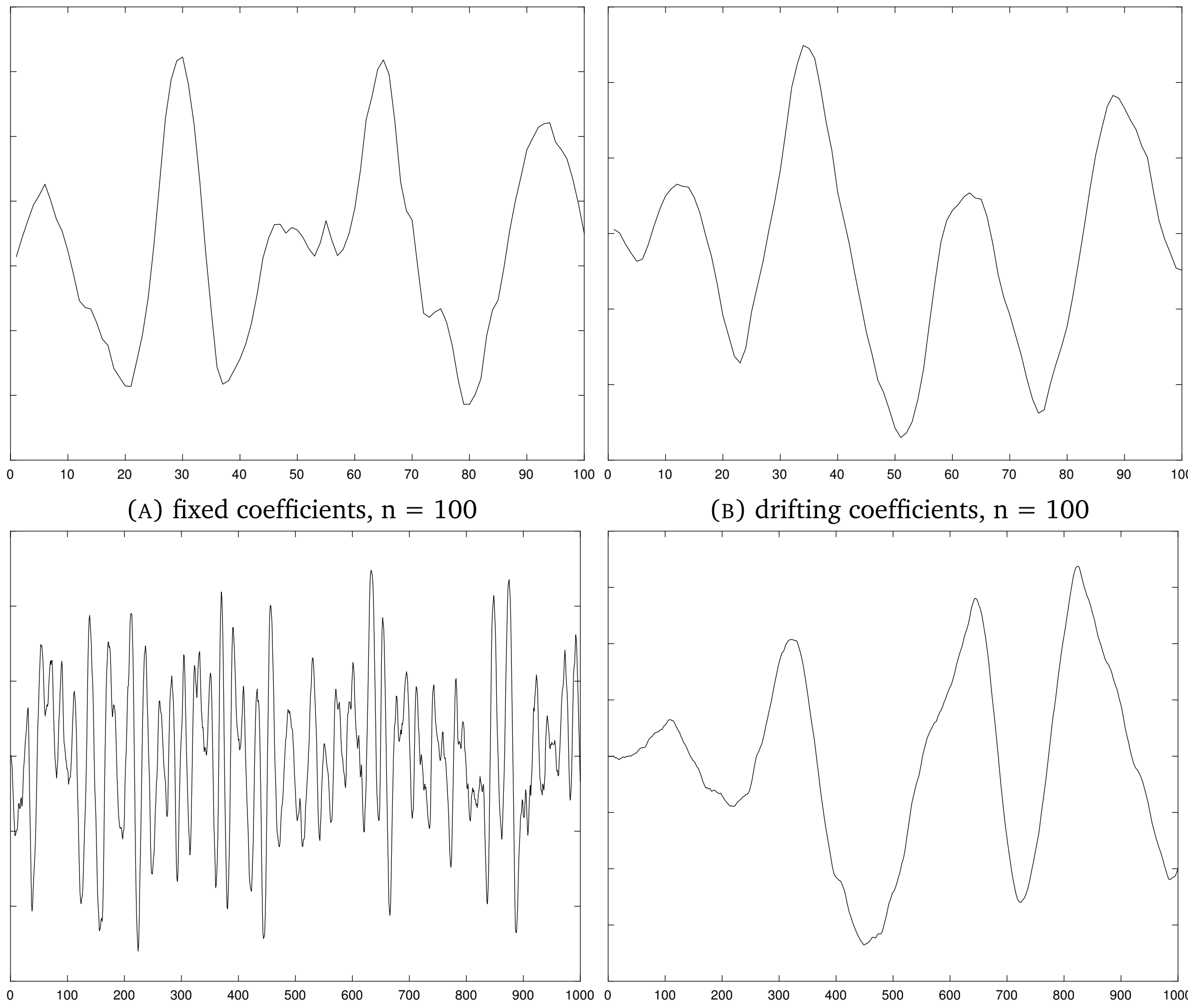
Borrowing from the insights of this literature, we model long cycles using autoregressive specifications with two complex roots as in Sargent (1987), however, the complex roots are drifting and local to one. The drifting specification allows us to preserve the long-cycle feature asymptotically: regardless of the size of the sample, the length of a cycle *as a fraction of the sample size* remains the same. This point is illustrated in Figure 1. It shows the difference between the simulated sample paths of cyclical processes generated using the fixed-coefficient and drifting-coefficient DGPs for small and large sample sizes. The figure demonstrates that in the model with fixed autoregressive coefficients, by relying on asymptotic approximations one would distort the cyclical properties of data. However, the long-cycle properties are preserved in the limit by relying on asymptotic approximations with drifting coefficients.

While we adopt the local-to-unity modeling approach of Phillips (1987, 1988), there are a number of important differences between our paper and the existing literature. First, since the roots are complex conjugates, the long-cycle process is near I(2) instead of near I(1) in the sense that there are two near-unity roots. Even though our processes are near I(2), they are stationary in finite samples. Moreover, our data generating process captures a crucial feature of macroeconomic and financial data: persistent low-frequency stochastic cycles. The latter point is demonstrated in the empirical application in Section 7. Second, due to the presence of complex roots, we obtain different asymptotic approximations from those in the literature. The previous results for local-to-unity processes cannot accommodate long cycles, and we develop a novel theory for such processes.

Thus, the third contribution of this paper is to the literature concerned with inference on autoregressive roots. Most of that literature is focused on the largest autoregressive roots (or the sum of the autoregressive coefficients).[3] In our paper, we also focus on another important and empirically relevant aspect of data: low-frequency stochastic oscillations. To fully analyze such processes, one has to consider complex autoregressive roots. Our results also lay out the foundation for a new econometric framework that besides the inference on cyclicality, can also be used to study cointegrating long cycles and phase shifts in macro-financial aggregates. However, the latter are outside the scope of this paper and left for future research.

Our paper is further related to the literature on complex unit roots (Bierens, 2001; Gregoir, 2006). Unlike Bierens (2001), our data generating process is stationary in finite samples, and persistent oscillations are achieved through local-to-unity modeling. Local-to-unity modeling with complex roots has been previously considered by Gregoir (2006). However, Gregoir (2006) only considers oscillations at fixed frequencies, while we focus on oscillations at local-to-zero frequencies. This crucial feature allows us to accommodate arbitrary long cycles with persistent oscillations at very low frequencies, which is an important attribute of many macroeconomic and financial time series as demonstrated in Section 7. Our emphasis on low-frequency oscillations is

[3] In a recent paper, Dou and Müller (2021) propose a generalized local-to-unity ARMA model with multiple local-to-unity autoregressive roots balanced local-to-unity roots in the moving average component. They do not consider cyclical behavior, and their limiting distributions are different from those arising in our case.



(A) fixed coefficients, n = 100

(B) drifting coefficients, n = 100

(C) fixed coefficients, n = 1000

(D) drifting coefficients, n = 1000

FIGURE 1. Time plots of AR(2) processes with fixed and drifting coefficients. In the standard AR(2) specification with fixed coefficients, the period stays the same as the sample size increases. In contrast, the period of an AR(2) process with drifting coefficients grows proportionally with the sample size.

not inconsequential. As we show in Section 5, the conventional asymptotic theory fails to approximate the finite-sample sample distributions of estimators and statistics not only when a process is persistent, but also when it oscillates at a low-frequency. The validity of asymptotic approximations for long-cycle data hinges on both of these dimensions and following the arguments of Andrews, Cheng, and Guggenberger (2020), one must verify the size properties of inferential procedures under drifting local-to-zero frequencies in combination with local-to-unity persistence parameters.

The fourth contribution of this paper is empirical, where we implement our procedure to study the cyclical properties of key macroeconomic and financial indicators using U.S. data. Recurrent boom-and-bust cycles are a salient feature of economic and financial history. A long-standing interest in understanding these ups and downs in the macro-financial aggregates has led to a vast body of literature on business cycles, and a resurgence of research on financial cycles post the financial crisis-induced Great Recession of 2008. Among these strands of work is the empirical characterization of business and financial cycles. The traditional approach to such characterization is to identify turning points or peaks and troughs in the time series using the dating algorithms of Bry and Boschan (1971) and Harding and Pagan (2002). Based on the turning-point analysis, Drehmann, Borio, and Tsatsaronis (2012) highlight the importance of medium-term cycles that last 18 years for credit, 11 years for GDP and 9 years for equity prices. These findings are in-line with studies using frequency-based bandpass filters (see Aikman, Haldane, and Nelson, 2015; Comin and Gertler, 2006).

The cyclical properties of data have also been formally examined in the literature using a variety of methods, including direct and indirect spectrum estimation (e.g. A'Hearn and Woitek, 2001; Strohsal, Proaño, and Wolters, 2019) and structural time series modeling (e.g. Harvey, 1985; Rünstler and Vlekke, 2018). However, they rely on the conventional asymptotic approximations that may produce misleading results with long cycle data, as we argue in this paper. For example, we show that the periodogram-based estimator is asymptotically biased in the case of long cycles.

Using our methodology, we find that long cycles cannot be ruled out for macroeconomic series such as the real GDP per capita, unemployment rates, and hours per capita. Our results suggest the possibility of much longer cycles than those previously reported in the literature. In addition, we find that financial variables such as credit to non-financial sector and home prices exhibit long cycles that are even longer than those for the macro variables. Our results support the position that financial cycles operate at a lower frequency than business cycles. However, our most striking result is that we decisively reject stochastic cycles for asset market variables such as the volatility index, credit risk premium, and equity prices. This suggests that the mechanism for asset market fluctuations is different from that of macroeconomic variables and financial variables such as credit and home prices. Importantly, this finding rejects a view suggested in the macro-finance literature that asset prices and economic fluctuations are driven by the same underlying forces: time-varying risk premiums and risk-bearing capacity (see Cochrane, 2017).

The rest of the paper is organized as follows. In Section 2, we present our modeling approach for long cycles. Section 3 presents our core asymptotic results. The results are extended in Section 4 to allow for linear time trends and deterministic cycles. In Section 5, we discuss the size distortions of the conventional inference approach. Section 6 describes our procedure for constructing confidence intervals for the cycle length. Section 7 presents our empirical results. In Appendix A, we discuss the asymptotic properties of the periodogram for long-cycle processes.

2. A model for long cycles

In this section, we present a model for processes exhibiting long cycles. Our objective is to develop a parsimonious modeling approach that allows for cycles with periods spanning non-negligible fractions of observed data. More formally, the model should allow for the period as a fraction of the sample size n to converge to a non-zero constant as $n \to \infty$.

As discussed in the introduction and following Sargent (1987), in the class of autoregressive models, a cyclical behavior requires complex roots that come as conjugate pairs, which makes AR(2) model with serially uncorrelated errors a natural starting point. Later in the paper, we will extend the approach and allow for serially correlated errors. Hence, consider a process $\{y_t\}$ generated according to

$$(1 - \phi_1 L - \phi_2 L^2) y_t = u_t, \tag{2.1}$$

where L denotes the lag operator and $\{u_t\}$ is a mean-zero iid sequence with a finite variance. Let λ_1 and λ_2 denote the roots of the characteristic equation for the lag polynomial in (2.1):

$$z^2 - \phi_1 z - \phi_2 = 0.$$

When $|\lambda_1| < 1$ and $|\lambda_2| < 1$, $\{y_t\}$ has the following MA(∞) representation:

$$y_t = \frac{1}{\lambda_1 - \lambda_2} \sum_{j=0}^{\infty} \left(\lambda_1^{j+1} - \lambda_2^{j+1} \right) u_{t-j}.$$

Suppose the roots λ_1, λ_2 are complex, and consider their polar form representation:

$$\lambda_1, \lambda_2 = r e^{\pm i\theta}, \tag{2.2}$$

where r denotes the modulus, and θ is the argument of the complex roots.

Given the polar coordinate representation for the roots and the MA(∞) representation for the process, we can now write $\{y_t\}$ as

$$y_t = \sum_{j=0}^{\infty} r^j \frac{\sin(\theta(j+1))}{\sin(\theta)} u_{t-j}. \tag{2.3}$$

According to (2.3), the realized value of y_t is a weighted infinite sum of past realizations of the innovation sequence $\{u_t\}$. When the characteristic roots are complex, the weights or impulse responses are given by a damped sine wave: the impulse response of y_t to u_{t-j} is

$$w_j = r^j \frac{\sin(\theta(j+1))}{\sin(\theta)},$$

where the modulus r indicates the rate of decay[4] or the persistence of the sine wave, and the argument θ corresponds to the angular frequency and determines the period of the sine wave.

The stochastic process $\{y_t\}$ inherits its oscillatory behaviour precisely from this damped periodic sine weighting function. The closer r is to one, the more persistent is $\{y_t\}$, and the closer θ is to zero, the lower is the oscillating frequency and the longer the length of cycles in $\{y_t\}$.

[4] In the more common exponential decay representation, $r^j = e^{\ln(r)j}$. Restricting to processes with non-explosive roots, i.e. $r \leq 1$, we have $\ln(r) \leq 0$ and $-\ln(r)$ is known as the decay constant.

Stochastic cycles are therefore conveniently captured in an AR(2) model with a pair of complex conjugate roots.

A cyclical process generated according to (2.1) with roots given by (2.2) has an expected cycle length of $2\pi/\theta$. With any fixed parameter value θ, the period as a fraction of the sample size becomes negligible for large n. Hence, asymptotic approximations assuming fixed values for the argument θ can produce distinctly different cyclical behavior from that observed in finite samples.[5] In other words, conventional asymptotics with a fixed complex root argument θ can distort the cyclical properties of the process. As a result, such asymptotic theory would provide a poor approximation to the actual behavior of the process in finite samples. Since the expected period of a process as a fraction of the sample size is given by $2\pi/(\theta n)$, to preserve the cyclical properties in the limit as $n \to \infty$, one has to consider a drifting sequence of the arguments $\{\theta_n\}$ and allow for $n\theta_n \to d \in [0, \infty]$.[6]

We re-write the AR(2) model in (2.1) as follows:[7]

$$(1 - \phi_{1,n}L - \phi_{2,n}L^2)y_t = u_t, \tag{2.4}$$

where $\phi_{1,n}$ and $\phi_{2,n}$ are now drifting coefficients that can change with n. We denote the corresponding characteristic roots as $\lambda_{1,n}$ and $\lambda_{2,n}$, and make the following assumption.

Assumption 2.1. The characteristic roots $\lambda_{1,n}, \lambda_{2,n}$ associated with the lag polynomial equation in (2.4) are given by

$$\lambda_{1,n} = e^{(c+id)/n}, \qquad \lambda_{2,n} = e^{(c-id)/n} \tag{2.5}$$

where $i = \sqrt{-1}$ is the imaginary number, and $c \leq 0$ and $d > 0$ are fixed localization parameters.

In the above assumption, we exclude positive values of c as they correspond to explosive roots. The negative values of d can be excluded as d and $-d$ define the same pair of roots. Note that the autoregressive coefficients are related to the characteristic roots through the following equations:

$$\phi_{1,n} = \lambda_{1,n} + \lambda_{2,n} = 2e^{c/n}\cos(d/n), \tag{2.6}$$

$$\phi_{2,n} = -\lambda_{1,n}\lambda_{2,n} = -e^{2c/n}. \tag{2.7}$$

Hence, the autoregressive coefficient $\phi_{1,n}$ is local to 2 while $\phi_{2,n}$ is local to -1. The sum of the autoregressive coefficients is local to one, and the process can be mistaken for those considered in the local-to-unity literature. As we discuss below, processes defined by (2.6)–(2.7) exhibit persistent stochastic oscillations and, as a result, their asymptotic properties are different from those considered in the local-to-unity literature.

The expressions for the characteristic roots in (2.2) and (2.5) are equivalent with r replaced by $r_n \equiv \exp(c/n)$, and θ replaced by $\theta_n \equiv d/n$. The modulus r_n in (2.5) has the same representation as the autoregressive parameter in the local-to-unity model in Phillips (1987, 1988). Therefore,

[5] This point is illustrated in Figure 1.

[6] The approach can still accommodate conventional asymptotics by allowing $n\theta_n \to \infty$.

[7] As in Phillips (1987, 1988), the solution to the difference equation (2.4) is a triangular array of the form $\{y_{n,t} : t = 1, \ldots, n; n \geq 1\}$. However, we suppress the subscript n to simplify the notation.

close to zero values of c correspond to persistent processes with two roots near unity. Hence, the process defined by Assumption 2.1 can be viewed as near I(2).

The parameter d controls the length of the cycle, where long cycles correspond to values of d near zero. Under Assumption 2.1, the expected period as a fraction of the sample size is given by

$$\tau_\theta \equiv \frac{2\pi}{n\theta_n} = \frac{2\pi}{d}. \tag{2.8}$$

With this parametrization, the length of the cycle as a fraction of the sample size is independent of the sample size, and resulting asymptotic approximations preserve the cyclical properties of the process in the limit.

