MODELING LONG CYCLES

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 inference on the cyclical properties of time series. We are particularly interested in stochastic cycles arising from 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 the data. We refer to such cycles as “long”.

Long cycles are prevalent in macroeconomic and financial data. In a recent paper, Beaudry et al. (2020) estimated that many variables have cycles of approximately 32–40 quarters, corresponding to 15% and even 20% of their observed samples. Using data from 1960 to 2011, Drehmann

Date: September 14, 2023.

*Corresponding author. Address: Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC Canada, V6T 1L4. Phone: +1 (604) 822-8217, Email: vadim.marmer@ubc.ca.
We thank our co-editor, Serena Ng, the associate editor, and an anonymous referee for their comments that helped us to improve the quality of the paper. We thank Paul Beaudry for numerous discussions and comments, Kevin Song for his contributions and discussions in the early stages of the project, and Jasmine Hao and Hiro Kasahara for their helpful comments and suggestions. Bruno Esposito Acosta provided outstanding research assistance. Vadim Marmer gratefully acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada under grants 435-2017-0329 and 435-2021-0189. This research was enabled in part by the support provided by Compute Canada.
et al. (2012) estimated that the length of the credit cycle is 18 years or approximately 35% of their sample. The first contribution of our paper is to show that statistical inference may be distorted in such cases. We find that substantial distortions occur when the cycle length exceeds 25% of the sample size.\footnote{See Appendix E.}

In our second contribution, we propose a new econometric procedure for 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 confidence intervals for the cycle length that have the following property: their asymptotic coverage probability is correct regardless of the cycle length. Thus, the confidence intervals are 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 a data-generating process (DGP) is acyclical, our procedure is expected to produce empty confidence intervals for the cycle length in large samples. Hence, the procedure can be used to rule out cyclical behavior.

Stochastic cycles arise naturally in the AR(2) model $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$ with complex roots. For example, Sargent (1987) shows that such a process has a peak in its spectrum in the interior of the $[0, \pi]$ range provided that the autoregressive coefficients satisfy $|\phi_1(1-\phi_2)/4\phi_2| < 1$.\footnote{The region with an interior spike in the spectrum is a subset of the region with complex roots (see Sargent, 1987, pages 261–265).} The period of such cycles is determined by the autoregressive coefficients through $2\pi/\omega$, where $\omega = \cos^{-1}(-\phi_1(1-\phi_2)/4\phi_2)$ is the spectrum peak frequency. According to this model, the cycle length would amount to only a negligible fraction of the sample size $n$ in large samples: $2\pi/\omega n \rightarrow 0$. Thus, the long-cycle characteristics of the 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.

To preserve long-cycle characteristics asymptotically, the spectrum peak frequency must be local to zero in the sense that $n\omega_n$ converges to a positive constant as $n \rightarrow \infty$, i.e. the spectrum peak frequency $\omega_n$ “drifts” closer to zero with the sample size. This property can be obtained by modeling the two conjugate complex roots of the AR(2) model as local to one, that is, the real and complex parts of the roots “drift” closer to one with the sample size. Following the literature (e.g. Andrews et al., 2020), we refer to such specifications and coefficients as “drifting”, while specifications with constant parameters independent of $n$ are referred to as “fixed”.\footnote{Drifting specifications can be used to verify the uniform size properties of inferential procedures (Andrews et al., 2020).} The corresponding autoregressive coefficients are too drifting.

Our resulting long-cycle model is a restricted version of the nearly-twice integrated model in Perron and Ng (1996), and the restriction is imposed to generate persistent oscillatory behavior. Although the resulting processes are near I(2), they are stationary in finite samples.
Figure 1 illustrates how drifting specifications allow us to preserve the long-cycle feature asymptotically. It shows the difference between the simulated sample paths of cyclical processes generated using the fixed-coefficient and drifting-coefficient AR(2) 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 the data. However, the long-cycle properties are preserved in the limit by relying on asymptotic approximations with drifting coefficients.

The problem is closely related to that in the literature concerned with inference on the largest autoregressive root (Stock, 1991; Andrews, 1993; Hansen, 1999; Elliott and Stock, 2001; Mikusheva, 2007, 2012; Dou and Müller, 2021). Equivalently, the sum of the autoregressive coefficients. It has been shown in this 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). However, despite many similarities, the existing results for local-to-unity
processes cannot accommodate long cycles because of the presence of two complex conjugate roots. Our third contribution is to develop a novel asymptotic theory for such processes. Our results also lay out the foundation for a new econometric framework that, besides the inference on cyclicity, can also be used to study cointegrating long cycles and phase shifts in macro-financial aggregates.

Our paper is also 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.

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. Long-standing interest in understanding these ups and downs in macro-financial aggregates has led to a vast body of literature on business cycles (see e.g. Bry and Boschan, 1971; Harvey, 1985; AHeearn and Woitek, 2001; Harding and Pagan, 2002; Comin and Gertler, 2006; Drehmann et al., 2012; Aikman et al., 2015; Strohsal et al., 2019; Rüstückler and Vlekke, 2018). Using our methodology, we find that long cycles cannot be ruled out for macroeconomic series, such as the real GDP per capita, the unemployment rates, and the hours per capita. Our results suggest the possibility of cycles that are much longer than those previously reported in the literature. In addition, we find that financial variables such as credit to the nonfinancial 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 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 the 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 remainder 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. Section 5 describes our procedure for constructing confidence intervals for the cycle length. Section 6 presents our empirical results. In Appendix A, we show that the periodogram is an asymptotically biased estimator in the presence of long cycles. In Appendix E, we discuss the size distortion from using conventional $\chi^2$.

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.
critical values for inference. In Appendix F, we show the consistency of the Bayesian Information Criterion (BIC) for the specification of models with long cycles, and then study the finite-sample properties of our proposed inferential procedures using Monte Carlo simulations in ??.

2. A model for long cycles

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

Following Sargent (1987), we consider the class of autoregressive models. Since in this class cyclical behavior requires complex roots that come as conjugate pairs, the AR(2) model with serially uncorrelated errors\(^6\) is a natural starting point. Thus, 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 i.i.d. sequence with a finite variance.