An alternative but related measure of the periodicity of a process can be constructed from its spectral properties. The advantage of this measure is that it takes into account the persistence of the process unlike that based solely on the argument θ_n of the complex roots. Let ω_n^* denote the frequency that maximizes the spectral density of the process in (2.4). As in Sargent (1987),

$$\omega_n^* = \cos^{-1}\left(-\frac{\phi_{1,n}(1-\phi_{2,n})}{4\phi_{2,n}}\right),$$

and the corresponding period of the process as a fraction of the sample size is given by

$$\tau_{\omega_n^*} \equiv \frac{2\pi/\omega_n^*}{n}.$$

The proposition below provides an asymptotic approximation to the length of a cycle as a fraction of the sample size when measured using the spectrum-based approach.

Proposition 2.1. *Suppose that $\{y_{n,t}\}$ is generated according to (2.4) with characteristic roots satisfying Assumption 2.1 and serially uncorrelated $\{u_t\}$ with a finite variance. Suppose further that $d \geq |c|$. Then its spectrum maximizing frequency ω_n^* satisfies*

$$n\omega_n^* = \sqrt{d^2 - c^2} + O(n^{-2}),$$

and its corresponding spectrum-based period as a fraction of the sample size satisfies

$$\tau_{\omega_n^*} = \frac{2\pi}{\sqrt{d^2 - c^2}} + O(n^{-2}).$$

The proposition shows that, when using spectrum-based measures of the period, the length of a cycle relatively to the sample size can be approximated by

$$\tau_\omega \equiv \frac{2\pi}{\sqrt{d^2 - c^2}}. \tag{2.9}$$

Unlike the angular frequency-based measure τ_θ, the spectrum-based measure τ_ω takes into account the persistence of the process as captured by the value of the localization parameter c. Note that larger negative values of $c \leq 0$ produce less persistent processes. In such cases, the spectrum's peak is closer to the origin, and as a result, less persistent processes may not exhibit any visible cyclical behavior.

In Appendix A, we show that the periodogram-based estimation approach produces biased estimates of τ_ω. We, therefore, develop below an inference procedure for τ_θ and τ_ω based on the estimates of the autoregressive coefficients $\phi_{1,n}$ and $\phi_{2,n}$. For that purpose we proceed in two steps. First, we develop a procedure for constructing asymptotically valid confidence sets for the autoregressive parameters $\phi_{1,n}$ and $\phi_{2,n}$. In the second step, we use projection arguments to build confidence intervals for the proposed measures of the length of a cycle τ_θ and τ_ω. The theory developed below relaxes the iid/serially uncorrelated assumptions on $\{u_t\}$.

3. Asymptotics for long-cycle processes

We now provide the asymptotic theory for the process defined in equations (2.4)–(2.5). The theory will be subsequently used for establishing the asymptotic distributions of regression-based statistics involving long-cycle time series. It is also required for developing robust and asymptotically valid inference about the cyclical properties of a process.

The specification proposed in equations (2.4)–(2.5) is akin to the first-order autoregressive local-to-unity root model in Phillips (1987). Assuming that a process $\{x_t\}$ is generated according to $x_t = a_n x_{t-1} + u_t$ with $a_n = \exp(c/n)$, Phillips (1987) shows that the distribution of $\{x_t\}$ can be approximated by an Ornstein-Uhlenbeck diffusion process:

$$n^{-1/2}x_{\lfloor nr\rfloor} = n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} e^{c(\lfloor nr\rfloor - t)/n}u_t \Rightarrow \sigma J_c(r), \tag{3.1}$$

where

$$J_c(r) \equiv \int_0^r e^{c(r-s)}\mathrm{d}W(s),$$

$r \in [0,1]$, $\lfloor x\rfloor$ denotes the largest integer less or equal to x, $W(\cdot)$ is a standard Brownian motion, σ^2 denotes the limit of the long-run variance of $\{u_t\}$, "$\Rightarrow$" denotes the weak convergence of probability measures, and it is assumed that $\{u_t\}$ satisfies a Functional Central Limit Theorem (FCLT). Note that the distribution of the Ornstein-Uhlenbeck process $J_c(r)$ depends on the localization parameter c. In what follows, we build on these insights.

We make the following assumption on the innovation sequence $\{u_t\}$.[8]

Assumption 3.1 (FCLT)**.** Let $W(\cdot)$ denote the standard Brownian motion, and let σ^2 be the limit of the long-run variance of $\{u_t\}$: $\sigma^2 \equiv \lim_{n\to\infty} Var(n^{-1/2}\sum_{t=1}^{n} u_t)$. Then for $r \in [0,1]$,

$$n^{-1/2}\sum_{t=1}^{\lfloor nr\rfloor} u_t \Rightarrow \sigma W(r).$$

In the case of long-cycles, a different limiting process arises from that of the local-to-unity case, with cyclicality reflected by the sine function. However, the process can be also described as an integral with respect to a Brownian motion, and it also depends on the localization parameters c and d. We define:

$$J_{c,d}(r) \equiv \frac{1}{d}\int_0^r e^{c(r-s)}\sin(d(r-s))\mathrm{d}W(s). \tag{3.2}$$

[8] Assumption 3.1 holds, for example, when $\{u_t\}$ is a mixing process such that $E(u_t) = 0$ for all t, $\sup_t E|u_t|^{\beta+\epsilon} \leq \infty$ for some $\beta < 2$ and $\epsilon > 0$, and $\{u_t\}$ is α-mixing of size $-\beta/(\beta-2)$ (Phillips, 1987). Alternatively, it holds when $\{u_t\}$ is a linear MA(∞) process satisfying the conditions in Phillips and Solo (1992, Theorem 3.15).

The next proposition shows that in large samples and after appropriate scaling, the distribution of a long-cycle process can be approximated by that of $J_{c,d}(\cdot)$.

Proposition 3.1. *Suppose that $\{y_t\}$ is generated according to equation (2.4), and Assumptions 2.1 and 3.1 hold. Then,*

$$n^{-3/2} y_{\lfloor nr \rfloor} \Rightarrow \sigma J_{c,d}(r).$$

Proof. The solution to (2.4) can be expressed in terms of the characteristic roots as

$$\begin{aligned} y_{n,t} &= \frac{1}{\lambda_{n,1}-\lambda_{n,2}} \sum_{k=1}^{t} \left(\lambda_{n,1}^{t-k+1} - \lambda_{n,2}^{t-k+1}\right) u_k \\ &= \frac{1}{2i \cdot e^{c/n} \sin(d/n)} \sum_{k=1}^{t} \left(e^{(c+id)(t-k+1)/n} - e^{(c-id)(t-k+1)/n}\right) u_k, \end{aligned}$$

where the second equality follows by Assumption 2.1. By Assumption 3.1 and as in (3.1),

$$\begin{aligned} & n^{-1/2} \sum_{k=1}^{\lfloor nr \rfloor} (e^{(c+id)(t-k+1)/n} - e^{(c-id)(t-k+1)/n}) u_k \\ \Rightarrow \quad & \sigma \int_0^r \left(e^{(c+id)(r-s)} - e^{(c-id)(r-s)}\right) \mathrm{d}W(s) \\ = \quad & 2i\sigma \int_0^r e^{c(r-s)} \sin(d(r-s)) \mathrm{d}W(s). \end{aligned}$$

The result follows since $\sin(d/n) = d/n + O(n^{-2})$. $\square$

The continuous time Gaussian process $J_{c,d}(\cdot)$ plays the central role in our analysis. It can be viewed as a continuous time version of the MA(∞) representation in (2.3): past shocks are weighted by a damped sine wave. Again, the localization parameter c controls the persistence, and the localization parameter d controls the cyclicality. Note also that the long-cycle process requires stronger scaling than that in the local-to-unity case: $n^{-3/2}$ instead of $n^{-1/2}$. This is a reflection of the fact that long-cycle processes are near I(2).

We now turn to the properties of the least-squares estimators and the corresponding test statistics for the second-order autoregressive model with long cycles. Let $\widehat{\phi}_{1,n}$ and $\widehat{\phi}_{2,n}$ denote the least-squares estimator of (2.4):

$$\begin{pmatrix} \widehat{\phi}_{1,n} \\ \widehat{\phi}_{2,n} \end{pmatrix} = \begin{pmatrix} \sum y_{t-1}^2 & \sum y_{t-1}y_{t-2} \\ \sum y_{t-1}y_{t-2} & \sum y_{t-2}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{t-1}y_t \\ \sum y_{t-2}y_t \end{pmatrix}. \tag{3.3}$$

As it turns out, the matrix on the right-hand side is asymptotically singular because all three elements $\sum y_{t-1}^2$, $\sum y_{t-2}^2$, and $\sum y_{t-1}y_{t-2}$ converge to the same random limit when properly scaled. This is because $\sum y_{t-1}y_{t-2} = \sum y_{t-1}^2 +$ smaller order terms, which follows formally from Lemma 3.1(b) below. The singularity complicates the derivation of the limiting distributions of the estimators and the corresponding test statistics.

To eliminate the singularity arising in the limit, we consider the following transformation of the equation in (2.4):

$$y_t = (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + u_t, \tag{3.4}$$

where $\Delta y_{t-1} = y_{t-1} - y_{t-2}$. Since (3.4) is obtained from the original equation through a non-singular linear transformation of the regressors and parameters, the OLS estimator of $\phi_{1,n} + \phi_{2,n}$ is given by $\widehat{\phi}_{1,n} + \widehat{\phi}_{2,n}$. Moreover, the usual Wald test statistic for testing joint hypotheses about ϕ_1 and ϕ_2 is the same for both regressions. Thus, we have:

$$\begin{aligned}&\begin{pmatrix} \widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n} \\ \widehat{\phi}_{2,n} - \phi_{2,n} \end{pmatrix} = \\ &\qquad \begin{pmatrix} \sum y_{t-1}^2 & -\sum y_{t-1}\Delta y_{t-1} \\ -\sum y_{t-1}\Delta y_{t-1} & \sum(\Delta y_{t-1})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{t-1}u_t \\ -\sum \Delta y_{t-1}u_t \end{pmatrix}. \end{aligned} \tag{3.5}$$

As we show below, the matrix on the right-hand side of (3.5) is no longer singular in the limit.

It follows from the representation in (3.5) that the asymptotic theory of the OLS estimator involves the sample moments of $(y_{t-1}, \Delta y_{t-1})$. Hence, in addition to the asymptotic approximation of y_{t-1}, we also need the asymptotic approximation for Δy_{t-1}. The latter involves two additional continuous time processes. We define:

$$\begin{aligned} K_{c,d}(r) &\equiv \frac{1}{d}\int_0^r e^{c(r-s)}\cos(d(r-s))\mathrm{d}W(s), \\ G_{c,d}(r) &\equiv c \cdot J_{c,d}(r) + d \cdot K_{c,d}(r). \end{aligned} \tag{3.6}$$

Note that the diffusion process $K_{c,d}(r)$ is akin to the process $J_{c,d}(r)$ except that it is defined with a cosine function instead of a sine function. The next proposition shows that in large samples and after scaling, the distribution of $\Delta y_{\lfloor nr \rfloor}$ can be approximated by that of $G_{c,d}(r)$.

Proposition 3.2. *Suppose that $\{y_t\}$ is generated according to equation (2.4), and Assumptions 2.1 and 3.1 hold. Then,*

$$n^{-1/2}\Delta y_{\lfloor nr \rfloor} \Rightarrow \sigma G_{c,d}(r),$$

where the result holds jointly with that in Proposition 3.1.

Note that in contrast to the $n^{-3/2}$ scaling applied to y_{t-1}, its first difference Δy_{t-1} requires scaling by $n^{-1/2}$. Hence, the first differences of long-cycle processes have convergence rates of $O(n^{1/2})$ tantamount to that of local-to-unity processes. However due to cyclicality, the large-sample distribution of Δy_{t-1} is different from that arising in the local-to-unity model.

Based on the results of Proposition 3.1 and 3.2, we can now provide the asymptotic theory for the sample moments of long-cycle processes. Parts of the lemma below require the following ergodicity property for $\{u_t\}$.

Assumption 3.2. Let $\sigma_u^2 \equiv \lim_{n\to\infty} n^{-1}\sum_{t=1}^n Eu_t^2$ be the average variance of $\{u_t\}$ over time. We assume that $n^{-1}\sum_{t=1}^n u_t^2 \to_p \sigma_u^2$.

Lemma 3.1. *Suppose that $\{y_t\}$ is generated according to equation (2.4), and Assumptions 2.1 and 3.1 hold. The following results hold jointly.*

(a) $n^{-4}\sum y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 J_{c,d}^2(r)\mathrm{d}r$.
(b) $n^{-3}\sum y_{t-1}\Delta y_{t-1} \Rightarrow \sigma^2 \int_0^1 J_{c,d}(r)G_{c,d}(r)\mathrm{d}r$.
(c) $n^{-2}\sum(\Delta y_{t-1})^2 \Rightarrow \sigma^2 \int_0^1 G_{c,d}^2(r)\mathrm{d}r$.

Suppose in addition that Assumption 3.2 holds. The following results hold jointly with (a)–(c).

(d) $n^{-2}\sum y_{t-1}u_t \Rightarrow \sigma^2 \int_0^1 J_{c,d}(r)\mathrm{d}W(r)$.
(e) $n^{-1}\sum \Delta y_{t-1}u_t \Rightarrow \sigma^2 \int_0^1 G_{c,d}(r)\mathrm{d}W(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2)$.

Note that in part (e) of the lemma, the limiting distribution of the sample covariance between Δy_{t-1} and u_t depends on the difference between the long-run and the average over time variances of $\{u_t\}$. This reflects the serial correlation in $\{u_t\}$ and is standard in the unit root literature. However despite the serial correlation, the difference $\sigma^2 - \sigma_u^2$ does not appear in the limiting expressions in part (d) for the sample covariance between y_{t-1} and u_t. This is because of the stronger scaling factor required for the near I(2) long-cycle process $\{y_t\}$.

To simplify the notation, in the rest of the paper we use $\int J_{c,d}^2$ to denote $\int_0^1 J_{c,d}^2(r)\mathrm{d}r$ and $\int J_{c,d}\mathrm{d}W$ to denote $\int_0^1 J_{c,d}(r)\mathrm{d}W(r)$. We use the same convention for the integral expressions with $G_{c,d}(r)$ with $J_{c,d}$ replaced by $G_{c,d}$. Lastly, we use $\int J_{c,d}G_{c,d}$ to denote $\int_0^1 J_{c,d}(r)G_{c,d}(r)\mathrm{d}r$.

Equipped with the results of Lemma 3.1, we can now describe the asymptotic distribution of the least-squares estimators of $\phi_{1,n}$ and $\phi_{2,n}$.

Proposition 3.3. *Suppose that $\{y_{n,t}\}$ is generated according to equation (2.4), and Assumptions 2.1, 3.1, and 3.2 hold. The following results hold jointly with the results of Lemma 3.1.*

(a) $n^2(\widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}) \Rightarrow$

$$\frac{\int G_{c,d}^2 \cdot \int J_{c,d}\mathrm{d}W - \left(\int G_{c,d}\mathrm{d}W + \frac{1}{2}(1-\sigma_u^2/\sigma^2)\right)\cdot \int J_{c,d}G_{c,d}}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - (\int J_{c,d}G_{c,d})^2}.$$

(b) $n\begin{pmatrix}\widehat{\phi}_{1,n} - \phi_{1,n} \\ \widehat{\phi}_{2,n} - \phi_{2,n}\end{pmatrix} \Rightarrow$

$$\begin{pmatrix}-1 \\ 1\end{pmatrix} \times \frac{\int J_{c,d}G_{c,d} \cdot \int J_{c,d}\mathrm{d}W - \left(\int G_{c,d}\mathrm{d}W + \frac{1}{2}(1-\sigma_u^2/\sigma^2)\right)\cdot \int J_{c,d}^2}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - \left(\int J_{c,d}G_{c,d}\right)^2}.$$

According to part (b) of the proposition, the joint asymptotic distribution of the least-squares estimators of $\phi_{1,n}$ and $\phi_{2,n}$ is singular and determined by the same random variable. Moreover, their rate of convergence is $O_p(n^{-1})$ despite the process $\{y_t\}$ being near I(2). This is a consequence of the asymptotic singularity in (3.3) as previously discussed on page 10. However, in part (a) of the proposition, the least-squares estimator of $\phi_{1,n} + \phi_{2,n}$ has the faster convergence rate $O_p(n^{-2})$ characteristic to I(2) processes. Note that the limiting distributions depend on the localization parameters c and d.