Let \( \lambda_1 \) and \( \lambda_2 \) denote the roots of the characteristic equation \( z^2 - \phi_1 z - \phi_2 = 0 \) for the lag polynomial in (2.1). When \( |\lambda_1| < 1 \) and \( |\lambda_2| < 1 \), \( \{y_t\} \) has the following MA(\( \infty \)) 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 \( \lambda_1, \lambda_2 \) are complex, and consider their polar form representation:

\[
\lambda_1, \lambda_2 = re^{\pm i\theta}, \tag{2.2}
\]

where \( r \) denotes the modulus, \( \theta \) is the argument of the complex roots, and \( i = \sqrt{-1} \) is the imaginary number.

Given the polar coordinate representation for the roots and the MA(\( \infty \)) representation for the process, we can 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\(^7\) or the persistence of the sine wave, and the argument \( \theta \) corresponds to the angular frequency and determines the period of the sine wave.

---

\(^6\)We extend the approach later in the paper to allow for serially correlated errors.

\(^7\)In a 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.
The latter has been used in the literature as a measure of the frequency of the cycle (Harvey, 1985).

The stochastic process \( \{y_t\} \) inherits its oscillatory behavior precisely from this damped periodic sine weighting function. The closer \( r \) is to one, the more persistent \( \{y_t\} \), and the closer \( \theta \) is to zero, the lower the oscillating frequency and the longer the length of cycles in \( \{y_t\} \). 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 the expected cycle length of \( 2\pi/\theta \). With any fixed parameter value \( \theta \), the period as a fraction of the sample size is negligible for large \( n \). Hence, asymptotic approximations assuming fixed values for the argument \( \theta \) can produce distinctly different cyclical behavior from that observed in finite samples.\(^8\) In other words, conventional asymptotics with a fixed complex root argument \( \theta \) 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 the 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 \( n\theta_n \to d \in [0, \infty] \).\(^9\)

We rewrite the AR(2) model in (2.1) as follows:\(^{10}\)

\[
(1 - \phi_{1,n}L - \phi_{2,n}L^2)y_t = u_t,
\]

where \( \phi_{1,n} \) and \( \phi_{2,n} \) are now drifting coefficients that can change with \( n \). We denote the corresponding characteristic roots by \( \lambda_{1,n} \) 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

\[
\lambda_{1,n} = e^{(c+id)/n}, \quad \lambda_{2,n} = e^{(c-id)/n}
\]

where \( c \leq 0 \) and \( d > 0 \) are fixed localization parameters.

**Definition 2.1.** In what follows, we refer to processes satisfying (2.4) and Assumption 2.1 as long-cycle.

Assumption 2.1 excludes positive values of \( c \) as they correspond to explosive roots. Negative values of \( d \) can be excluded as \( d \) and \( -d \) define the same pair of roots. The autoregressive coefficients are related to the characteristic roots through the following equations:

\[
\phi_{1,n} = \lambda_{1,n} + \lambda_{2,n} = 2e^{c/n}\cos(d/n),
\]

\[
\phi_{2,n} = -\lambda_{1,n}\lambda_{2,n} = -e^{2c/n}.
\]

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

---

\(^{8}\)This point is illustrated in Figure 1.

\(^{9}\)The approach can still accommodate conventional asymptotics by allowing \( n\theta_n \to \infty \).

\(^{10}\)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.
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 since one can
replace \( r \) and \( \theta \) with \( r_n \equiv \exp(c/n) \) and \( \theta_n \equiv d/n \) respectively. 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). Since there are two roots near one, the process defined by Assumption 2.1 can be viewed
as near I(2). The model is a restricted version of the nearly-twice integrated model of Perron and
Ng (1996), where the restriction is imposed to generate complex roots and persistent cycles.\(^{11}\)

The localization parameter \( d \) controls the length of the cycle, where long cycles correspond to
values of \( d \) near zero, while \( c \) controls its persistence. Under Assumption 2.1, the expected period
as a fraction of the sample size is given by

\[
\tau \equiv \frac{2\pi}{n\theta_n} = \frac{2\pi}{d}. \tag{2.8}
\]

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

A process with a cyclical factor in its impulse response coefficients may not have visible cyclical
oscillations when the damping effect of \( r^j \) in (2.3) is too strong. An alternative but related
measure of the periodicity of a process can be constructed from its spectral properties (see e.g.
Kaiser and Maravall, 2001). The advantage of this measure is that it also takes into account the
existence properties, unlike those based solely on the argument \( \theta_n \) of the complex roots. Let \( \omega_n^* \)
denote the frequency that maximizes the spectral density of the process in (2.4). As in Sargent
(1987),

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

provided that

\[
\left| \frac{\phi_{1,n}(1 - \phi_{2,n})}{4\phi_{2,n}} \right| < 1. \tag{2.9}
\]

The condition in (2.9) together with (2.6)–(2.7) imply that for sufficiently large sample sizes \( n \),
the spectrum has a peak away from the origin if\(^{12}\)

\[
d > |c|, \tag{2.10}
\]

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

\[
\tau_n^* = \frac{2\pi}{\omega_n^*}. \tag{2.8}
\]

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

\(^{11}\)Perron and Ng (1996) are primarily concerned with unit root testing and propose modifications designed to improve
the size properties of some commonly used tests.

\(^{12}\)The result in (2.10) follows from (2.9) and (2.6)-(2.7) using second-order expansions of \( \cos(d/n) \) and \( \exp(c/n) \).
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 \( \omega^*_n \) 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

\[
\tau_\omega \equiv \frac{2\pi}{\sqrt{d^2 - c^2}}.
\] (2.11)

Unlike the angular frequency-based measure \( \tau_\theta \), the spectrum-based measure \( \tau_\omega \) takes into account the persistence of the cycle as captured by the localization parameter \( c \). Note that larger negative values of \( c < 0 \) produce less persistent cycles. 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 this paper, we consider both \( \tau_\theta \) and \( \tau_\omega \) since both types of cyclicity measures are used in the literature (see e.g. Harvey, 1985; Sargent, 1987; Kaiser and Maravall, 2001). If one is only concerned with the presence of a cyclical factor in the impulse response coefficients, \( \tau_\theta \) is appropriate. However, if in addition one is interested in cyclicity with persistence strong enough to produce a visible spectrum spike away from zero, then more stringent conditions on the autoregressive coefficients are required, and \( \tau_\omega \) is more appropriate. The restriction \( d \geq |c| \), which is required for \( \tau_\omega \) to be defined, ensures that the cycle is persistent enough to produce a peak in the spectrum.

In Appendix A, we show that the periodogram-based estimation approach produces biased estimates of \( \tau_\omega \). Therefore, in the following we develop an inference procedure for \( \tau_\theta \) and \( \tau_\omega \) based on the estimates of the autoregressive coefficients \( \phi_{1,n} \) and \( \phi_{2,n} \). For this purpose, we proceed in two steps. First, we develop a procedure to construct 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 \( \tau_\theta \) and \( \tau_\omega \) measures of the length of a cycle. The theory developed below also allows serially correlated \( \{u_t\} \).

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 to establish the asymptotic distributions of regression-based statistics involving long-cycle time series. It is also necessary to develop robust and asymptotically valid inference about the cyclical properties of a process.
The specification proposed in equations (2.4)–(2.5) is similar 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:

\[
 n^{-1/2} x_{[nr]} = n^{-1/2} \sum_{t=1}^{[nr]} e^{c([nr]-t)/n} u_t \Rightarrow \sigma J_c(r),
\]

where

\[
 J_c(r) = \int_0^r e^{c(r-s)} dW(s),
\]

\( r \in [0,1] \), \( \lfloor x \rfloor \) denotes the largest integer less or equal to \( x \), \( W(\cdot) \) is a standard Brownian motion, \( \sigma^2 \) 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 \( \{u_t\} \).

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

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

In the case of long cycles, a different limiting process arises from that of the local-to-unity case, with cyclicity reflected by the sine function. However, similarly to the results in Phillips (1987), the process can be described as an integral with respect to a Brownian motion and depends on the localization parameters \( c \) and \( d \). We define:

\[
 J_{c,d}(r) = \frac{1}{d} \int_0^r e^{c(r-s)} \sin(d(r-s)) dW(s).
\]

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,

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

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

\[
 y_t = \frac{1}{\lambda_{n,1} - \lambda_{n,2}} \sum_{k=1}^{t} \left( \lambda_{n,1}^{t-k+1} - \lambda_{n,2}^{t-k+1} \right) u_k
\]

\[13^{13}\text{Assumption 3.1 holds, for example, when } \{u_t\} \text{ is a mixing process such that } E(u_t) = 0 \text{ for all } t, \sup_{t} E|u_t|^{\beta + \epsilon} \leq \infty \text{ for some } \beta < 2 \text{ and } \epsilon > 0, \text{ and } \{u_t\} \text{ is } \alpha \text{-mixing of size } -\beta/(\beta - 2) \text{ (Phillips, 1987). Alternatively, it holds when } \{u_t\} \text{ is a linear MA}(\infty) \text{ process satisfying the conditions in Phillips and Solo (1992, Theorem 3.15).} \]
where the second equality follows by Assumption 2.1. By Assumption 3.1 and as in (3.1),

\[ n^{-1/2} \sum_{k=1}^{[nr]} (e^{(c+id)(t-k+1)/n} - e^{(c-id)(t-k+1)/n}) u_k \]

\[ \Rightarrow \sigma \int_0^r (e^{(c+id)(r-s)} - e^{(c-id)(r-s)}) \, dW(s) \]

\[ = 2i\sigma \int_0^r e^{c(r-s)} \sin(d(r-s)) \, dW(s). \]

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

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(\( \infty \)) representation in (2.3): past shocks are weighted by a damped sine wave. Again, the parameters \( c \) and \( d \) control the persistence and frequency of the cycle, respectively. Note also that long-cycle processes require stronger scaling than local-to-unity: \( n^{-3/2} \) 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 \( \hat{\phi}_{1,n} \) and \( \hat{\phi}_{2,n} \) denote the least-squares estimator of (2.4):

\[
\begin{pmatrix}
\hat{\phi}_{1,n} \\
\hat{\phi}_{2,n}
\end{pmatrix} = \left( \sum y_{t-1}^2 \quad \sum y_{t-1} y_{t-2} \quad \sum y_{t-2}^2 \right)^{-1} \left( \sum y_{t-1} y_t \quad \sum y_{t-1} y_{t-2} \quad \sum y_{t-2} y_t \right).
\]

(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 \( \sum y_{t-1} y_{t-2} = \sum y_{t-2}^2 + \) 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):

\[ y_t = (\phi_{1,n} + \phi_{2,n}) y_{t-1} - \phi_{2,n} \Delta y_{t-1} + u_t, \]

(3.4)

where \( \Delta y_{t-1} = y_{t-1} - y_{t-2} \). Since (3.4) is obtained from the original equation through a nonsingular linear transformation of the regressors and parameters, the OLS estimator of \( \phi_{1,n} + \phi_{2,n} \) is given by \( \hat{\phi}_{1,n} + \hat{\phi}_{2,n} \). Moreover, the usual Wald test statistic for testing joint hypotheses about \( \phi_1 \) and \( \phi_2 \) is the same for both regressions. Thus, we have the following:

\[
\begin{pmatrix}
\hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n} \\
\hat{\phi}_{2,n} - \phi_{2,n}
\end{pmatrix}
\]
ergodicity property for the sample moments of long-cycle processes. Parts of the lemma below require the following
\[ \Delta \]
sample distribution of \( y_t \).

Assumption 3.2.

Let \( n \) scaling by \( \Delta \) after scaling, the distribution of \( y_t \) requires the cosine function instead of a sine function. The next proposition shows that in large samples and 3.1 hold. Then, \( n^{-1/2} \Delta y_{nr} \Rightarrow \sigma G_{c,d}(r) \), where the result holds jointly with that in Proposition 3.1.