Next, we discuss inference for the autoregressive coefficients. Consider testing a joint hypothesis $H_0 : \phi_1 = \phi_{2,0}, \phi_2 = \phi_{2,0}$ against $H_1 : \phi_1 \neq \phi_{2,0}$ or $\phi_2 \neq \phi_{2,0}$. The usual Wald statistic is given by

$$W_n(\phi_{1,0},\phi_{2,0}) \equiv \begin{pmatrix} \widehat{\phi}_{1,n}+\widehat{\phi}_{2,n}-\phi_{1,0}-\phi_{2,0} \\ \widehat{\phi}_{2,n}-\phi_{2,0} \end{pmatrix}^{\top} \widehat{V}_n^{-1} \begin{pmatrix} \widehat{\phi}_{1,n}+\widehat{\phi}_{2,n}-\phi_{1,0}-\phi_{2,0} \\ \widehat{\phi}_{2,n}-\phi_{2,0} \end{pmatrix}, \qquad (3.7)$$

where

$$\widehat{V}_n \equiv \widehat{\sigma}_n^2 \begin{pmatrix} \sum y_{t-1}^2 & -\sum y_{t-1}\Delta y_{t-1} \\ -\sum y_{t-1}\Delta y_{t-1} & \sum(\Delta y_{t-1})^2 \end{pmatrix}^{-1},$$

and $\widehat{\sigma}_n^2$ is a consistent estimator of the long-run variance σ^2 constructed using $\hat{u}_t = y_t - \widehat{\phi}_{1,n}y_{t-1} - \widehat{\phi}_{2,n}y_{t-2}$, see Newey and West (1987) and Andrews (1991). The infeasible estimator of σ^2 that uses u_t is constructed as

$$\tilde{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n} u_t^2 + 2\sum_{h=1}^{m_n} w_n(h)n^{-1}\sum_{t=h+1}^{n} u_t u_{t-h},$$

where $m_n = o(n)$ is the lag truncation parameter, and $w_n(\cdot)$ is a bounded weight function such that $\lim_{n\to\infty} w_n(h) = 1$ for all h. The feasible estimator $\widehat{\sigma}_n^2$ is constructed similarly using the estimated residuals $\hat{u}_t$ in place of u_t. We make the following assumption.

Assumption 3.3. The infeasible estimator $\tilde{\sigma}_n^2$ of the long-run variance σ^2 is consistent: $\tilde{\sigma}_n^2 \to_p \sigma^2$.

The conditions for consistency of the infeasible estimator can be found in Newey and West (1987) and Andrews (1991). Our next result describes the asymptotic null distribution of the Wald statistic for long-cycle processes.

Proposition 3.4. *Suppose that $\{y_{n,t}\}$ is generated according to equation (2.4), Assumptions 2.1 and 3.1-3.3 hold, and $m_n = o(n)$. Then,*

$$W_n(\phi_{1,n},\phi_{2,n}) \Rightarrow \frac{\int \left\{ J_{c,d}\cdot\left(\int G_{c,d}\mathrm{d}W + \frac{1}{2}(1-\sigma_u^2/\sigma^2)\right) - G_{c,d}\cdot\int J_{c,d}\mathrm{d}W \right\}^2}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - \left(\int J_{c,d}G_{c,d}\right)^2}.$$

The asymptotic null distribution of the Wald statistic is non-standard and non-pivotal: it depends on the ratio of the average over time and long-run variances σ_u^2/σ^2, and on the unknown localization parameters c and d. While the ratio σ_u^2/σ^2 plays no role when $\{u_t\}$ are serially uncorrelated and can be estimated consistently otherwise,[9] the dependence on c and d remains. Hence, the quantiles of the limiting distribution can only be simulated given the values of c and d.

In Section 5, we discuss the differences between the usual χ_2^2 critical values and the quantiles of the asymptotic distribution in Proposition 3.4. Depending on the values of c and d, the differences can be substantial especially when the model includes deterministic components discussed in the next section.

[9] See the proof of Proposition 3.4.

4. Extensions to models with deterministic components

For practical applications, it is important to allow the DGP to include non-zero means, trends, and deterministic cycles. We discuss such adjustments in this section. As the results below show, the limiting distributions of the regression estimators and test statistics take a similar form to those in Section 3, but with $J_{c,d}$ and $G_{c,d}$ replaced with their residuals from appropriate continuous time projections. This property is standard in the unit root literature and continues to hold in our case.

Formally, we assume that the data $\{y_t : t = 1, \ldots, n\}$ are generated according to

$$(1 - \phi_{1,n}L - \phi_{2,n}L^2)(y_t - D_t) = u_t, \tag{4.1}$$

where D_t is non-random, can vary with t, and depends on unknown parameters. To control for the deterministic part, estimation of the autoregressive coefficients requires projecting against the components of D_t. The asymptotic distributions of the estimators and test statistics change accordingly. We consider the following three formulations of D_t:

(i) Constant mean: $D_t = \mu$ for some unknown parameter μ.

(ii) Deterministic cycles: $D_t = \mu + \sum_k \{\theta_{1k} \cos(2\pi kt/n) + \theta_{2k} \sin(2\pi kt/n)\}$, where k's are known positive integers, and θ_{1k}, θ_{2k} are unknown coefficients.

(iii) Linear time trend: $D_t = \mu + \xi t/n$, where ξ is the unknown coefficient.

The specification in (i) allows $\{y_t\}$ to have a constant over time non-zero mean. The DGP in (ii) can be used, for example, to distinguish between very low frequency fluctuations and long cycles, as many time series in economics exhibit such patterns, see Beaudry, Galizia, and Portier (2020).[10] The D_t component in (ii) generates cosine and sine oscillations at frequencies $2\pi k/n$. The period of such oscillations relatively to the sample size is $1/k$, and they can capture very low-frequency cycles in data that are outside the range of the econometrician's interest. For practical purposes, we consider $k = 1, 2, 3$. Inclusion of such components can be viewed as de-trending of data by removing fluctuations at the frequencies corresponding to the values of k. The asymptotic results developed in this section can be used to account for de-trending in inferential procedures.

The DGP in (iii) allows for linear time trends, and such adjustments have a long history in the unit root literature. The division by n is required for deriving the asymptotic properties and can be absorbed into the unknown coefficient ξ. Hence, observationally the model in (iii) is identical to the model with no adjustment by n.

The empirical application in Section 7 also considers the case where D_t consists of seasonal dummies and a constant. However, as shown in Phillips and Jin (2002) for unit root testing, the arising asymptotic distributions have the same form as those in the constant mean case.

As in the previous section and to avoid singularities in the limit, we use the transformed version of the model with y_{t-1} and Δy_{t-1}:

$$y_t = (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + (1 - \phi_{1,n}L - \phi_{2,n}L^2)D_t + u_t. \tag{4.2}$$

[10] We thank Paul Beaudry for pointing our attention to this fact.

4.1. Constant mean

In this section, we consider case (i) of a constant unknown mean. When $D_t = \mu$, equation (4.2) becomes

$$y_t = \alpha_n + (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + u_t, \tag{4.3}$$

where $\alpha_n \equiv (1 - \phi_{1,n} - \phi_{2,n})\mu = O(n^{-2})$.[11] Let $\widehat{\phi}_{1,n}$ and $\widehat{\phi}_{2,n}$ be the least-squares estimator of the corresponding coefficients in (4.3), and define $\widetilde{y}_{t-1} = y_{t-1} - \bar{y}$ and $\widetilde{\Delta y}_{t-1} = \Delta y_{t-1} - \overline{\Delta y}$, where $\bar{y}_n$ and $\overline{\Delta y}_n$ denote the sample averages of y_{t-1} and Δy_{t-1} respectively. Then,

$$\begin{pmatrix} \widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n} \\ \widehat{\phi}_{2,n} - \phi_{2,n} \end{pmatrix} = \begin{pmatrix} \sum \widetilde{y}_{t-1}^2 & -\sum \widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1} \\ -\sum \widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1} & \sum \widetilde{\Delta y}_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \widetilde{y}_{t-1}u_t \\ -\sum \widetilde{\Delta y}_{t-1}u_t \end{pmatrix}.$$

and we have the following analogue of Lemma 3.1.

Lemma 4.1. *Suppose that $\{y_t\}$ is generated according to equation (4.3), and Assumptions 2.1 and 3.1 hold. Define $\widetilde{J}_{c,d}(r) \equiv J_{c,d}(r) - \int_0^1 J_{c,d}(s)\mathrm{d}s$ and $\widetilde{G}_{c,d}(r) \equiv G_{c,d}(r) - \int_0^1 G_{c,d}(s)\mathrm{d}s$. The following results hold jointly.*

- *(a)* $n^{-4}\sum \widetilde{y}_{t-1}^2 \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}^2$.
- *(b)* $n^{-2}\sum \widetilde{\Delta y}_{t-1}^2 \Rightarrow \sigma^2 \int \widetilde{G}_{c,d}^2$.
- *(c)* $n^{-3}\sum \widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1} \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}\widetilde{G}_{c,d}$.

Suppose in addition that Assumption 3.2 holds. The following results hold jointly with (a)–(c).

- *(d)* $n^{-2}\sum \widetilde{y}_{t-1}u_t \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}\mathrm{d}W$.
- *(e)* $n^{-1}\sum \widetilde{\Delta y}_{t-1}u_t \Rightarrow \sigma^2 \int \widetilde{G}_{c,d}\mathrm{d}W + \frac{1}{2}(\sigma^2 - \sigma_u^2)$.

The results in Lemma 4.1 are parallel to those in Lemma 3.1. However, instead of $J_{c,d}$ and $G_{c,d}$, the distributions arising in the limit depend on $\widetilde{J}_{c,d}$ and $\widetilde{G}_{c,d}$. Note that the latter processes are obtained from $J_{c,d}$ and $G_{c,d}$ by subtracting their respective continuous time averages, which matches the construction of $\widetilde{y}_t$ and $\widetilde{\Delta y}_t$ in finite samples.

4.2. Deterministic cycles

In this section, we consider case (ii) of deterministic cycles. When D_t includes deterministic cycles, equation (4.2) takes the form

$$\begin{aligned} y_t = \alpha_n + \sum_k \left\{\gamma_{1k,n}\cos\left(\frac{2\pi kt}{n}\right) + \gamma_{2k,n}\sin\left(\frac{2\pi kt}{n}\right)\right\} \\ + (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + u_t, \end{aligned} \tag{4.4}$$

where the intercept α_n is as defined in the case of a constant mean. The lags of the cosine and sine components can be written as linear combinations of $\cos(2\pi kt/n)$ and $\sin(2\pi kt/n)$ with coefficients depending on n and, therefore, can be omitted. The least-squares estimators $\widehat{\phi}_{1,n}$ and

[11] See Lemma C.1 in the Appendix.

$\widehat{\phi}_{2,n}$ can be obtained by estimating

$$y_t = (\phi_{1,n} + \phi_{2,n})\widetilde{y}_{t-1} - \phi_{2,n}\Delta\widetilde{y}_{t-1} + u_t,$$

where $\widetilde{y}_{t-1}$ and $\widetilde{\Delta y}_{t-1}$ are the residuals from the regressions of y_{t-1} and Δy_{t-1} respectively on $\cos(2\pi kt/n)$, $\sin(2\pi kt/n)$, and a constant.

The following result describes the asymptotic distributions of the sample moments of $\widetilde{y}_{t-1}$, $\widetilde{\Delta y}_{t-1}$, and u_t.

Lemma 4.2. *Suppose that $\{y_t\}$ is generated according to equation (4.4), and Assumptions 2.1 and 3.1 hold. Define*

$$\widetilde{J}_{c,d}(r) \equiv J_{c,d}(r) - \int_0^1 J_{c,d}(s)\mathrm{d}s - \sum_k \left\{\psi_{1k}\cos(2\pi kr) - \psi_{2k}\sin(2\pi kr)\right\},$$

$$\widetilde{G}_{c,d}(r) \equiv G_{c,d}(r) - \int_0^1 G_{c,d}(s)\mathrm{d}s - \sum_k \left\{\varphi_{1k}\cos(2\pi kr) - \varphi_{2k}\sin(2\pi kr)\right\},$$

where

$$\psi_{1k} \equiv 2\int_0^1 \cos(2\pi ks)J_{c,d}(s)\mathrm{d}s, \quad \psi_{2k} \equiv 2\int_0^1 \sin(2\pi ks)J_{c,d}(s)\mathrm{d}s,$$

$$\varphi_{1k} \equiv 2\int_0^1 \cos(2\pi ks)G_{c,d}(s)\mathrm{d}s, \quad \varphi_{2k} \equiv 2\int_0^1 \sin(2\pi ks)G_{c,d}(s)\mathrm{d}s.$$

The following results hold jointly.

(a) $n^{-4}\sum\widetilde{y}_{t-1}^2 \Rightarrow \sigma^2\int\widetilde{J}_{c,d}^2$.

(b) $n^{-2}\sum\Delta\widetilde{y}_{t-1}^2 \Rightarrow \sigma^2\int\widetilde{G}_{c,d}^2$.

(c) $n^{-3}\sum\widetilde{y}_{t-1}\Delta\widetilde{y}_{t-1} \Rightarrow \sigma^2\int\widetilde{J}_{c,d}\widetilde{G}_{c,d}$.

Suppose in addition that Assumption 3.2 holds. The following results hold jointly with (a)-(c).

(d) $n^{-2}\sum\widetilde{y}_{t-1}u_t \Rightarrow \sigma^2\int\widetilde{J}_{c,d}\mathrm{d}W$.

(e) $n^{-1}\sum\Delta\widetilde{y}_{t-1}u_t \Rightarrow \sigma^2\int\widetilde{G}_{c,d}\mathrm{d}W + \frac{1}{2}(\sigma^2 - \sigma_u^2)$.

Lemma 4.2 is the analogue of Lemma 3.1 for the model with deterministic cycles. The coefficients $\psi_{1,k}$ and $\psi_{2,k}$ can be viewed as the least-squares coefficients in the continuous time regression of $J_{c,d}(s)$ against $\cos(2\pi ks)$, $\sin(2\pi ks)$, and a constant with s varying over the $[0,1]$ interval. The coefficients φ_{1k} and φ_{2k} have a similar interpretation with $J_{c,d}$ replaced by $G_{c,d}$. The processes $\widetilde{J}_{c,d}$ and $\widetilde{G}_{c,d}$ are therefore the residuals from the corresponding continuous time regressions. They are the continuous time versions of $\widetilde{y}_{t-1}$ and $\widetilde{\Delta y}_{t-1}$ respectively. Hence, the results of Lemma 3.1 continue to hold with the processes $J_{c,d}$ and $G_{c,d}$ replaced by their respective residuals from the continuous time regressions.

4.3. Linear time trend

In this section, we consider case (iii) of a linear time trend. The model in equation (4.2) now takes the form

$$y_t = \delta_n + \beta_n(t/n) + (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + u_t \tag{4.5}$$

where $\delta_n \equiv \alpha_n + (\phi_{1,n} + 2\phi_{2,n})\xi/n = O(n^{-2})$, and $\beta_n \equiv \xi(1 - \phi_{1,n} - \phi_{2,n}) = O(n^{-2})$.[12] Similarly to the previous cases, the least-squares estimators $\widehat{\phi}_{1,n}$ and $\widehat{\phi}_{2,n}$ can be obtained by estimating

$$\widetilde{y}_t = (\phi_{1,n} + \phi_{2,n})\widetilde{y}_{t-1} - \phi_{2,n}\widetilde{\Delta y}_{t-1} + \widetilde{u}_t$$

where $\widetilde{y}_{t-1}$ and $\widetilde{\Delta y}_{t-1}$ are now the residuals from the regressions of y_{t-1} and Δy_{t-1} respectively against t/n and a constant.

Lemma 4.3. *Suppose that $\{y_t\}$ is generated according to equation (4.5), and Assumptions 2.1 and 3.1 hold. Define*

$$\widetilde{J}_{c,d}(r) \equiv J_{c,d}(r) - (4 - 6r)\int_0^1 J_{c,d}(s)\mathrm{d}s - (12r - 6)\int_0^1 sJ_{c,d}(s)\mathrm{d}s,$$
$$\widetilde{G}_{c,d}(r) \equiv G_{c,d}(r) - (4 - 6r)\int_0^1 G_{c,d}(s)\mathrm{d}s - (12r - 6)\int_0^1 sG_{c,d}(s)\mathrm{d}s.$$

The following results hold jointly.

(a) $n^{-4}\sum \widetilde{y}_{t-1}^2 \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}^2$.

(b) $n^{-2}\sum \widetilde{\Delta y}_{t-1}^2 \Rightarrow \sigma^2 \int \widetilde{G}_{c,d}^2$.

(c) $n^{-3}\sum \widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1} \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}\widetilde{G}_{c,d}$.

Suppose in addition that Assumption 3.2 holds. The following results hold jointly with (a)-(c).

(d) $n^{-2}\sum \widetilde{y}_{t-1}u_t \Rightarrow \sigma^2 \int \widetilde{J}_{c,d}\mathrm{d}W$.

(e) $n^{-1}\sum \widetilde{\Delta y}_{t-1}u_t \Rightarrow \sigma^2 \int \widetilde{G}_{c,d}\mathrm{d}W + \frac{1}{2}(\sigma^2 - \sigma_u^2)$.

Lemma 4.3 is the analogue of Lemmas 4.1 and 4.2 for the case of the linear time trend. The processes $\widetilde{J}_{c,d}(r)$ and $\widetilde{G}_{c,d}(r)$ can be similarly interpreted as the residuals from the continuous time regressions of $J_{c,d}(r)$ and $G_{c,d}(r)$ respectively against a constant and r varying over the interval $[0, 1]$.

4.4. Asymptotic distributions of the estimators and test statistics

The results of Lemmas 4.1–4.3 can now be used to describe the asymptotic distributions of the least-squares estimators of the autoregressive coefficients and the corresponding Wald statistics for the models with a constant mean, deterministic cycles, and a linear time trend respectively.

Under the same assumptions as those in Proposition 3.3, however with the model in equation (2.4) replaced by that in either (4.3), (4.4), or (4.5), the asymptotic distribution of the least-squares estimators of the autoregressive coefficients now satisfies

$$n^2(\widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}) \Rightarrow$$
$$\frac{\int \widetilde{G}_{c,d}^2 \cdot \int \widetilde{J}_{c,d}\mathrm{d}W - \left(\int \widetilde{G}_{c,d}\mathrm{d}W + \frac{1}{2}(1 - \sigma_u^2/\sigma^2)\right) \cdot \int \widetilde{J}_{c,d}\widetilde{G}_{c,d}}{\int \widetilde{J}_{c,d}^2 \cdot \int \widetilde{G}_{c,d}^2 - (\int \widetilde{J}_{c,d}\widetilde{G}_{c,d})^2},$$

[12] See Lemma C.1 in the Appendix.

$$n\begin{pmatrix}\widehat{\phi}_{1,n}-\phi_{1,n}\\ \widehat{\phi}_{2,n}-\phi_{2,n}\end{pmatrix}\Rightarrow$$

$$\begin{pmatrix}-1\\ 1\end{pmatrix}\times\frac{\int\widetilde{J}_{c,d}\widetilde{G}_{c,d}\cdot\int\widetilde{J}_{c,d}\mathrm{d}W-\left(\int\widetilde{G}_{c,d}\mathrm{d}W+\frac{1}{2}(1-\sigma_u^2/\sigma^2)\right)\cdot\int\widetilde{J}_{c,d}^2}{\int\widetilde{J}_{c,d}^2\cdot\int\widetilde{G}_{c,d}^2-\left(\int\widetilde{J}_{c,d}\widetilde{G}_{c,d}\right)^2},\tag{4.6}$$

where the convergence holds jointly with the results of either Lemma 4.1, 4.2, or 4.3 respectively with the correspondingly defined residual processes $\widetilde{J}_{c,d}^2$ and $\widetilde{G}_{c,d}^2$.

For all three specifications in Sections 4.1–4.3, the Wald statistic for testing $H_0:\phi_1=\phi_{2,0},\phi_2=\phi_{2,0}$ against $H_1:\phi_1\neq\phi_{2,0}$ or $\phi_2\neq\phi_{2,0}$ takes the same form as in equation (3.7). However, $\widehat{V}_n$ is now given by

$$\widehat{V}_n=\widehat{\sigma}_n^2\begin{pmatrix}\sum\widetilde{y}_{t-1}^2 & -\sum\widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1}\\ -\sum\widetilde{y}_{t-1}\widetilde{\Delta y}_{t-1} & \sum(\widetilde{\Delta y}_{t-1})^2\end{pmatrix}^{-1},$$

with $\widetilde{y}_{t-1}$ and $\widetilde{\Delta y}_{t-1}$ defined respectively for each specification. Provided that the assumptions of Proposition 3.4 hold with the model in (2.4) replaced by that in either (4.3), (4.4), or (4.5), the asymptotic null distribution of the Wald statistic is given by

$$W_n(\phi_{1,n},\phi_{2,n})\Rightarrow\frac{\int\left\{\widetilde{J}_{c,d}\cdot\left(\int\widetilde{G}_{c,d}\mathrm{d}W+\frac{1}{2}(1-\sigma_u^2/\sigma^2)\right)-\widetilde{G}_{c,d}\cdot\int\widetilde{J}_{c,d}\mathrm{d}W\right\}^2}{\int\widetilde{J}_{c,d}^2\cdot\int\widetilde{G}_{c,d}^2-\left(\int\widetilde{J}_{c,d}\widetilde{G}_{c,d}\right)^2}\tag{4.7}$$

with the correspondingly defined residual processes $\widetilde{J}_{c,d}^2$ and $\widetilde{G}_{c,d}^2$.

As in the base case with no deterministic components, the asymptotic null distributions of the Wald statistics are non-standard and depend on the unknown parameters c and d. The differences between the quantiles of these asymptotic distributions and the χ_2^2 critical values are discussed in Section 5. In comparison with the base case, inclusion of the deterministic components may result in more substantial deviations from the χ_2^2 critical values.

5. Size distortions of conventional tests

In this section, we discuss the size distortions one would see if the econometrician were to use conventional χ_2^2 critical values in place of the quantiles of the distributions derived in equation (4.7) in the previous section. For the purpose of this exercise, we assume that there is no serial correlation in $\{u_t\}$ and, as a result, non-centrality term $0.5(1-\sigma_u^2/\sigma^2)$ is equal to zero. Note that if $\{u_t\}$ is serially correlated, one can expect more substantial size distortions due to the presence of the non-centrality term in the asymptotic distribution.

Let $F_{c,d}(\cdot)$ denote the CDF of the asymptotic null distribution in (4.7). Note that the CDF depends on the unknown localization parameters c and d. Consider a test that rejects the null hypothesis when the Wald statistic exceeds the conventional $\chi_{2,1-\alpha}^2$ critical value, where $\chi_{2,1-\alpha}^2$ is the $1-\alpha$ quantile of the χ_2^2 distribution. The asymptotic size of this test is $1-F_{c,d}(\chi_{2,1-\alpha}^2)$, and size distortion are given by the difference between the asymptotic size and the nominal size

TABLE 1. Asymptotic size of the conventional Wald test with $\chi^2_{2,1-\alpha}$ critical values for $\alpha = 0.05$ for different values of the localization parameters c, d and different specifications of the deterministic component

	d					
c	5	15	25	35	45	55
			constant mean			
-1	.116	.059	.070	.054	.051	.050
-5	.110	.060	.052	.062	.054	.050
-10	.089	.064	.053	.049	.056	.051
-15	.077	.064	.055	.051	.049	.061
-20	.070	.062	.056	.051	.049	.049
-70	.051	.051	.050	.050	.049	.048
			deterministic cycle: $k = 1$			
-1	.746	.103	.071	.054	.051	.050
-5	.641	.161	.070	.063	.054	.050
-10	.441	.192	.090	.059	.056	.051
-15	.317	.186	.104	.069	.054	.059
-20	.242	.170	.109	.075	.058	.050
-130	.052	.052	.051	.051	.050	.048
			linear time trend			
-1	.317	.071	.072	.053	.051	.050
-5	.257	.089	.058	.062	.054	.051
-10	.188	.101	.066	.053	.056	.051
-15	.146	.101	.072	.057	.051	.060
-20	.121	.096	.074	.060	.052	.049
-100	.053	.052	.052	.052	.050	.049
τ_θ	1.26	0.42	0.25	0.18	0.14	0.11

$\tau_\theta = 2\pi/d$: cycle length as a fraction of the sample size.

α. We next examine the extent of size distortions for different values of c and d in the case of the three specifications for the deterministic component D_t in Section 4.

Table 1 reports the asymptotic size for $\alpha = 0.05$ and different values of c and d for each of the three specifications of D_t. The CDF $F_{c,d}(\cdot)$ is computed by Monte Carlo simulation with 100,000 replications and $J_{c,d}$ and $G_{c,d}$ processed generated using the Euler-Maruyama method with a time step $\Delta t = 0.01$. The table also reports the length of the cycle as a fraction of the sample size measured by $\tau_\theta = 2\pi/d$. The smaller is the value of d, the lower is the oscillation frequency, and the longer is the cycle length relative to the sample size.

In the case of all three specifications for D_t, the table shows similar patterns: the asymptotic size deviates from the nominal 0.05 values for values of c and d closer to zero. However, as c becomes more negative and d becomes more positive, the asymptotic size starts to approach the nominal value.

For example, in model with a constant mean, the asymptotic size at $c = -1$ and $d = 5$ is 0.116, which means that the Wald test based on the conventional $\chi^2_{2,1-\alpha}$ critical value over rejects the null by 0.066. As we move down the rows and across the columns of Table 1, the process becomes less persistent and with a shorter cycle period, and as a result size distortions become negligible. Note, however, that the relationship can be non-monotone.

While in the case of the constant mean model, the size distortions are relatively minor, they are much more prominent in the case of the models with deterministic cycles and linear trends. In particular, the usage of conventional χ^2_2 critical values may result in severe size distortions in the case of deterministic cycles. For example when $c = -1$ and $d = 5$, the null rejection probability is approximately 75% instead of 5%. It is approximately 32% in the case of the linear time trend specification. While $d = 5$ corresponds to very long cycles as measured by τ_θ, size distortions remain substantial even for shorter cycles. For example, in the model with deterministic cycles, the size of the conventional test is approximately 19% for $c = -10, d = 15$. These values correspond to $\tau_\theta = 0.42$ and $\tau_\omega = 0.56$.

Note again that the size distortions can be non-monotone across the rows/columns. However, for large negative values of c or large values of d, size distortions disappear. This is consistent with the results in Phillips (1987), who shows that in the local-to-unity model, the null distribution of the t-statistic for the autoregressive coefficient converges to the standard normal as $c \to -\infty$.

To conclude, depending on the values of c and d, the expression on the right-hand side of (4.7) can generate a wide range of different asymptotic distributions. The distributions can deviate substantially from the χ^2_2 for values of c, d sufficiently close to zeros. Such specifications correspond to longer cycles. For values of c, d sufficiently far from zero, which correspond to shorter cycles, the distributions converge to χ^2_2. In particular, across all specifications of the deterministic component, the size distortions from using χ^2_2 critical values become negligible for $\tau_\theta \leq 0.14$. However, when the length of the cycle as measured by τ_θ exceeds 14% of the sample size, conventional inference procedures can results in size distortions. The distortions are typically more pronounced for longer cycles.

6. Inference for cyclicality

In this section, we propose a procedure for inference on the cycle length in terms of the angular frequency-based measure τ_θ and the spectrum-based measure τ_ω that were introduced in Section 2. Recall that the two measures can be deduced from the autoregressive coefficients ϕ_{1n} and $\phi_{2,n}$ through the relationships in (2.6)–(2.9). Therefore, we first construct confidence sets for the autoregressive parameters by collecting values (ϕ_1, ϕ_2) consistent with cyclical behavior and not rejected by data. In the second step, we use projection arguments to construct confidence intervals for τ_θ and τ_ω. By multiplying the values of τ_θ and τ_ω in the confidence intervals by n, the length of the cycle can be also expressed in time units instead of fractions of the sample size.

The proposed confidence sets have the following property: If the true DGP is indeed cyclical, the coverage probability is at least $1 - \alpha$ asymptotically whether the roots of the autoregressive equation are close to one or far from it. However, if the true DGP is inconsistent with cyclical behavior, we expect the confidence sets to be empty in large samples. Hence, the proposed procedure can be used to detect cyclical specifications consistent with data. It can also be used to rule out cyclical behavior. However, our procedure is not designed for ruling out non-cyclical specifications.

When the roots of the autoregressive polynomial are local to unity as in Assumption 2.1, the least-squares estimators of the autoregressive coefficients are consistent regardless of whether $\{u_t\}$ is serially correlated or not. This is established in Proposition 3.3 for the base case and in (4.6) for the cases with deterministic components. Serial correlation in $\{u_t\}$ and the resulting correlation between $(y_{t-1}, \Delta y_{t-1})$ and u_t is reflected by the non-centrality term $0.5(1-\sigma_u^2/\sigma^2)$ in the asymptotic distributions. The non-centrality proliferates from the estimators into the asymptotic null distribution of the Wald statistic. This is standard for the unit root literature and continues to hold in our framework.

However, when the roots of the autoregressive polynomial are sufficiently far from unity, the least-squares estimators of the autoregressive coefficients are no longer consistent and suffer from first order bias. Therefore, to design an inferential procedure that remains valid regardless of the magnitude of the roots, we explicitly control for potential serial correlation in $\{u_t\}$.

We proceed as follows. First, in Section 6.1 we discuss how to construct confidence intervals for τ_θ and τ_ω when $\{u_t\}$ are serially uncorrelated. Then, in Section 6.2 we extend the procedure to a serially correlated innovation process $\{u_t\}$ by assuming that it satisfies an AR(p) formulation with real roots bounded away from one. We employ the BIC selection procedure to choose the appropriate number of lags p as well as the specification for the deterministic part D_t.

6.1. Serially uncorrelated $\{u_t\}$

Suppose that $\{u_t\}$ is serially uncorrelated and, therefore, the least-squares estimators of the autoregressive coefficients $\phi_{1,n}$ and $\phi_{2,n}$ are consistent whether the roots are close to unity or far from it. Recall that the expression on the right-hand side of (4.7) with $1-\sigma_u^2/\sigma^2 = 0$ approximates well the asymptotic distribution of the Wald statistic for any configuration of the localization parameters c and d. Moreover, recall that given the sample size n, there is a one-to-one relationship between $(\phi_{1,n}, \phi_{2,n})$ and the localization parameters (c, d), and let

$$\phi_{1,n} = \Phi_{1,n}(c,d) \text{ and } \phi_{2,n} = \Phi_{2,n}(c,d),$$

where the functions $\Phi_{1,n}(c,d)$ and $\Phi_{2,n}(c,d)$ are defined according to (2.6) and (2.7) respectively. Because the relationship is one-to-one for any given n, confidence sets for (ϕ_1, ϕ_2) can be equivalently represented as confidence sets in terms of (c, d).

By running the regressions of y_t against y_{t-1} and y_{t-2} with different specifications of the deterministic part D_t, one can use the Bayesian Information Criterion (BIC) based selection procedure to consistently choose the appropriate specification between the constant mean, deterministic cycle, and linear time trend. Once the specification for D_t has been selected, consider the corresponding Wald statistic $W_n(\Phi_{1,n}(c,d), \Phi_{2,n}(c,d))$. Let $\mathcal{W}_{1-\alpha}(c,d)$ denote the $1-\alpha$ quantiles of the asymptotic distribution in (4.7) with $1-\sigma_u^2/\sigma^2 = 0$, i.e. $\mathcal{W}_{1-\alpha}(c,d)$ is the $1-\alpha$ quantile of the distribution of

$$\mathcal{W}(c,d) \sim \frac{\int \left\{ \widetilde{J}_{c,d} \cdot \int \widetilde{G}_{c,d} \mathrm{d}W - \widetilde{G}_{c,d} \cdot \int \widetilde{J}_{c,d} \mathrm{d}W \right\}^2}{\int \widetilde{J}_{c,d}^2 \cdot \int \widetilde{G}_{c,d}^2 - \left(\int \widetilde{J}_{c,d} \widetilde{G}_{c,d} \right)^2}, \tag{6.1}$$

where the definitions of $\widetilde{J}_{c,d}$ and $\widetilde{G}_{c,d}$ correspond to the specification for D_t. The confidence set for (c,d) can now be constructed by test inversion as

$$CS_{n,1-\alpha} \equiv \left\{ (c,d) : W_n\big(\Phi_{1,n}(c,d), \Phi_{2,n}(c,d)\big) \leq \mathcal{W}_{1-\alpha}(c,d) \right\}.$$

The confidence set $CS_{n,1-\alpha}$ is bounded as $\mathcal{W}_{1-\alpha}(c,d) \to \chi^2_{2,1-\alpha}$ when $c \to -\infty$ or $d \to \infty$. In practice, the confidence set can be approximated by choosing a dense two-dimensional grid of values c and d. We use a grid with $c \leq 0$ and $2\pi < d < n\pi$, where the lower bound of 2π is imposed to rule out cycles longer than the sample size when measured by τ_θ.