Note that in contrast to the scaling \( n^{-3/2} \) applied to \( y_t \), its first difference \( \Delta y_{t-1} \) requires scaling by \( n^{-1/2} \). Therefore, the first differences of long-cycle processes have convergence rates of \( O(n^{1/2}) \) tantamount to those of local-to-unity processes. However, due to cyclicality, the large-sample distribution of \( \Delta y_{t-1} \) differs from that 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 \( \sigma_u^2 \equiv \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} E u_t^2 \) be the average variance of \( \{u_t\} \) over time. We assume that \( n^{-1} \sum_{t=1}^{n} u_t^2 \to \sigma_u^2 \).

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

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

(d) \( n^{-2} \sum y_{t-1} u_t \Rightarrow \sigma^2 \int_0^1 J_{c,d}(r)dW(r) \).
(e) \( n^{-1} \sum \Delta y_{t-1} u_t \Rightarrow \sigma^2 \int_1^1 G_{c,d}(r)dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2) \).

Note that in part (e) of the lemma, the limiting distribution of the sample covariance between \( \Delta y_{t-1} \) 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 due to 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 \(\int_0^1 J_{c,d}^2(r)dr\) and \(\int J_{c,d}dW\) to denote \(\int_0^1 J_{c,d}(r)dW(r)\). 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 \(\int_0^1 J_{c,d}(r)G_{c,d}(r)dr\).

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) \(n^2(\hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}) \Rightarrow \frac{\int G_{c,d}^2 \cdot \int J_{c,d}dW - \left( \int G_{c,d}dW + \frac{1}{2}(1 - \sigma_u^2/\sigma^2) \right) \cdot \int J_{c,d}G_{c,d}}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - (\int J_{c,d}G_{c,d})^2}.\)

(b) \(n \left( \frac{\hat{\phi}_{1,n} - \phi_{1,n}}{\hat{\phi}_{2,n} - \phi_{2,n}} \right) \Rightarrow \left( \frac{-1}{1} \right) \times \frac{\int J_{c,d}G_{c,d} \cdot \int J_{c,d}dW - \left( \int G_{c,d}dW + \frac{1}{2}(1 - \sigma_u^2/\sigma^2) \right) \cdot \int J_{c,d}^2}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - (\int J_{c,d}G_{c,d})^2}.\)

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. Furthermore, their convergence rate is \(O_p(n^{-1})\) despite that the process \(\{y_t\}\) is 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 of I(2) processes. Note that the limiting distributions depend on the localization parameters \(c\) and \(d\).

In Section 5 below and following Hansen (1999), we consider a grid-based approach for constructing confidence sets for the autoregressive coefficients. Since Hansen (1999) relies on the \(t\)-statistic for his procedure for the largest autoregressive root, we construct our procedure around the Wald statistic.\(^{14}\)

Consider testing a joint hypothesis \(H_0: \phi_1 = \phi_{1,0}, \phi_2 = \phi_{2,0}\) against \(H_1: \phi_1 \neq \phi_{1,0}\) or \(\phi_2 \neq \phi_{2,0}\). Construction of grid-based confidence sets involves testing a sequence of hypothesis with different null values \(\phi_{1,0}, \phi_{2,0}\) and then collecting those that are not rejected. The usual Wald statistic is given by

\[
W_n(\phi_{1,0}, \phi_{2,0}) \equiv \left( \begin{array}{c} \hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,0} - \phi_{2,0} \\ \hat{\phi}_{2,n} - \phi_{2,0} \end{array} \right) \top V_{n-1}^{-1} \left( \begin{array}{c} \hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,0} - \phi_{2,0} \\ \hat{\phi}_{2,n} - \phi_{2,0} \end{array} \right), \tag{3.7}
\]

\(^{14}\)The approach in Elliott and Stock (2001) can be used to construct alternative statistics.
where
\[ \hat{V}_n = \tilde{\sigma}_n^2 \left( \frac{\sum y_t^2 - \sum y_{t-1} \Delta y_{t-1}}{\sum (\Delta y_{t-1})^2} \right)^{-1}, \]

and \( \tilde{\sigma}_n^2 \) is a consistent estimator of the long-run variance \( \sigma^2 \) constructed using \( \hat{u}_t = y_t - \hat{\phi}_{1,n} y_{t-1} - \hat{\phi}_{2,n} y_{t-2} \), see Newey and West (1987) and Andrews (1991). The infeasible estimator of \( \sigma^2 \) that uses \( u_t \) is constructed as \( \tilde{\sigma}_n^2 = n^{-1} \sum_{t=1}^n u_t^2 + 2 \sum_{h=1}^{m_n} n^{-1} \sum_{t=h+1}^n u_t u_{t-h}, \) where \( m_n = o(n) \) is the lag truncation parameter, and \( w_n(h) \) is a bounded weight function such that \( \lim_{n \to \infty} w_n(h) = 1 \) for all \( h \). The feasible estimator \( \hat{\sigma}_n^2 \) is constructed similarly using the estimated residuals \( \hat{u}_t \) instead of \( u_t \).

We make the following assumption.

**Assumption 3.3.** The infeasible estimator \( \tilde{\sigma}_n^2 \) of the long-run variance \( \sigma^2 \) 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,
\[
W_n(\phi_{1,n}, \phi_{2,n}) \Rightarrow \frac{\int \left\{ J_{c,d} \cdot \left( \int G_{c,d} dW + \frac{1}{2} (1 - \sigma_u^2 / \sigma^2) \right) - G_{c,d} \cdot \int J_{c,d} dW \right\}^2}{\int J_{c,d}^2 \cdot \int G_{c,d}^2 - \left( \int J_{c,d} G_{c,d} \right)^2}.
\]

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 \( \sigma_u^2 / \sigma^2 \), and on the unknown localization parameters \( c \) and \( d \). While the ratio \( \sigma_u^2 / \sigma^2 \) does not play a role when \( \{u_t\} \) are serially uncorrelated and can be estimated consistently otherwise, 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 Appendix E, we discuss the differences between the conventional \( \chi^2_2 \) 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 cycle length exceeds 25% of the sample size and the model includes deterministic components that are introduced in the next section.