The construction of $CS_{n,1-\alpha}$ is akin to the grid bootstrap procedure of Hansen (1999), however, we use the asymptotic critical values instead of their bootstrap approximation. Note that the critical values must be adjusted for every considered point (c,d). The validity of $CS_{n,1-\alpha}$ is due to the following facts. First, (c,d) is included in the confidence set only if the null hypothesis $H_0 : \phi_{1,n} = \Phi_{1,n}(c,d), \phi_{2,n} = \Phi_{2,n}(c,d)$ cannot be rejected by the Wald test with the critical value $\mathcal{W}_{1-\alpha}(c,d)$. Second, the critical values are computed using the same values (c,d) as those specified under H_0. Third, the distribution in (6.1) nests the χ^2_2 distribution, which arises under the fixed (ϕ_1, ϕ_2) asymptotics, as a limiting case. Note that having correct size under both drifting and fixed parameters specifications is required for the uniform validity (Andrews, Cheng, and Guggenberger, 2020).

We construct confidence intervals for τ_θ and τ_ω from $CS_{n,1-\alpha}$ by projection:

$$CI^{\tau_\theta}_{n,1-\alpha} \equiv \left[\inf_{d:(c,d)\in CS_{n,1-\alpha}} \frac{2\pi}{d}, \sup_{d:(c,d)\in CS_{n,1-\alpha}} \frac{2\pi}{d} \right], \tag{6.2}$$

$$CI^{\tau_\omega}_{n,1-\alpha} \equiv \left[\inf_{(c,d)\in CS_{n,1-\alpha}} \frac{2\pi}{\sqrt{d^2-c^2}}, \sup_{(c,d)\in CS_{n,1-\alpha}} \frac{2\pi}{\sqrt{d^2-c^2}} \right]. \tag{6.3}$$

The confidence interval for τ_θ is bounded as long as the grid of d values used to construct $CS_{n,1-\alpha}$ excludes zero. On the other hand, the confidence interval for τ_ω can be unbounded if pairs (c,d) with $c = d$ are included in $CS_{n,1-\alpha}$.

6.2. Serially correlated $\{u_t\}$

In this section we assume that the innovations process $\{u_t\}$ is generated as AR(p):

$$(1 - \rho_1 L - \ldots - \rho_p L^p) u_t = \varepsilon_t, \tag{6.4}$$

where $\{\varepsilon_t\}$ are iid $(0, \sigma^2_\varepsilon)$, and the roots of the polynomial $1 - \rho_1 L - \ldots - \rho_p L^p$ are real and bounded away from unity. By running the regressions of y_t against different specifications of the deterministic part D_t and $y_{t-1}, y_{t-1}, \ldots, y_{t-2-m}$ for some $m > p$, one can again use the BIC selection procedure to consistently estimate the specification for D_t and the number of lags p. We now proceed assuming that the model for D_t and the number of lags p are known.

Let $\widetilde{y}_t$ denote the residuals from projection of y_t against the components of D_t. Under H_0 : $\phi_{1,n} = \phi_{1,0}, \phi_{2,n} = \phi_{2,0}$, the values $\phi_{1,n}$ and $\phi_{2,n}$ are known and can be computed from the values of c and d. Let

$$\widetilde{u}_{t,0} \equiv \left(1 - \phi_{1,0}L - \phi_{2,0}L^2\right)\widetilde{y}_t.$$

Using the null-restricted residuals $\widetilde{u}_{t,0}$, one can estimate the autoregressive coefficients $\rho_1, \ldots, \rho_p$. Let $\widehat{\rho}_{1,0}, \ldots, \widehat{\rho}_{p,0}$ denote their least-squares estimators. Note that under H_0, these estimators are consistent. We can now remove the autoregressive part in u_t:

$$\widehat{x}_{t,0} \equiv (1 - \widehat{\rho}_{1,0}L - \ldots - \widehat{\rho}_{p,0}L^p)\widetilde{y}_t.$$

Thus, to construct the process $\widehat{x}_{t,0}$, we have filtered out the deterministic part D_t and serial correlation in $\{u_t\}$. Note that under the null, the population counterpart of $\widehat{x}_{t,0}$ satisfies:

$$\widetilde{x}_t \equiv (1 - \rho_1 L - \ldots - \rho_p L^p)\widetilde{y}_t = \frac{\widetilde{\varepsilon}_t}{1 - \phi_{1,n}L - \phi_{2,n}L^2},$$

where $\widetilde{\varepsilon}_t$ is the residual from the projection of ε_t against the components of D_t.

One can now use $\{\widehat{x}_{t,0}\}$ for inference on the cyclical properties of $\{y_t\}$, however, additional adjustments are required to account for estimation of $\rho_1, \ldots, \rho_p$. The main purpose of the adjustments discussed below is to ensure that the modified Wald statistic has the correct asymptotic null distributions both under the long-cycle asymptotics proposed in the paper as well as under the standard asymptotics with ϕ_1 and ϕ_2 fixed in the stationary range.[13]

Let $\widehat{\phi}_{1,n}$ and $\widehat{\phi}_{1,n}$ now denote the least-squares estimators of ϕ_1 and ϕ_2 respectively from the regression of $\widehat{x}_{t,0}$ against $\widehat{x}_{t-1,0}$ and $\widehat{x}_{t-2,0}$:

$$\widehat{x}_{t,0} = \widehat{\phi}_{1,n}\widehat{x}_{t-1,0} + \widehat{\phi}_{2,n}\widehat{x}_{t-2,0} + \widehat{\varepsilon}_{t,0},$$

where $\widehat{\varepsilon}_{t,0}$ denotes the least-squares residuals. The modified Wald statistic takes the form

$$W_{n,p}(\phi_{1,0}, \phi_{2,0}) \equiv \frac{1}{\widehat{\sigma}^2_{\varepsilon,n}} \begin{pmatrix} \widehat{\phi}_{1,n} - \phi_{1,0} \\ \widehat{\phi}_{2,n} - \phi_{2,0} \end{pmatrix}^\top M_n \Sigma_n^{-1} M_n \begin{pmatrix} \widehat{\phi}_{1,n} - \phi_{1,0} \\ \widehat{\phi}_{2,n} - \phi_{2,0} \end{pmatrix},$$

where $\widehat{\sigma}^2_{\varepsilon,n} \equiv n^{-1} \sum \widehat{\varepsilon}^2_{t,0}$, and the matrix M_n is given by

$$M_n \equiv \begin{pmatrix} \sum \widehat{x}^2_{t-1,0} & \sum \widehat{x}_{t-1,0}\widehat{x}_{t-2,0} \\ \sum \widehat{x}_{t-1,0}\widehat{x}_{t-2,0} & \sum \widehat{x}^2_{t-2,0} \end{pmatrix}.$$

To construct Σ_n, we first define $\dot{x}_{t,0}$ and $\ddot{x}_{t,0}$ as the residual from the least-squares regression of $\widehat{x}_{t,0}$ and $\widehat{x}_{t-1,0}$ respectively against $\widetilde{u}_{t,0}, \ldots, \widetilde{u}_{t-p+1,0}$:

$$\begin{aligned} \widehat{x}_{t,0} &= \dot{\zeta}_{1,n}\widetilde{u}_{t,0} + \ldots + \dot{\zeta}_{p,n}\widetilde{u}_{t-p+1,0} + \dot{x}_{t,0}, \\ \widehat{x}_{t-1,0} &= \ddot{\zeta}_{1,n}\widetilde{u}_{t,0} + \ldots + \ddot{\zeta}_{p,n}\widetilde{u}_{t-p+1,0} + \ddot{x}_{t-1,0}, \end{aligned} \tag{6.5}$$

[13] Recall that having correct size under both drifting and fixed parameters specifications is required for the uniform validity (Andrews, Cheng, and Guggenberger, 2020).

where $\dot{\zeta}_{1,n}, \ldots, \dot{\zeta}_{p,n}$ and $\ddot{\zeta}_{1,n}, \ldots, \ddot{\zeta}_{p,n}$ are the OLS estimators. The matrix Σ_n is given by

$$\Sigma_n \equiv \begin{pmatrix} \sum \dot{x}_{t-1,0}^2 & \sum \dot{x}_{t-1,0}\ddot{x}_{t-2,0} \\ \sum \dot{x}_{t-1,0}\ddot{x}_{t-2,0} & \sum \ddot{x}_{t-2,0}^2 \end{pmatrix}.$$

The next proposition shows that under the conventional stationary asymptotics, the asymptotic null distribution of the Wald statistic is the usual χ_2^2 distribution.

Proposition 6.1. *Suppose that $\{y_t\}$ is generated according to $(1 - \phi_1 L - \phi_2 L^2)y_t = u_t$ with the coefficients ϕ_1 and ϕ_2 fixed in the stationary range, and $\{u_t\}$ satisfying (6.4) with the coefficients $\rho_1, \ldots, \rho_p$ in the stationary range and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Then,*

$$W_{n,p}(\phi_1, \phi_2) \Rightarrow \chi_2^2.$$

In the case of a long-cycle specification, the null asymptotic distribution of the modified Wald statistic is the same as in (6.1).

Proposition 6.2. *Suppose that $\{y_t\}$ is generated according to (4.1), where $\{u_t\}$ satisfies (6.4) with the coefficients $\rho_1, \ldots, \rho_p$ in the stationary range and $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$. Suppose further that Assumption 2.1 holds. Then,*

$$W_{n,p}(\phi_{1,n}, \phi_{2,n}) \Rightarrow \mathcal{W}(c, d),$$

where $W(c, d)$ is defined in (6.1) with $\widetilde{J}_{c,d}$ and $\widetilde{G}_{c,d}$ defined according to the specification of D_t.

Using the results of Propositions 6.1 and 6.2, one can now construct confidence sets for (c, d) using the modified Wald statistic as

$$CS_{n,p,1-\alpha} \equiv \Big\{(c, d) : W_{n,p}\big(\Phi_{1,n}(c, d), \Phi_{2,n}(c, d)\big) \leq \mathcal{W}_{1-\alpha}(c, d)\Big\}.$$

Similarly to the construction in (6.2) and (6.3), the confidence set $CS_{n,p,1-\alpha}$ can be projected to construct confidence intervals for τ_θ and τ_ω.

7. Cyclical properties of macroeconomic and financial variables

In this section, we apply our inference procedure to the quarterly series of a set of macroeconomic and financial variables for the U.S. All data are publicly available from FRED, Federal Reserve Bank of St. Louis. A detailed description of the data is summarized in Table 3 in Appendix B. All the series are measured in natural logs except for the credit-to-GDP ratio (for the private non-financial sector), which is in percentage points, and the interest rate spread between Moody's seasoned BAA corporate bond yield and the 10-year treasury constant maturity, which is expressed in levels. For each series, we take the longest and the most updated sample ending in 2020. Depending on the series, our samples span periods ranging from 34 to 73 years.

We use the empirical models in (4.1). Let y_t denote the observed data series such that

$$y_t = y_t^c + D_t,$$
$$y_t^c = \phi_{1,n} y_{t-1}^c + \phi_{2,n} y_{t-2}^c + u_t,$$

TABLE 2. Length of cycle in quarters

	$n\tau_\theta$	$n\tau_\omega$	n	Linear time trend	Deterministic cycles	Autocorr. u_t
Macroeconomic variables						
Real GDP per capita	— (23, 264)	— (25, 512)	294	Yes	No	No
Unemployment rate	52 (22, 260)	— (27, 185)	290	No	No	No
Hours per capita	22 (18, 64)	25 (18, 198)	294	Yes	$k = 1$	No
Financial variables						
VXO S&P 100 volatility index	— $\emptyset$	— $\emptyset$	139	No	$k = 3$	No
Credit risk premium BAA to 10Y	— $\emptyset$	— $\emptyset$	269	Yes	No	No
Equity price index	— $\emptyset$	— $\emptyset$	197	Yes	No	No
Private non-financial sector credit % GDP	— (50, 245)	— (54, 358)	273	Yes	No	AR(1)
Home price index	63 (42, 120)	77 (43, 234)	134	Yes	No	No

[1] All data series are sampled at quarterly frequencies with the sample size of each series given by n.
[2] Columns 1 and 2 indicate respectively the length of cycles measured based on angular frequency $n\tau_\theta$ and spectrum-maximizing frequency $n\tau_\omega$.
[3] In columns 1 and 2, the numbers on top indicate the point estimates of the cycle length. A dash line "—" is used when the point estimate corresponds to an acyclical process and when the point estimate is unavailable in case of autocorrelation. Enclosed in the parentheses are the minimum and maximum cycle lengths implied by the 95% confidence intervals of τ_θ and τ_ω. When the interval is empty, it is indicated by the symbol "$\emptyset$". All numbers are in quarters.
[4] The intercept is included in all specifications.
[5] Credit to private non-financial sector (% GDP) are seasonally adjusted by including seasonal dummies in the regression.

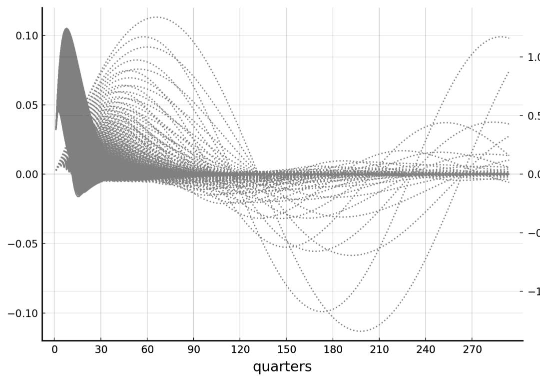
where y_t^c is the latent cyclical part, the innovations $\{u_t\}$ are potentially serially correlated according to an AR(p) specification with unknown p, and D_t may contain linear deterministic trends and deterministic cycles as discussed in Section 4. In all specifications, the intercept (constant) is included by default. The raw data for the credit-to-GDP ratio is not seasonally adjusted and, therefore, its specification for D_t also allows for seasonal dummies.

We use the BIC to select the appropriate specification for D_t (i.e. whether to include a linear time trend, deterministic cycles or seasonal dummies). We also rely on the BIC to determine the presence of autocorrelation in $\{u_t\}$. Note that a BIC estimate of the lag order $p + 2 \geq 3$ for $\{y_t\}$ implies an autocorrelation of order p for $\{u_t\}$. In our empirical application, most of the time series do not exhibit autocorrelation in $\{u_t\}$ except for credit to GDP ratio, as indicated by the BIC. Moreover, credit-to-GDP ratio is the only series where we have included the seasonal dummies.

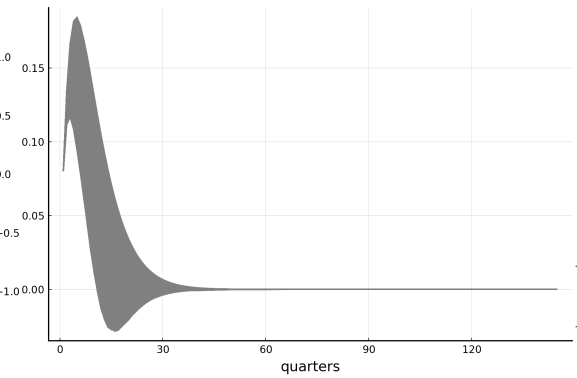
Table 2 presents our results. Note that the last three columns of the table describe the specification selected by the BIC for D_t and the order of autocorrelation for $\{u_t\}$. For example, hours per capita contains a linear time trend and deterministic cycles of cosine and sine waves with $k = 1$, which corresponds to a periodicity of $n/k = n$. According to the BIC, the errors $\{u_t\}$ are serially uncorrelated. Columns 1 and 2 of the table report respectively the angular frequency-based measure $n\tau_\theta$ and the spectrum-maximizing frequency-based measure $n\tau_\omega$ for the cycle length. The point estimates for $n\tau_\theta$ and $n\tau_\omega$ are indicated as "—" when the autoregressive coefficient estimates of $\phi_{1,n}$ and $\phi_{2,n}$ correspond to acyclical processes, or when they are not available as in the case of autocorrelation. The minimum and maximum cycle length implied by the 95% confidence intervals of $n\tau_\theta$ and $n\tau_\omega$ are given in the parentheses.

The two alternative measures of cycle length generally produce similar lower bound estimates. Based on the 95% confidence intervals, we are unable to reject the null that macroeconomic variables, such as real GDP per capita, unemployment rate and hours per capita, contain stochastic cycles with periodicity of at least 5-6 years. Partly due to the projection-based construction of the $n\tau_\theta$ and $n\tau_\omega$ confidence intervals, the implied range of the cycle length is typically wide. The upper bound confidence estimates usually are large and differ considerably between the two measures. Nevertheless, our results point to the presence of cyclicality among macroeconomic variables, conforming to the view of endogenous business cycles (Beaudry, Galizia, and Portier, 2020). On the financial side, we find that credit to private non-financial sector as a percent of GDP and home prices exhibit cycles of at least 10 years in duration, twice as long as the minimum detected cycle length in the macroeconomic variables.

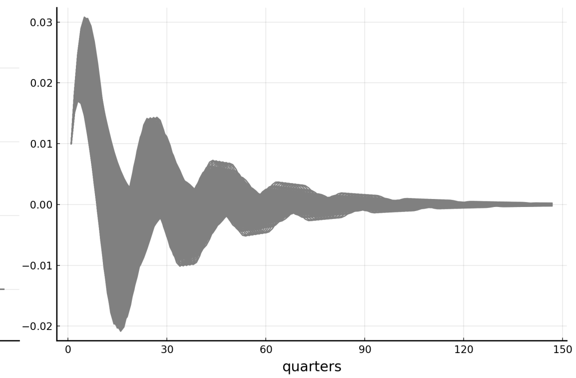
The most striking finding of this section is that for the asset market variables (the volatility index, credit risk premium, and equity prices), our procedure returns empty confidence sets. This suggests that the underlying mechanism for asset market fluctuations is different from that of the macro variables and the financial variables such as the credit and home prices. Our results are in favour of the dichotomy between the asset market and the real economy. Moreover, the results do not support the view that the recessions are driven by risk perception, risk premiums and risk-bearing capacity suggested in the macro-finance literature (see Cochrane, 2017). Note that the S&P 100 Volatility Index, a measure of market uncertainty, has a deterministic cycle of approximately 46 quarters in length according to the BIC model selection. However, it is different in nature from stochastic cycles detected in the other variables.



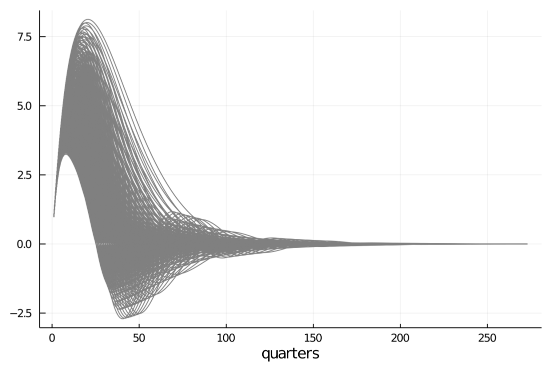
(A) Real GDP per capita



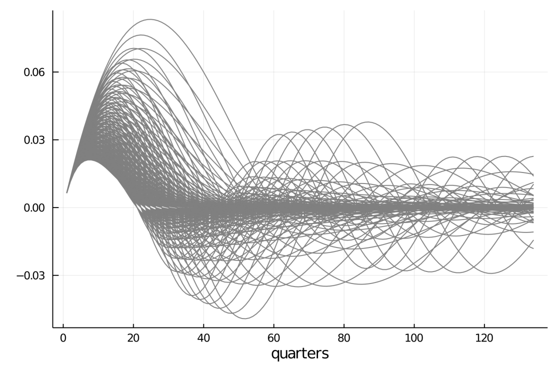
(B) Unemployment rate



(C) Hours per capita



(D) Credit to private non-financial sector % GDP



(E) Home price index

FIGURE 2. Impulse responses to a one-standard-deviation shock to innovations

To better visualize the cyclical dynamics consistent with the data, for each series y_t, we plot in Figure 2 the impulse responses to a one-standard-deviation shock to the innovation u_0 of all cyclical specifications in the 95% confidence sets $CS_{n,1-\alpha}$.[14] The dynamics shown in the figure resonate with the results in Table 2. Financial variables such as credit to private non-financial sector and home prices exhibit much longer cycles than the business cycle variables. The duration from peaks to troughs is at least 25-30 quarters in financial cycles and at least 15 quarters in business cycles. In addition, the financial cycles are also more pronounced. For a one-standard deviation shock, the initial amplitude of the cyclical response is approximately 3 to 7 times the standard deviation for credit and home prices, and about 1.5 to 3 times for unemployment rate and hours per capita.

For real GDP per capita, the impulse responses are split into two parts. On the left, the axis corresponds the set of impulse responses similar to those observed in unemployment rate and hours per capita. On the right, the axis maps to the set of cyclical impulse responses with large amplitudes and high persistences. Note that the scale of the axis on the right has increased by 10-fold. While the possibility of having much longer and highly persistent stochastic cycles cannot be rejected, real GDP per capita do also share similar dynamics to unemployment rate and hours per capita.[15]

In sum, our results suggest that business cycles as marked by the expansions and contractions of the aggregate economic activity are not just recurrent but periodic with an average duration of at least 5-6 years. Furthermore, financial cycles as characterized by the booms and busts in credit and home prices are much longer than the business cycles: at least 10 years in duration. In addition, these financial cycles have more prominent oscillations with much larger amplitudes than business cycles. Moreover, we find that equity prices, though commonly included in the characterization of financial cycles, do not exhibit stochastic cycles, and therefore merits separate consideration from credit and home prices. Lastly, our results suggest that asset market fluctuations are a different phenomenon from the changes in real economic activities.

Appendix A. The periodogram of long-cycle processes

A.1. Asymptotic properties

Periodogram-based nonparametric estimators are commonly used to infer on cyclical behaviour of time series. In this section, we derive the asymptotic properties of the periodogram in the case of long-cycle processes. For $-\pi \leq \omega \leq \pi$, the periodogram of $\{y_t\}$ is defined as

$$I_n(\omega) \equiv \frac{1}{2\pi n} \left| \sum_{t=1}^{n} y_t e^{-i\omega t} \right|^2, \tag{A.1}$$

[14] For credit to private non-financial sector, the standard deviation of the innovation is computed assuming no serial correlation.

[15] Note also that hours per capita is much more persistent than unemployment rate and real GDP per capita.

see, for example, equation (6.1.24) in Priestley (1981). In the case of covariance stationary processes with continuous spectral densities, it is well known that the periodogram is an asymptotically unbiased estimator of the spectral density at ω (see equation (6.2.12) in Priestley, 1981).

Given the results of Proposition 2.1, in the case of long-cycle processes, we are interested in the spectrum near the origin at frequencies of the form $\omega_n = h/n$ for a constant $h \in \mathbb{R}$. Suppose that $\{y_{n,t}\}$ is generated according to the DGP in equation (2.4) with the roots as in Assumption 2.1. Assume also that $\{u_t\}$ are serially uncorrelated with a zero mean and the variance σ_u^2. In this case, the spectral density of $\{y_t\}$, denoted $f_n(\omega)$, satisfies[16]

$$\begin{aligned} n^{-4} f_n(h/n) &= \frac{\sigma_u^2}{2\pi n^4} \frac{1}{|1-\lambda_{1,n} e^{ih/n}|^2 |1-\lambda_{2,n} e^{ih/n}|^2} \\ &\to \frac{\sigma_u^2}{2\pi} \frac{1}{(c^2+(d+h)^2)(c^2+(d-h)^2)}. \end{aligned}$$

We show below that, in the case of long-cycle processes and near the origin frequencies, the periodogram is a biased estimator.

Proposition A.1. *Suppose that $\{y_{n,t}\}$ is generated according to equation (2.4) and Assumption 2.1, and $\{u_t\}$ are serially uncorrelated with a zero mean and the variance $\sigma_u^2 > 0$. Then for a constant h, the periodogram of $\{y_{n,t}\}$ satisfies*

$$\lim_{n\to\infty} n^{-4} E\left[I_n\left(\frac{h}{n}\right)\right] = \frac{\sigma_u^2}{\pi} \int_{-\infty}^{\infty} \frac{1-\cos(h-x)}{(h-x)^2} \frac{1}{(c^2+(d+x)^2)(c^2+(d-x)^2)} dx. \tag{A.2}$$

The result can be extended to allow for strictly stationary and serially correlated $\{u_t\}$, when the spectral density $\varphi(\omega)$ of $\{u_t\}$ is bounded, bounded away from zero and continuously differentiable with the derivative satisfying $\sup_{x\in[-\pi n,\pi n]} |\varphi'(x/n)| = O(n^{-1})$. For example, the condition holds when $\{u_t\}$ is an MA(p) process. In that case, σ_u^2 in equation (A.2) should be replaced with $\varphi(0)$.

The result in Proposition A.1 can be used to assess the magnitude of the bias implied by the periodogram as we illustrate below. When the cyclical properties of a process are assessed using its spectrum, the appropriate measure of the cycle length is τ_ω. The solid line in Figure 3 plots the limiting expression for the expected values of the periodogram of a long-cycle process at near the origin frequencies. Its maximizing frequency is shown by the solid vertical line. The dashed line displays the limit of the true spectral density. The vertical dashed line indicates the true spectrum maximizing frequency $\sqrt{d^2-c^2}$ derived in Proposition 2.1. To construct the plot, we use the following values of the localization parameters: $c = 4$ and $d = 10$.

The numerical results displayed in the figure demonstrate that the periodogram may under estimate the spectrum maximizing frequency and, as a result, over estimate the length of the cycle. According to the true spectrum, the cycle length relatively to the sample size is $\tau_\omega = 0.69$, while according to the periodogram it is $\tau_\omega = 0.73$. For quarterly data and a sample size $n = 200$, this corresponds to the upward bias of 8 quarters for the cycle length.

[16] See the proof of Proposition A.1.

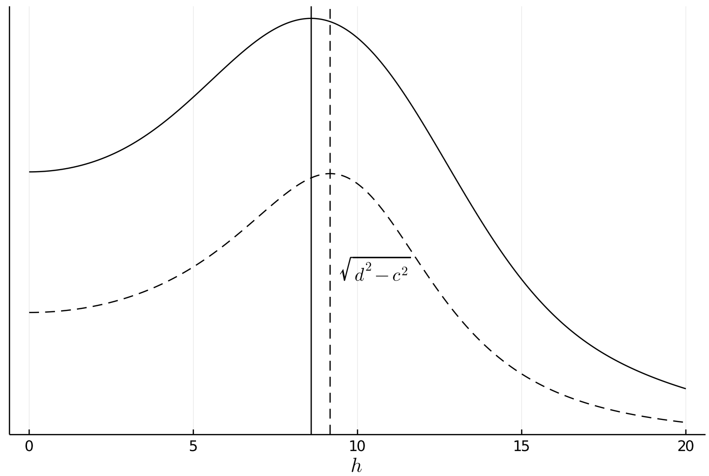


FIGURE 3. The limits of the expected value of the periodogram (solid line) and the true spectrum (dashed line) for $c = 4$ and $d = 10$. The corresponding vertical lines indicate the maximizing frequencies in terms of h, where h is determined by $\omega_n = h/n$, and ω_n denotes frequencies

Proposition 3.1 can be used to describe the asymptotic distribution of the periodogram of a long-cycle process at near-the-origin frequencies. The next result shows that asymptotic distribution of the periodogram depends on a continuous time Fourier transform of the asymptotic approximation of the long-cycle process.

Corollary A.1. *Suppose that $\{y_{n,t}\}$ is generated according to equation (2.4), and Assumptions 2.1 and 3.1 hold. Then,*

$$n^{-4} I_n\left(\frac{h}{n}\right) \Rightarrow \frac{1}{2\pi}\left|\int_0^1 J_{c,d}(r)e^{-ihr}dr\right|^2,$$

where the process $J_{c,d}(r)$ is defined in equation (3.2).

A.2. Proofs of the asymptotic properties of the periodogram

Proof of Proposition A.1. The spectral density of $\{y_{n,t}\}$ is given by

$$f_n(\omega) = \frac{\sigma_u^2}{2\pi}\frac{1}{|1-\lambda_{n,1}e^{i\omega}|^2|1-\lambda_{n,2}e^{i\omega}|^2}. \tag{A.3}$$

By the results in Priestley (1981), equations (6.2.10)–(6.2.11),

$$EI_n\left(\frac{h}{n}\right) = \int_{-\pi}^{\pi} f_n(x)F_n\left(x-\frac{h}{n}\right)dx = \frac{1}{n}\int_{-\pi n}^{\pi n} f_n\left(\frac{x}{n}\right)F_n\left(\frac{x-h}{n}\right)dx, \tag{A.4}$$

where the second result holds by the change of variable, and

$$F_n(x) = \frac{\sin^2(nx/2)}{n\sin^2(x/2)} = \frac{1-\cos(nx)}{n(1-\cos(x))}.$$

Applying a series expansion $\cos((h-x)/n) = 1 - 0.5((h-x)/n)^2 + O((h-x)/n)^4$, we obtain

$$F_n\left(\frac{h-x}{n}\right) = \frac{2n(1-\cos(h-x))}{(h-x)^2\left(1+O\left(\frac{h-x}{n}\right)^2\right)}. \tag{A.5}$$

Next, we consider an expansion of the elements of $f_n(x/n)$.

$$\begin{aligned}
& 1-\lambda_{n,1}e^{ix/n} \\
= \; & 1-e^{c/n}\left(\cos\left(\frac{d+x}{n}\right)+i\sin\left(\frac{d+x}{n}\right)\right) \\
= \; & 1-\left(1+\frac{c}{n}+O\left(\frac{1}{n^2}\right)\right) \\
& \times\left(1-\frac{1}{2}\left(\frac{d+x}{n}\right)^2+O\left(\frac{d+x}{n}\right)^4+i\left(\frac{d+x}{n}+O\left(\frac{d+x}{n}\right)^3\right)\right) \\
= \; & -\frac{c}{n}+O\left(\left(\frac{d+x}{n}\right)^2+\frac{(d+x)^2}{n^3}+\left(\frac{d+x}{n}\right)^4+\frac{1}{n^2}\right) \\
& +i\left(\frac{d+x}{n}+O\left(\frac{d+x}{n^2}+\frac{(d+x)^3}{n^4}\right)\right).
\end{aligned}$$

Hence,

$$\left|1-\lambda_{n,1}e^{ix/n}\right|^2 = \frac{c^2+(d+x)^2}{n^2}+O\left(\frac{d+x}{n}\right)^4. \tag{A.6}$$

Similarly,

$$\left|1-\lambda_{n,2}e^{ix/n}\right|^2 = \frac{c^2+(d-x)^2}{n^2}+O\left(\frac{d-x}{n}\right)^4. \tag{A.7}$$

By (A.3) and (A.6)–(A.7),

$$\frac{1}{n}f_n\left(\frac{x}{n}\right) = \frac{n^3\sigma^2}{2\pi\left(c^2+(d+x)^2\left(1+O\left(\frac{d+x}{n}\right)^2\right)\right)\left(c^2+(d-x)^2\left(1+O\left(\frac{d-x}{n}\right)^2\right)\right)}. \tag{A.8}$$

The result of the proposition follows from (A.4), (A.5), and (A.8).

□

Proof of Corollary A.1. Since

$$\int_{t/n}^{(t+1)/n} e^{-ihs}ds = \frac{e^{-iht/n}}{n}\left(1+O(n^{-1})\right),$$

we have

$$\begin{aligned}
n^{-5/2}\sum_{t=1}^{n} y_t e^{-iht/n} &= \frac{1}{1+O(n^{-1})}\sum_{t=1}^{n} n^{-3/2}y_t\int_{t/n}^{(t+1)/n} e^{-ihs}\mathrm{d}s \\
&= \frac{1}{1+O(n^{-1})}\sum_{t=1}^{n}\int_{t/n}^{(t+1)/n} n^{-3/2}y_{\lfloor ns\rfloor}e^{-ihs}\mathrm{d}s \\
&= \frac{1}{1+O(n^{-1})}\int_0^1 n^{-3/2}y_{\lfloor ns\rfloor}e^{-ihs}\mathrm{d}s+O_p(n^{-1})
\end{aligned}$$

$$\Rightarrow \int_0^1 J_{c,d} e^{-ihs} \mathrm{d}s, \tag{A.9}$$

where the equality in the third line holds because $y_{\lfloor ns \rfloor} = y_t$ for $t/n \leq s < (t+1)/n$, and the equality in the last line holds by the Continuous Mapping Theorem (CMT) and Proposition 3.1. The result of the corollary follows by (A.9) and the CMT. □

Appendix B. Description of the data in Section 7

Table 3 in this appendix provides a description of the variables used in the empirical application in Section 7. The description includes the source with exact identifiers for each variable, any transformations applied to the raw data, and the sample periods.