4. Extensions to models with deterministic components

For practical applications, it is important to allow the DGP to include nonzero means, trends, and deterministic cycles. We discuss such extensions 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.

15See the proof of Proposition 3.4.
Formally, we assume that the data \{y_t : t = 1, \ldots, n\} are generated according to
\[
(1 - \phi_{1,n}L - \phi_{2,n}L^2)(y_t - D_t) = u_t,
\]
where $D_t$ is non-random, can vary with $t$, and depends on unknown parameters. To control for deterministic regressors, 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 $\mu$.

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

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

The specification\(^{16}\) in (i) allows \{\(y_t\)\} to have a constant over time with a nonzero 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 et al. (2020).\(^{17}\) The $D_t$ component in (ii) generates cosine and sine oscillations at frequencies $2\pi k/n$. The period of such oscillations relative to the sample size is $1/k$, and they can capture very low-frequency cycles in data that are outside the range of interest of the econometrician. For practical purposes, we consider $k = 1, 2, 3$. Inclusion of such components can be viewed as detrending 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 detrending 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 to derive the asymptotic properties and can be absorbed into the unknown coefficient $\xi$. Hence, observationally, the model in (iii) is identical to the model with no adjustment by $n$.

The empirical application in Section 6 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 $\Delta y_{t-1}$:
\[
y_t = (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + (1 - \phi_{1,n}L - \phi_{2,n}L^2)D_t + u_t.
\]

\[\text{(4.2)}\]

4.1. Constant mean

In this section, we consider case (i) of a constant unknown mean. When $D_t = \mu$, equation (4.2) becomes
\[
y_t = \alpha_n + (\phi_{1,n} + \phi_{2,n})y_{t-1} - \phi_{2,n}\Delta y_{t-1} + u_t,
\]
\[\text{(4.3)}\]

\(^{16}\)We do not consider specifications in which deterministic components may dominate the stochastic long-cycle component.

\(^{17}\)We thank Paul Beaudry for pointing our attention to this fact.
where \( \alpha_n \equiv (1 - \phi_{1,n} - \phi_{2,n}) \mu = O(n^{-2}) \).

Let \( \hat{\phi}_{1,n} \) and \( \hat{\phi}_{2,n} \) be the least-squares estimator of the corresponding coefficients in (4.3), and define \( \bar{y}_{t-1} = y_{t-1} - \bar{y} \) and \( \Delta \bar{y}_{t-1} = \Delta y_{t-1} - \Delta \bar{y} \), where \( \bar{y}_n \) and \( \Delta \bar{y}_n \) denote the sample averages of \( y_{t-1} \) and \( \Delta y_{t-1} \) respectively. Then,

\[
\left( \hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n} \right) = \left( \begin{array}{cc} \sum \bar{y}_{t-1}^2 & -\sum \bar{y}_{t-1} \Delta \bar{y}_{t-1} \\ -\sum \bar{y}_{t-1} \Delta \bar{y}_{t-1} & \sum \Delta \bar{y}_{t-1} \end{array} \right)^{-1} \left( \begin{array}{c} \sum \bar{y}_{t-1} u_t \\ -\sum \Delta \bar{y}_{t-1} u_t \end{array} \right),
\]

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 \( \tilde{J}_{c,d}(r) \equiv J_{c,d}(r) - \int_0^r J_{c,d}(s)ds \) and \( \tilde{G}_{c,d}(r) \equiv G_{c,d}(r) - \int_0^r G_{c,d}(s)ds \). The following results hold jointly.

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

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

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

In addition, suppose that Assumption 3.2 holds. The following results hold jointly with (a)–(c).

(d) \( n^{-2} \sum \bar{y}_{t-1} u_t \Rightarrow \sigma^2 \int \tilde{J}_{c,d} dW \).

(e) \( n^{-1} \sum \Delta \bar{y}_{t-1} u_t \Rightarrow \sigma^2 \int \tilde{G}_{c,d} dW + \frac{1}{2} (\sigma^2 - \sigma_0^2) \).

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 \( \tilde{J}_{c,d} \) and \( \tilde{G}_{c,d} \). 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 \( \bar{y}_t \) and \( \Delta \bar{y}_t \) 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

\[
y_t = \alpha_n + \sum_k \left\{ \gamma_{1,k,n} \cos \left( \frac{2\pi kt}{n} \right) + \gamma_{2,k,n} \sin \left( \frac{2\pi kt}{n} \right) \right\} + (\phi_{1,n} + \phi_{2,n}) y_{t-1} - \phi_{2,n} \Delta y_{t-1} + u_t,
\]

where the intercept \( \alpha_n \) 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 \( \hat{\phi}_{1,n} \) and \( \hat{\phi}_{2,n} \) can be obtained by estimating

\[
y_t = (\phi_{1,n} + \phi_{2,n}) \bar{y}_{t-1} - \phi_{2,n} \Delta \bar{y}_{t-1} + u_t,
\]

where \( \bar{y}_{t-1} \) and \( \Delta \bar{y}_{t-1} \) are the residuals from the regressions of \( y_{t-1} \) and \( \Delta y_{t-1} \) 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 \( \bar{y}_{t-1} \), \( \Delta \bar{y}_{t-1} \), and \( u_t \).

\(^{18}\)See Lemma C.1 in the Appendix.
Lemma 4.2. Suppose that \{y_t\} is generated according to equation (4.4), and Assumptions 2.1 and 3.1 hold. Define

\[
\tilde{J}_{c,d}(r) \equiv J_{c,d}(r) - \int_0^1 J_{c,d}(s) \, ds - \sum_k \{\psi_{1k} \cos(2\pi kr) - \psi_{2k} \sin(2\pi kr)\},
\]

\[
\tilde{G}_{c,d}(r) \equiv G_{c,d}(r) - \int_0^1 G_{c,d}(s) \, ds - \sum_k \{\varphi_{1k} \cos(2\pi kr) - \varphi_{2k} \sin(2\pi kr)\},
\]

where

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

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

The following results hold jointly.

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

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

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

In addition, suppose that Assumption 3.2 holds. The following results hold jointly with (a)-(c).

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

(e) \( n^{-1} \sum \varDelta \tilde{y}_{t-1} u_t \Rightarrow \sigma^2 \int \tilde{G}_{c,d} W + \frac{1}{2}(\sigma^2 - \sigma_0^2) \).

Lemma 4.2 is the analogue of Lemma 3.1 for the model with deterministic cycles. The coefficients \( \psi_{1k} \) and \( \psi_{2k} \) 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 interval \([0, 1]\). The coefficients \( \varphi_{1k} \) and \( \varphi_{2k} \) have a similar interpretation with \( J_{c,d} \) replaced by \( G_{c,d} \). Therefore, the processes \( \tilde{J}_{c,d} \) and \( \tilde{G}_{c,d} \) are the residuals from the corresponding continuous-time regressions. They are continuous-time versions of \( \tilde{y}_{t-1} \) and \( \varDelta \tilde{y}_{t-1} \) respectively. Therefore, the results of Lemma 3.1 continue to hold with processes \( J_{c,d} \) and \( G_{c,d} \) replaced by their respective residuals from 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

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

where \( \delta_n \equiv \alpha_n + (\phi_{1,n} + 2\phi_{2,n}) \xi/n = O(n^{-2}) \), and \( \beta_n \equiv \xi(1 - \phi_{1,n} - \phi_{2,n}) = O(n^{-2}) \).\(^{19}\) Similarly to the previous cases, the least-squares estimators \( \hat{\phi}_{1,n} \) and \( \hat{\phi}_{2,n} \) can be obtained by estimating

\[
\hat{y}_t = (\phi_{1,n} + \phi_{2,n}) \tilde{y}_{t-1} - \phi_{2,n} \varDelta \tilde{y}_{t-1} + \tilde{u}_t
\]

where \( \tilde{y}_{t-1} \) and \( \varDelta \tilde{y}_{t-1} \) are now the residuals from the regressions of \( y_{t-1} \) and \( \varDelta y_{t-1} \) respectively against \( t/n \) and a constant.

\(^{19}\)See Lemma C.1 in the Appendix.
Lemma 4.3. Suppose that \( \{y_t\} \) is generated according to equation (4.5), and Assumptions 2.1 and 3.1 hold. Define
\[
\tilde{J}_{c,d}(r) \equiv J_{c,d}(r) - (4 - 6r) \int_0^1 J_{c,d}(s) ds - (12r - 6) \int_0^1 sJ_{c,d}(s) ds,
\]
\[
\tilde{G}_{c,d}(r) \equiv G_{c,d}(r) - (4 - 6r) \int_0^1 G_{c,d}(s) ds - (12r - 6) \int_0^1 sG_{c,d}(s) ds.
\]
The following results hold jointly.
(a) \( n^{-4} \sum \tilde{y}_{t-1}^2 \Rightarrow \sigma^2 \int \tilde{J}_{c,d}^2 \).
(b) \( n^{-2} \sum \tilde{\Delta}y_{t-1} \Rightarrow \sigma^2 \int \tilde{G}_{c,d}^2 \).
(c) \( n^{-3} \sum \tilde{y}_{t-1} \tilde{\Delta}y_{t-1} \Rightarrow \sigma^2 \int \tilde{J}_{c,d} \tilde{G}_{c,d} \).

In addition, suppose that Assumption 3.2 holds. The following results hold jointly with (a)-(c).
(d) \( n^{-2} \sum \tilde{y}_{t-1}u_t \Rightarrow \sigma^2 \int \tilde{J}_{c,d}^2 W \).
(e) \( n^{-1} \sum \tilde{\Delta}y_{t-1}u_t \Rightarrow \sigma^2 \int \tilde{G}_{c,d}^2 W + \frac{1}{2}(\sigma^2 - \sigma_u^2) \).

Lemma 4.3 is the analogue of Lemmas 4.1 and 4.2 for the case of the linear time trend. The processes \( \tilde{J}_{c,d}(r) \) and \( \tilde{G}_{c,d}(r) \) can be interpreted similarly 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 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
\[
n^2(\hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \hat{\phi}_{1,n} - \hat{\phi}_{2,n}) \Rightarrow \frac{\int \tilde{G}_{c,d}^2 \cdot \int \tilde{J}_{c,d} dW - \left( \int \tilde{G}_{c,d} dW + \frac{1}{2}(1 - \sigma_u^2/\sigma^2) \right) \cdot \int \tilde{J}_{c,d} \tilde{G}_{c,d} \cdot \int \tilde{J}_{c,d}^2 - (\int \tilde{J}_{c,d} \tilde{G}_{c,d})^2}{\int \tilde{J}_{c,d}^2 \cdot \int \tilde{G}_{c,d}^2 - (\int \tilde{J}_{c,d} \tilde{G}_{c,d})^2},
\]
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 \( \tilde{J}_{c,d}^2 \) and \( \tilde{G}_{c,d}^2 \).
For all three specifications in Sections 4.1–4.3, the Wald statistic for testing $H_0 : \phi_1 = \phi_{1,0}, \phi_2 = \phi_{2,0}$ against $H_1 : \phi_1 \neq \phi_{1,0}$ or $\phi_2 \neq \phi_{2,0}$ takes the same form as in equation (3.7). However, $\tilde{V}_n$ is now given by

$$\tilde{V}_n = \tilde{\sigma}^2_n \left( \frac{\sum \tilde{y}_{t-1}^2 - \sum \tilde{y}_{t-1} \tilde{\Delta}y_{t-1} \sum (\Delta y_{t-1})^2}{\sum (\Delta y_{t-1})^2} \right)^{-1},$$

with $\tilde{y}_{t-1}$ and $\tilde{\Delta}y_{t-1}$ 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

$$W_n(\phi_{1,n}, \phi_{2,n}) \Rightarrow \int \left\{ \tilde{J}_{c,d} \cdot \left( \int \tilde{G}_{c,d} dW + \frac{1}{2}(1 - \sigma^2_\alpha/\sigma^2) \right) - \tilde{G}_{c,d} \cdot \int \tilde{J}_{c,d} dW \right\}^2$$

$$\frac{\int \tilde{J}_{c,d}^2 \cdot \int \tilde{G}_{c,d}^2 - (\int \tilde{J}_{c,d} \tilde{G}_{c,d})^2}{\int \tilde{J}_{c,d}^2 \cdot \int \tilde{G}_{c,d}^2 - (\int \tilde{J}_{c,d} \tilde{G}_{c,d})^2}$$

(4.7)

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

As in the base case with no deterministic components, the asymptotic null distributions of the Wald statistics are nonstandard and depend on the unknown parameters $c$ and $d$. Differences between the quantiles of these asymptotic distributions and the critical values of the $\chi^2$ distribution are discussed in Appendix E. Compared to the base case, the inclusion of deterministic components may result in more substantial deviations from the $\chi^2$ critical values.