TABLE 3. Data description

	Source	Identifier	Construction	Sample
Real GDP per capita	FRED	A939RX0Q048SBEA	Natural logarithm	1947Q1 to 2020Q2
Unemployment rate	FRED	UNRATE	Natural logarithm	1948Q1 to 2020Q2
Hours per capita	FRED	HOANBS B230RC0Q173SBEA	Ratio of non-farm business hours to population, Natural logarithm	1947Q1 to 2020Q2
VXO S&P 100 volatility index	FRED	VXOCLS	Natural logarithm	1986Q1 to 2020Q3
Credit risk premium BAA to 10Y	FRED	BAA10YM	—	1953Q2 to 2020Q2
Credit to non-financial sector % GDP	FRED	QUSPAM770A	—	1952Q1 to 2020Q1
Home price index	FRED	CSUSHPISA	S&P/Case-Shiller U.S. National Home Price Index, Natural logarithm	1987Q1 to 2020Q2
Equity price index	FRED	WILL5000IND CPALTT01USQ661S	Wilshire 5000 Total Market Index divided by CPI, Natural logarithm	1971Q2 to 2020Q2

In the case where aggregation is needed, the end of period values are used.

Appendix C. Proofs of the main results

Proof of Proposition 2.1. By Assumption 2.1 and (2.6)–(2.7),

$$\begin{aligned} -\frac{\phi_{1,n}(1-\phi_{2,n})}{4\phi_{2,n}} &= 0.5\cos\bigl(dn^{-1}\bigr)\bigl(\exp\bigl(cn^{-1}\bigr) + \exp\bigl(-cn^{-1}\bigr)\bigr) \\ &= (1 - 0.5d^2n^{-2} + O(n^{-4}))(1 + 0.5c^2n^{-2} + O(n^{-4})) \\ &= 1 - 0.5(d^2 - c^2)n^{-2} + O(n^{-4}). \end{aligned}$$

Since the argument of $\cos^{-1}$ converges to one, it follows that

$$\omega_n^* = \cos^{-1}\bigl(1 - 0.5(d^2 - c^2)n^{-2} + O(n^{-4})\bigr) = o(1). \tag{C.1}$$

Consider $\cos^{-1}(1-s) = t$ or $1 - s = \cos(t)$, where s and t are small. Expanding $\cos(t)$ around $t = 0$, we obtain $s = t^2/2 + O(t^4)$. Hence, $2s = t^2(1 + O(t^2))$, and it follows that

$$\begin{aligned} t &= \sqrt{2s}(1 + O(t^2)) \\ &= \sqrt{2s}(1 + O(2s(1 + O(t^2)))) \\ &= \sqrt{2s} + O(s^{3/2}). \end{aligned}$$

Therefore,

$$n\omega_n^* = \sqrt{d^2 - c^2} + O(n^{-2}).$$

□

Lemmas C.1, C.2, and C.3 below present auxiliary results that are needed for the proof of Proposition 3.2 and Lemma 3.1. In particular, Lemma C.3 establishes the properties of the diffusion processes that appear in the limiting expressions for the estimators and test statistics.

Lemma C.1. *Suppose that Assumption 2.1 holds. The following approximation holds for the long-cycle autoregressive coefficients in* (2.6) *and* (2.7)*:*

(a) $\phi_{1,n} = 2 + \frac{2c}{n} + \frac{c^2-d^2}{n^2} + O(n^{-3})$.
(b) $-\phi_{2,n} = 1 + \frac{2c}{n} + \frac{2c^2}{n^2} + O(n^{-3})$.
(c) $\frac{\phi_{1,n}}{1-\phi_{2,n}} = 1 - \frac{d^2+c^2}{2n^2} + O(n^{-3})$.
(d) $\frac{2}{1-\phi_{2,n}} = 1 - \frac{c}{n} - \frac{c^2}{n^2} + O(n^{-3})$.
(e) $\phi_{1,n} + \phi_{2,n} = 1 - \frac{c^2+d^2}{n^2} + O(n^{-3})$.
(f) $-(\phi_{1,n} + \phi_{2,n})\phi_{2,n} = 1 + \frac{2c}{n} + \frac{c^2-d^2}{n^2} + O(n^{-3})$.
(g) $1 - (\phi_{1,n} + \phi_{2,n})^2 = \frac{2(c^2+d^2)}{n^2} + O(n^{-3})$.
(h) $\phi_{2,n}^2 = 1 + \frac{4c}{n} + \frac{8c^2}{n^2} + O(n^{-3})$.

Lemma C.2. *Suppose* $\{y_{n,t}\}$ *is generated according to* (2.4)*. We have:*

(a) $\sum_{t=2}^n y_{t-2}^2 = \sum_{t=1}^n y_{t-1}^2 - y_{n-1}^2$.
(b) $\sum_{t=3}^n y_{t-1}y_{t-2} = \frac{\phi_{1,n}}{1-\phi_{2,n}} \sum_{t=2}^n y_{t-1}^2 + \frac{1}{1-\phi_{2,n}}(\sum_{t=2}^n y_{t-1}u_t - y_n y_{n-1}) + \frac{\phi_{2,n}}{1-\phi_{2,n}} y_0 y_1$.
(c) $\sum_{t=2}^n y_{t-2}u_t = \sum_{t=2}^n y_{t-1}u_t - \sum_{t=2}^n (y_{t-1} - y_{t-2})u_t$.
(d) $y_n = y_{n-1} + (y_n - y_{n-1})$.
(e) $y_n y_{n-1} = y_n^2 - y_n(y_n - y_{n-1})$.
(f) $y_{n-1}^2 = y_n^2 - 2y_n(y_n - y_{n-1}) + (y_n - y_{n-1})^2$.
(g) $\sum y_{t-1}(\Delta y_t - \Delta y_{t-1}) = y_{n-1}\Delta y_n - \sum(\Delta y_{t-1})^2$.

Lemma C.3. *The diffusion processes* $J_{c,d}(\cdot)$*,* $K_{c,d}(\cdot)$*, and* $G_{c,d}(\cdot)$ *have the following properties:*

(a) $\mathrm{d}J_{c,d}(r) = c \cdot J_{c,d}(r)\mathrm{d}r + d \cdot K_{c,d}(r)\mathrm{d}r = G_{c,d}(r)\mathrm{d}r$.
(b) $\mathrm{d}K_{c,d}(r) = c \cdot K_{c,d}(r)\mathrm{d}r - d \cdot J_{c,d}(r)\mathrm{d}r + \frac{1}{d}\mathrm{d}W(r)$.
(c) $\int_0^r e^{2c(r-s)} J_{c,d}(r)\mathrm{d}s = \frac{1}{c^2+d^2}\Big\{ \int_0^r e^{2c(r-s)}\mathrm{d}W(s) - \Big(c \cdot J_{c,d}(r) + d \cdot K_{c,d}(r)\Big)\Big\}$.
(d) $\mathrm{d}\big(J_{c,d}(r) \cdot K_{c,d}(r)\big) = 2c \cdot J_{c,d}(r) \cdot K_{c,d}(r)\mathrm{d}r + d \cdot (K_{c,d}^2(r) - J_{c,d}^2(r))\mathrm{d}r + \frac{1}{d} J_{c,d}(r)\mathrm{d}W(r)$.
(e) $\int_0^1 G_{c,d}^2(r)\mathrm{d}r = (c^2 + d^2)\int_0^1 J_{c,d}^2(r)\mathrm{d}r + J_{c,d}(1)G_{c,d}(1) - \int_0^1 J_{c,d}(r)\mathrm{d}W(r) - c \cdot J_{c,d}^2(1)$
(f) $J^2(1) = 2\int_0^1 J_{c,d}(r)G_{c,d}(r)\mathrm{d}r$

(g) $(G_{c,d}^2(1)-1)/2 = c\int_0^1 G_{c,d}^2(r)\mathrm{d}r + cd\int_0^1 K_{c,d}(r)G_{c,d}(r)\mathrm{d}r - d^2\int_0^1 J_{c,d}(r)G_{c,d}(r)\mathrm{d}r + \int_0^1 G_{c,d}(r)\mathrm{d}W(r)$.

Proof of Lemma C.3. To prove part (a) and (b), note that by applying trigonometric identities, we have

$$\begin{aligned} J_{c,d}(r) &= \frac{1}{d}\int_0^r e^{c(r-s)}\left\{\sin(dr)\cos(ds) - \cos(dr)\sin(ds)\right\}\mathrm{d}W(s), \\ K_{c,d}(r) &= \frac{1}{d}\int_0^r e^{c(r-s)}\left\{\cos(dr)\cos(ds) + \sin(dr)\sin(ds)\right\}\mathrm{d}W(s). \end{aligned}$$

By applying stochastic differentiation of $J_{c,d}(r)$ and $K_{c,d}(r)$,

$$\begin{aligned} d\cdot\mathrm{d}J_{c,d}(r) &= (c\cdot e^{cr}\sin(dr) + d\cdot e^{cr}\cos(dr))\int_0^r e^{-cs}\cos(ds)\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + e^{cr}\sin(dr)e^{-cr}\cos(dr)\mathrm{d}W(r) \\ &\quad - (c\cdot e^{cr}\cos(dr) - d\cdot e^{cr}\sin(dr))\int_0^r e^{-cs}\sin(ds)\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad - e^{cr}\sin(dr)e^{-cr}\cos(dr)\mathrm{d}W(r) \\ &= c\int_0^r e^{c(r-s)}\left\{\sin(dr)\cos(ds) - \cos(dr)\sin(ds)\right\}\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + d\int_0^r e^{c(r-s)}\left\{\cos(dr)\cos(ds) + \cos(dr)\sin(ds)\right\}\mathrm{d}W(s)\cdot\mathrm{d}r, \\ d\cdot\mathrm{d}K_{c,d}(r) &= (c\cdot e^{cr}\cos(dr) - d\cdot e^{cr}\sin(dr))\int_0^r e^{-cs}\cos(ds)\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + e^{cr}\cos(dr)e^{-cr}\cos(dr)\mathrm{d}W(r) \\ &\quad + (c\cdot e^{cr}\sin(dr) + d\cdot e^{cr}\cos(dr))\int_0^r e^{-cs}\sin(ds)\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + e^{cr}\sin(dr)e^{-cr}\sin(dr)\mathrm{d}W(r) \\ &= c\int_0^r e^{c(r-s)}\left\{\cos(dr)\cos(ds) + \sin(dr)\sin(ds)\right\}\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + d\int_0^r e^{c(r-s)}\left\{\sin(dr)\cos(ds) - \cos(dr)\sin(ds)\right\}\mathrm{d}W(s)\cdot\mathrm{d}r \\ &\quad + \mathrm{d}W(r). \end{aligned}$$

Parts (a) and (b) now follow from the trigonometric identities. To prove part (c), use the results from (a) and (b) and evaluate the following integrals using integration by parts:

$$\begin{aligned} \int_0^r e^{2c(r-s)}J_{c,d}(s)\mathrm{d}s &= \frac{d}{c}\int_0^r e^{2c(r-s)}K_{c,d}(s)\mathrm{d}s - \frac{1}{c}J_{c,d}(r), \\ \int_0^r e^{2c(r-s)}K_{c,d}(s)\mathrm{d}s &= \frac{1}{cd}\int_0^r e^{2c(r-s)}\mathrm{d}W(s) - \frac{d}{c}\int_c^r e^{2c(r-s)}J_{c,d}(s)\mathrm{d}s - \frac{1}{c}K_{c,d}(r). \end{aligned}$$

With some algebraic manipulations, we obtain part (c).

By Ito's lemma,

$$\mathrm{d}\big(J_{c,d}(r)\cdot K_{c,d}(r)\big) = \mathrm{d}J_{c,d}(r)\cdot K_{c,d}(r) + J_{c,d}(r)\cdot\mathrm{d}K_{c,d}(r).$$

Note that the quadratic covariation is negligible in this case. Using (a) and (b), part (d) follows immediately.

Next we proceed to prove (e). From (d), it follows that

$$d \cdot J_{c,d}(1)K_{c,d}(1) = 2cd \int_0^1 J_{c,d}(r)K_{c,d}(r)\mathrm{d}r + d^2 \int_0^1 (K_{c,d}^2(r) - J_{c,d}^2(r))\mathrm{d}r + \int_0^1 J_{c,d}(r)\mathrm{d}W(r),$$

By the definition of $G_{c,d}(\cdot)$,

$$J_{c,d}(1)G_{c,d}(1) = c \cdot J_{c,d}^2(1) + d \cdot J_{c,d}(1)K_{c,d}(1).$$

By applying the two results from above, we obtain the result in (e):

$$\begin{aligned}
\int_0^1 G_{c,d}^2(r)\mathrm{d}r &= c^2 \int_0^1 J_{c,d}^2(r)\mathrm{d}r + 2cd \int_0^1 J_{c,d}(r)K_{c,d}(r)\mathrm{d}r + d^2 \int_0^1 K_{c,d}^2(r)\mathrm{d}r \\
&= (c^2 + d^2) \int_0^1 J_{c,d}^2(r)\mathrm{d}r + d \cdot J_{c,d}(1)K_{c,d}(1) - \int_0^1 J_{c,d}(r)\mathrm{d}W(r) \\
&= (c^2 + d^2) \int_0^1 J_{c,d}^2(r)\mathrm{d}r + \cdot J_{c,d}(1)G_{c,d}(1) - c \cdot J_{c,d}^2(1) - \int_0^1 J_{c,d}(r)\mathrm{d}W(r).
\end{aligned}$$

To prove (f) and (g), we use stochastic differentiation of $J_{c,d}^2(r)$ and $G_{c,d}^2(r)$, respectively:

$$\begin{aligned}
\mathrm{d}J_{c,d}^2(r) &= 2J_{c,d}(r)\mathrm{d}J_{c,d}(r) = 2J_{c,d}(r)G_{c,d}(r)\mathrm{d}r, \\
\mathrm{d}G_{c,d}^2(r) &= 2G_{c,d}(r)\mathrm{d}G_{c,d}(r) + (\mathrm{d}G_{c,d}(r))^2 \\
&= 2G_{c,d}(r)(c \cdot \mathrm{d}J_{c,d}(r) + d \cdot \mathrm{d}K_{c,d}(r)) + \mathrm{d}r \\
&= 2c \cdot G_{c,d}(r)G_{c,d}(r)\mathrm{d}r + 2cd \cdot G_{c,d}(r)K_{c,d}(r)\mathrm{d}r - 2d^2 \cdot G_{c,d}(r)J_{c,d}(r) \\
&\quad +2G_{c,d}(r)\mathrm{d}W(r) + \mathrm{d}r.
\end{aligned}$$

The results in (f) and (g) follows by integrating both sides of the stochastic differential equations above with respect to r over $[0, 1]$.