5. Inference for cyclicity

In this section, we propose a procedure for inference on the cycle length in terms of the angular frequency-based measure $\tau_\theta$ and the spectrum-based measure $\tau_\omega$ that were introduced in Section 2. Recall that the two measures can be deduced from the autoregressive coefficients $\phi_{1,n}$ and $\phi_{2,n}$ through the relationships in (2.6)–(2.11). Therefore, we first construct confidence sets for the autoregressive parameters by collecting values $(\phi_1, \phi_2)$ consistent with cyclical behavior and not rejected by data. In the second step, we use projection arguments to construct confidence intervals for $\tau_\theta$ and $\tau_\omega$. By multiplying the values of $\tau_\theta$ and $\tau_\omega$ in the confidence intervals by $n$, the length of the cycle can also be expressed in time units instead of fractions of 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 or far from one. However, if the true DGP is inconsistent with the cyclical behavior, we expect the confidence sets to be empty in large samples. Therefore, the proposed procedure can be used to detect cyclical specifications consistent with the data or to rule out cyclical behavior. However, our procedure is not designed for inference on acyclical 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. The serial correlation in $\{u_t\}$ and the resulting correlation between $(y_{t-1}, \Delta y_{t-1})$ and $u_t$ is reflected by the noncentrality term $0.5(1 - \alpha)$.
\( \sigma_n^2 / \sigma^2 \) 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, that is, under I(0) specifications, the least squares estimators of the autoregressive coefficients are no longer consistent if \( \{ u_t \} \) is serially correlated. Although the process is I(0), due to the inconsistency of the least-squares estimators, the null asymptotic distribution of the Wald statistic is no longer a central \( \chi^2 \). Therefore, to design an inferential procedure that remains valid regardless of the magnitude of the roots, we need to be able to accommodate a potential serial correlation in \( \{ u_t \} \).

We proceed as follows. First, in Section 5.1 we discuss how to construct confidence intervals for \( \tau_\theta \) and \( \tau_\omega \) when \( \{ u_t \} \) are serially uncorrelated. Then, in Section 5.2 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 use the BIC selection procedure to choose the appropriate number of lags \( p \) and the specification for the deterministic part \( D_t \).

5.1. 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 \( 1 - \sigma_n^2 / \sigma^2 = 0 \) 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 \( (\phi_1, \phi_2) \) can be equivalently represented as confidence sets in terms of \( (c, d) \).

After running the regressions of \( y_t \) against \( y_{t-1} \) and \( y_{t-2} \) with different specifications of the deterministic part \( D_t \), the BIC can be used to consistently choose the appropriate specification between the constant mean, the deterministic cycle, or the linear time trend. After selecting \( D_t \), consider the corresponding Wald statistic \( W_n(\Phi_{1,n}(c, d), \Phi_{2,n}(c, d)) \). Let \( W_{1-\alpha}(c, d) \) denote the \( 1 - \alpha \) quantile of the asymptotic distribution in (4.7) with \( 1 - \sigma_n^2 / \sigma^2 = 0 \). I.e., \( W_{1-\alpha}(c, d) \) is the \( 1 - \alpha \) quantile of the distribution of

\[
W(c, d) \sim \frac{\int \left\{ J_{c,d} \cdot \bar{G}_{c,d} \cdot dW - \bar{G}_{c,d} \cdot dW \right\}^2}{\int J_{c,d}^2 \cdot \bar{G}_{c,d}^2 - (\int J_{c,d} \bar{G}_{c,d})^2}, \quad (5.1)
\]
where the definitions of \( \bar{J}_{c,d} \) and \( \bar{G}_{c,d} \) correspond to the selected specification for \( D_t \). The confidence set for \( (c, d) \) can now be constructed by test inversion as
\[
CS_{n,1-\alpha} \equiv \left\{ (c, d) : W_n(\Phi_{1,n}(c,d), \Phi_{2,n}(c,d)) \leq \chi^2_{2,1-\alpha} \right\}.
\]

The confidence set \( CS_{n,1-\alpha} \) is bounded as \( \chi^2_{1-\alpha}(c,d) \rightarrow \chi^2_{2,1-\alpha} \) when \( c \rightarrow -\infty \) or \( d \rightarrow \infty \). 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 \( c \leq c \leq 0 \), and \( 2\pi < d < n\pi \), where the lower bound \( c \) is chosen by the econometrician and the lower bound of \( 2\pi \) is imposed to rule out cycles longer than the sample size when measured by \( \tau_\theta \).

The construction of \( CS_{n,1-\alpha} \) is similar 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 point \( (c, d) \) considered. 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 \( \chi^2_{1-\alpha}(c,d) \). Second, the critical values are computed using the same values \( (c, d) \) as those specified in \( H_0 \). Third, the distribution in (5.1) nests the \( \chi^2_2 \) distribution, which arises under the fixed \( (\phi_1, \phi_2) \) asymptotics, as a limiting case. Note that having the correct size under both drifting and fixed parameter specifications is required for uniform validity (Andrews et al., 2020).