$\square$

Proof of Proposition 3.2. By Lemma C.1(a) and (b),

$$\begin{aligned}
y_t &= \left(2 + \frac{2c}{n} + \frac{c^2 - d^2}{n^2} + O(n^{-3})\right)y_{t-1} - \left(1 + \frac{2c}{n} + \frac{2c^2}{n^2} + O(n^{-3})\right)y_{t-2} + u_t, \text{ and} \\
\Delta y_t &= \left(1 + \frac{2c}{n}\right)\Delta y_{t-1} + \left(\frac{c^2 - d^2}{n^2} + O(n^{-3})\right) - y_{t-1}\left(\frac{2c^2}{n^2} + O(n^{-3})\right)y_{t-2} + u_t \\
&= \sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} u_j + \left(\frac{c^2 - d^2}{n^2} + O(n^{-3})\right) \sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} y_{j-1} \\
&\quad -\left(\frac{2c^2}{n^2} + O(n^{-3})\right)\Bigg) \sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} y_{j-2}.
\end{aligned}$$

Define $S_n(r) \equiv \sum_{t=1}^{\lfloor nr \rfloor} u_t$. We have:

$$
\begin{aligned}
\Delta y_{\lfloor nr \rfloor} \;=\;& \sum_{j=0}^{\lfloor nr \rfloor}\left(1+\frac{2c}{n}\right)^{\lfloor nr \rfloor - j}\int_{\frac{j-1}{n}}^{\frac{j}{n}} \mathrm{d}S_n(s) \\
&+\left(\frac{c^2-d^2}{n}+O(n^{-2})\right)\sum_{j=0}^{\lfloor nr \rfloor}\int_{\frac{j-1}{n}}^{\frac{j}{n}}\left(1+\frac{2c}{n}\right)^{\lfloor nr \rfloor - j} y_{\lfloor n\frac{j-1}{n}\rfloor}\mathrm{d}s \\
&-\left(\frac{2c^2}{n}+O(n^{-2})\right)\sum_{j=0}^{\lfloor nr \rfloor}\int_{\frac{j-1}{n}}^{\frac{j}{n}}\left(1+\frac{2c}{n}\right)^{\lfloor nr \rfloor - j} y_{\lfloor n\frac{j-2}{n}\rfloor}\mathrm{d}s.
\end{aligned}
$$

By the CMT and Proposition 3.1,

$$
\begin{aligned}
n^{-1/2}\Delta y_{\lfloor nr \rfloor} &\Rightarrow \sigma\int_0^r e^{2c(r-s)}\mathrm{d}W(s) - \sigma(c^2+d^2)\int_0^r e^{2c(r-s)}J_{c,d}(s)\mathrm{d}s, \\
&= \sigma(c\cdot J_{c,d}(r) + d\cdot K_{c,d}(r)),
\end{aligned}
$$

where the result in the last line follows by Lemma C.3(c). The result of the proposition now follows by the definition of $G_{c,d}(r)$ in (3.6). □

Proof of Lemma 3.1. Parts (a)–(c) follow immediately from Propositions 3.1 and 3.2 by the CMT. To prove the result in part (d), by squaring both sides of equation (2.4) and summing over t, we obtain:

$$
\begin{aligned}
\sum y_t^2 =& (\phi_{1,n}+\phi_{2,n})^2\sum y_{t-1}^2 + \phi_{2,n}^2\sum(\Delta y_{t-1})^2 + \sum u_t^2 \\
&- 2(\phi_{1,n}+\phi_{2,n})\phi_{2,n}\sum y_{t-1}\Delta y_{t-1} + 2(\phi_{1,n}+\phi_{2,n})\sum y_{t-1}u_t - 2\phi_{2,n}\sum \Delta y_{t-1}u_t.
\end{aligned}
$$

After rearranging and applying the results of Lemmas C.1–C.2, we have:

$$
\sum y_{t-1}u_t = \frac{c^2+d^2}{n^2}\sum y_{t-1}^2 + y_n\Delta y_n - \sum(\Delta y_{t-1}^2) - \frac{2c}{n}\sum y_{t-1}\Delta y_{t-1} + O_p(n).
$$

By the results in parts (a)–(c) of the lemma, and using the shortened notation as explained on page 12,

$$
\begin{aligned}
n^{-2}\sum y_{t-1}u_t &\Rightarrow \sigma^2\left((c^2+d^2)\int J_{c,d}^2 + J_{c,d}(1)G_{c,d}(1) - \int G_{c,d}^2 - 2c\int J_{c,d}G_{c,d}\right) \\
&= \sigma^2\int J_{c,d}\mathrm{d}W,
\end{aligned}
$$

where the result in the last line is by Lemma C.3(e) and (f).

To prove part (e), we follow the same steps as in part (d) using (3.4) to obtain

$$
\begin{aligned}
\frac{1}{n}\sum \Delta y_{t-1}u_t &= \frac{c^2+d^2}{n^3}\sum y_{t-1}\Delta y_{t-1} - \frac{2c}{n^2}\sum(\Delta y_{t-1})^2 - \frac{1}{2n}\sum u_t^2 + \frac{1}{2n}(\Delta y_n)^2 + O(n^{-1}) \\
&\Rightarrow \sigma^2(c^2+d^2)\int J_{c,d}G_{c,d} - 2c\sigma^2\int G_{c,d}^2 - \frac{1}{2}\sigma_u^2 + \frac{1}{2}\sigma^2 G_{c,d}^2(1) \\
&= \sigma^2\int G_{c,d}\mathrm{d}W + \frac{1}{2}(\sigma^2 - \sigma_u^2),
\end{aligned}
$$

where the equality in the last line is by part (g) of Lemma C.3 and the definition of $G_{c,d}$. □

Proof of Proposition 3.3. By (3.5),

$$\begin{pmatrix} \widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n} \\ \widehat{\phi}_{2,n} - \phi_{2,n} \end{pmatrix} = \frac{1}{\sum y_{t-1}^2 \sum (\Delta y_{t-1})^2 - (\sum y_{t-1}\Delta y_{t-1})^2} \times \begin{pmatrix} \sum (\Delta y_{t-1})^2 & \sum y_{t-1}\Delta y_{t-1} \\ \sum y_{t-1}\Delta y_{t-1} & \sum y_{t-1}^2 \end{pmatrix} \begin{pmatrix} \sum y_{t-1} u_t \\ -\sum \Delta y_{t-1} u_t \end{pmatrix}.$$

The result in part (a) and the result in part (b) for $\widehat{\phi}_{2,n}$ follow immediately by Lemma 3.1 and the CMT. The result in part (b) for $\widehat{\phi}_{1,n}$ follows since

$$\begin{aligned} n(\widehat{\phi}_{1,n} - \phi_{1,n}) &= n(\widehat{\phi}_{1,n} + \widehat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}) - n(\widehat{\phi}_{2,n} - \phi_{2,n}) \\ &= O_p(n^{-1}) - n(\widehat{\phi}_{2,n} - \phi_{2,n}), \end{aligned}$$

where the second equality holds by the result in part (a).

□

Proof of Proposition 3.4. The result follows from Lemma 3.1(a)-(c) and Proposition 3.3, provided that $\hat{\sigma}_n^2 \to_p \sigma^2$. The long-run variance estimator $\widehat{\sigma}_n^2$ is given by

$$\widehat{\sigma}_n^2 = \hat{\sigma}_{u,n}^2 + 2\sum_{h=1}^{m_n} w_n(h) n^{-1} \sum_{t=h+1}^{n} \hat{u}_t \hat{u}_{t-h}, \text{ where } \widehat{\sigma}_{u,n}^2 = n^{-1}\sum_{t=1}^{n} \hat{u}_t^2.$$

Denote $\phi_{12,n} \equiv \phi_{1,n} + \phi_{2,n}$ and $\widehat{\phi}_{12,n} \equiv \widehat{\phi}_{1,n} + \widehat{\phi}_{2,n}$. We have:

$$\begin{aligned} \widehat{\sigma}_{u,n}^2 &= \frac{1}{n}\sum u_t^2 - (\widehat{\phi}_{12,n} - \phi_{12,n})\frac{2}{n}\sum y_{t-1}u_t + (\widehat{\phi}_{2,n} - \phi_{2,n})\frac{2}{n}\sum \Delta y_{t-1} u_t \\ &\quad + \frac{1}{n}\sum \left((\widehat{\phi}_{12,n} - \phi_{12,n}) y_{t-1} + (\widehat{\phi}_{2,n} - \phi_{2,n}) \Delta y_{t-1} \right)^2 \\ &= \frac{1}{n}\sum u_t^2 + O_p(n^{-1}) \\ &\to_p \sigma_u^2, \end{aligned}$$

where the equality in the line before the last holds by Lemma Lemma 3.1(d),(e) and Proposition 3.3, and the result in the last line holds by Assumption 3.2. By the same arguments and since the weight function $w_n(\cdot)$ is bounded,

$$n^{-1}\sum_{t=h+1}^{n} \hat{u}_t \hat{u}_{t-h} = n^{-1}\sum_{t=h+1}^{n} u_t u_{t-h} + O_p(n^{-1}).$$

Hence,

$$\widehat{\sigma}_n^2 = \tilde{\sigma}_n^2 + O_p(m_n/n),$$

and the result follows by Assumption 3.3. □

Proof of Lemma 4.1. By the results of Propositions 3.1, 3.2, and the CMT,

$$n^{-3/2}\bar{y}_n/\sigma \Rightarrow \int_0^1 J_{c,d}(s)\mathrm{d}s, \quad n^{-1/2}\overline{\Delta y}_n/\sigma \Rightarrow \int_0^1 G_{c,d}(s)\mathrm{d}s.$$

Hence,

$$n^{-3/2}(y_{\lfloor nr\rfloor} - \bar{y}_n)/\sigma \Rightarrow J_{c,d}(s) - \int_0^1 J_{c,d}(r)\mathrm{d}s = \widetilde{J}_{c,d}(r),$$

$$n^{-1/2}(\Delta y_{\lfloor nr\rfloor} - \overline{\Delta y}_n)/\sigma \Rightarrow G_{c,d}(r) - \int_0^1 G_{c,d}(s) = \widetilde{G}_{c,d}(r)\mathrm{d}s.$$

The results of the lemma now follow by the CMT using the same arguments as those in the proof of Lemma 3.1

□

Proof of Lemma 4.2. The results of the lemma follow by the same arguments as those in the proofs of Lemma 3.1 and 4.1 after observing that $\int_0^1 \cos^2(2\pi ks)\mathrm{d}s = \int_0^1 \sin^2(2\pi ks)\mathrm{d}s = 1/2$. □

Proof of Lemma 4.3. The results of the lemma follow by the same arguments as those in the proofs of Lemma 3.1 and 4.1 after observing that

$$\begin{pmatrix} 1 & \int_0^1 s\mathrm{d}s \\ \int_0^1 s\mathrm{d}s & \int_0^1 s^2\mathrm{d}s \end{pmatrix}^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix}.$$

□

Proof of Proposition 6.1. To simplify the presentation, we prove the result for $p = 1$. For the general case, the proof is similar but requires more a complicated notation. Under H_0, $\widetilde{u}_{t,0} = \widetilde{u}_t$, where $\{\widetilde{u}_t\}$ are the residuals from the projection of $\{u_t\}$ against the components of D_t. Since

$$(1 - \phi_1 L - \phi_2 L^2)\widehat{x}_{t,0} = \widetilde{\varepsilon}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{u}_{t-1},$$

$$\widehat{\rho}_{1,0} - \rho_1 = \frac{\sum \widetilde{u}_{t-1}\varepsilon_t}{\sum \widetilde{u}_{t-1}^2},$$

the estimators of ϕ_1 and ϕ_2 satisfy:

$$\begin{pmatrix} \widehat{\phi}_{1,n} - \phi_1 \\ \widehat{\phi}_{2,n} - \phi_2 \end{pmatrix} = \begin{pmatrix} \sum \widehat{x}_{t-1,0}^2 & \sum \widehat{x}_{t-1,0}\widehat{x}_{t-2,0} \\ \sum \widehat{x}_{t-1,0}\widehat{x}_{t-2,0} & \sum \widehat{x}_{t-2,0}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \widehat{x}_{t-1,0}(\widetilde{\varepsilon}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{u}_{t-1}) \\ \sum \widehat{x}_{t-2,0}(\widetilde{\varepsilon}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{u}_{t-1}) \end{pmatrix},$$

with

$$\begin{pmatrix} \sum \widehat{x}_{t-1,0}(\widetilde{\varepsilon}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{u}_{t-1}) \\ \sum \widehat{x}_{t-2,0}(\widetilde{\varepsilon}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{u}_{t-1}) \end{pmatrix} = \begin{pmatrix} \sum \left(\widehat{x}_{t-1,0} - \frac{\sum \widehat{x}_{s-1,0}\widetilde{u}_{s-1}}{\sum \widetilde{u}_{s-1}^2}\widetilde{u}_{t-1}\right)\widetilde{\varepsilon}_t \\ \sum \left(\widehat{x}_{t-2,0} - \frac{\sum \widehat{x}_{s-2,0}\widetilde{u}_{s-1}}{\sum \widetilde{u}_{s-1}^2}\widetilde{u}_{t-1}\right)\widetilde{\varepsilon}_t \end{pmatrix}. \tag{C.2}$$

The result follows since under the null, $\widehat{\rho}_{1,0} - \rho_1 = O_p(n^{-1/2})$ and $\widehat{x}_{t,0} = \widetilde{x}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{y}_{t-1}$. □

Proof of Proposition 6.2. Similarly to the proof of Proposition 6.1, we prove the result for $p = 1$. For the general case, the proof is analoguous, but requires more a complicated notation. Consider

$\dot{\zeta}_{1,n}$ in (6.5):

$$\dot{\zeta}_{1,n} = \frac{\sum \widetilde{u}_t(\widetilde{x}_t - (\widehat{\rho}_{1,0} - \rho_1)\widetilde{y}_{t-1})}{\sum \widetilde{u}_t^2} = O_p(n),$$

where the second equality holds by the lemmas in Section 4. Next, consider the elements of the matrix Σ_n:

$$\begin{aligned} n^{-4} \sum \dot{x}_{t-1,0}^2 &= n^{-4} \sum (\widehat{x}_{t,0} - \dot{\zeta}_{1,n}\widetilde{u}_{t,0})^2 \\ &= n^{-4} \sum \widetilde{x}_t^2 + o_p(1) \\ &\Rightarrow \sigma_\varepsilon^2 \int \widetilde{J}_{c,d}^2, \end{aligned}$$

where the results in the second and third lines hold again by the lemmas in Section 4. After applying the same arguments to the other elements in Σ_n, the elements of M_n, and the expressions on the right-hand side of (C.2), the result of the proposition follows by the CMT.

□

References

A'HEARN, B., AND U. WOITEK (2001): "More international evidence on the historical properties of business cycles," *Journal of Monetary Economics*, 47(2), 321–346.

AIKMAN, D., A. G. HALDANE, AND B. D. NELSON (2015): "Curbing the credit cycle," *Economic Journal*, 125(585), 1072–1109.

ANDREWS, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

——— (1993): "Exactly median-unbiased estimation of first order autoregressive/unit root models," *Econometrica*, pp. 139–165.

ANDREWS, D. W. K., X. CHENG, AND P. GUGGENBERGER (2020): "Generic results for establishing the asymptotic size of confidence sets and tests," *Journal of Econometrics*, 218, 496–531.

BEAUDRY, P., D. GALIZIA, AND F. PORTIER (2020): "Putting the cycle back into business cycle analysis," *American Economic Review*, 110(1), 1–47.

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

BRY, G., AND C. BOSCHAN (1971): "Cyclical analysis of time series: Selected procedures and computer programs," NBER technical paper.

COCHRANE, J. H. (2017): "Macro-finance," *Review of Finance*, 21(3), 945–985.

COMIN, D., AND M. GERTLER (2006): "Medium-term business cycles," *American Economic Review*, 96(3), 523–551.

DOU, L., AND U. K. MÜLLER (2021): "Generalized local-to-unity models," *Econometrica*, forthcoming.

DREHMANN, M., C. E. BORIO, AND K. TSATSARONIS (2012): "Characterising the financial cycle: don't lose sight of the medium term!," BIS working paper.

ELLIOTT, G., AND J. H. STOCK (2001): "Confidence intervals for autoregressive coefficients near one," *Journal of Econometrics*, 103(1-2), 155–181.

GREGOIR, S. (2006): "Efficient tests for the presence of a pair of complex conjugate unit roots in real time series," *Journal of Econometrics*, 130(1), 45–100.

HANSEN, B. E. (1999): "The grid bootstrap and the autoregressive model," *Review of Economics and Statistics*, 81(4), 594–607.

HARDING, D., AND A. PAGAN (2002): "Dissecting the cycle: a methodological investigation," *Journal of Monetary Economics*, 49(2), 365–381.

HARVEY, A. C. (1985): "Trends and cycles in macroeconomic time series," *Journal of Business & Economic Statistics*, 3(3), 216–227.

MIKUSHEVA, A. (2007): "Uniform inference in autoregressive models," *Econometrica*, 75(5), 1411–1452.

——— (2012): "One-dimensional inference in autoregressive models with the potential presence of a unit root," *Econometrica*, 80(1), 173–212.

NEWEY, W. K., AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariane matrix," *Econometrica*, 55, 703–708.

PHILLIPS, P. C. B. (1987): "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74(3), 535–547.

——— (1988): "Regression theory for near-integrated time series," *Econometrica*, 56(5), 1021–1043.

PHILLIPS, P. C. B., AND S. JIN (2002): "The KPSS test with seasonal dummies," *Economics Letters*, 77(2), 239–243.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20(2), 971–1001.

PRIESTLEY, M. B. (1981): *Spectral Analysis and Time Series*. Academic press.

RÜNSTLER, G., AND M. VLEKKE (2018): "Business, housing, and credit cycles," *Journal of Applied Econometrics*, 33(2), 212–226.

SARGENT, T. J. (1987): *Macroeconomic Theory*, Economic Theory, Econometrics, and Mathematical Economics Series. Emerald Group Publishing Limited, Bingley, UK, 2nd edn.

STOCK, J. H. (1991): "Confidence intervals for the largest autoregressive root in US macroeconomic time series," *Journal of Monetary Economics*, 28(3), 435–459.

STROHSAL, T., C. R. PROAÑO, AND J. WOLTERS (2019): "Characterizing the financial cycle: Evidence from a frequency domain analysis," *Journal of Banking & Finance*, 106, 568–591.