We construct confidence intervals for \( \tau_\theta \) and \( \tau_\omega \) from \( CS_{n,1-\alpha} \) by projection:
\[
CI_{n,1-\alpha}^{\tau_\theta} \equiv \left[ \inf_{d: (c,d) \in CS_{n,1-\alpha}} \frac{2\pi}{d}, \sup_{d: (c,d) \in CS_{n,1-\alpha}} \frac{2\pi}{d} \right],
\]
(5.2)
\[
CI_{n,1-\alpha}^{\tau_\omega} \equiv \left[ \inf_{(c,d) \in CS_{n,1-\alpha}} \frac{2\pi}{\sqrt{d^2 - c^2}}, \sup_{(c,d) \in CS_{n,1-\alpha}} \frac{2\pi}{\sqrt{d^2 - c^2}} \right].
\]
(5.3)

The confidence interval for \( \tau_\theta \) 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 \( \tau_\omega \) can be unbounded if pairs \( (c, d) \) with \( c = d \) are included in \( CS_{n,1-\alpha} \).

5.2. Serially correlated \( \{u_t\} \)

In this section we assume that the innovations process \( \{u_t\} \) is generated as AR(\( p \)):
\[
(1 - \rho_1 L - \ldots - \rho_p L^p)u_t = \varepsilon_t,
\]
(5.4)
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. In this case, \( \{y_t\} \) is AR(\( p + 2 \)), and by running the regressions of \( y_t \) against different specifications of the deterministic part \( D_t \) and \( y_{t-1}, y_{t-1}, \ldots, y_{t-(m+2)} \) for some \( m > p \), one can again use the BIC to consistently estimate the specification for \( D_t \) and the number of lags \( p \) in (5.4).

\footnote{In our empirical application in Section 6, the largest considered value of \( d \) is 678, and \( \xi = -678 \). Note also that the grid only needs to cover the stationary region corresponding to complex roots (see e.g. Sargent, 1987, pages 187–189).}
Let \( \tilde{y}_t \) denote the residuals from the 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 calculated from the values of \( e \) and \( d \). Let
\[
\tilde{u}_{t,0} \equiv (1 - \phi_{1,0}L - \phi_{2,0}L^2)\tilde{y}_t.
\]
Using the null-restricted residuals \( \tilde{u}_{t,0} \), one can estimate the autoregressive coefficients \( \rho_1, \ldots, \rho_p \). Let \( \hat{\rho}_1, \ldots, \hat{\rho}_p \) denote their least-squares estimators. Note that under \( H_0 \), these estimators are consistent. We can now remove the autoregressive part in \( u_t \):
\[
\tilde{x}_{t,0} \equiv (1 - \hat{\rho}_1 L - \ldots - \hat{\rho}_p L^p)\tilde{y}_t.
\]
Thus, to construct the process \( \tilde{x}_{t,0} \), we have filtered the deterministic part \( D_t \) and the serial correlation in \( \{u_t\} \). Note that under the null, the population counterpart of \( \tilde{x}_{t,0} \) satisfies:
\[
\tilde{x}_t \equiv (1 - \rho_1 L - \ldots - \rho_p L^p)\tilde{y}_t = \frac{\tilde{\varepsilon}_t}{1 - \phi_{1,n} L - \phi_{2,n} L^2},
\]
where \( \tilde{\varepsilon}_t \) is the residual from the projection of \( \varepsilon_t \) against the components of \( D_t \).

Now we can use \( \{\tilde{x}_{t,0}\} \) for inference on the cyclical properties of \( \{y_t\} \), however, additional adjustments are required to account for the 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 and under the standard asymptotics with \( \phi_1 \) and \( \phi_2 \) fixed in the stationary range.\(^{21}\)

Let \( \hat{\phi}_{1,n} \) and \( \hat{\phi}_{1,n} \) now denote the least-squares estimators of \( \phi_1 \) and \( \phi_2 \) respectively from the regression of \( \tilde{x}_{t,0} \) against \( \tilde{x}_{t-1,0} \) and \( \tilde{x}_{t-2,0} \):
\[
\tilde{x}_{t,0} = \hat{\phi}_{1,n} \tilde{x}_{t-1,0} + \hat{\phi}_{2,n} \tilde{x}_{t-2,0} + \tilde{\varepsilon}_{t,0},
\]
where \( \tilde{\varepsilon}_{t,0} \) denotes the least-squares residuals. The modified Wald statistic takes the form
\[
W_{n,p}(\phi_{1,0}, \phi_{2,0}) \equiv \frac{1}{\hat{\sigma}_{\varepsilon,n}^2} \begin{pmatrix} \hat{\phi}_{1,n} - \phi_{1,0} \\ \hat{\phi}_{2,n} - \phi_{2,0} \end{pmatrix}^\top M_n \Sigma_n^{-1} M_n \begin{pmatrix} \hat{\phi}_{1,n} - \phi_{1,0} \\ \hat{\phi}_{2,n} - \phi_{2,0} \end{pmatrix},
\]
where \( \hat{\sigma}_{\varepsilon,n}^2 \equiv n^{-1} \sum \tilde{\varepsilon}_{t,0}^2 \), and the matrix \( M_n \) is given by
\[
M_n \equiv \begin{pmatrix} \sum \tilde{x}_{t-1,0}^2 & \sum \tilde{x}_{t-1,0} \tilde{x}_{t-2,0} \\ \sum \tilde{x}_{t-1,0} \tilde{x}_{t-2,0} & \sum \tilde{x}_{t-2,0}^2 \end{pmatrix}.
\]
To construct \( \Sigma_n \), we first define \( \tilde{x}_{t,0} \) and \( \tilde{x}_{t-1,0} \) as the residual from the least-squares regression of \( \tilde{x}_{t,0} \) and \( \tilde{x}_{t-1,0} \) respectively against \( \tilde{u}_{t,0}, \ldots, \tilde{u}_{t-p+1,0} \):
\[
\begin{align*}
\tilde{x}_{t,0} &= \hat{\zeta}_{1,n} \tilde{u}_{t,0} + \ldots + \hat{\zeta}_{p,n} \tilde{u}_{t-p+1,0} + \tilde{x}_{t,0}, \\
\tilde{x}_{t-1,0} &= \hat{\zeta}_{1,n} \tilde{u}_{t,0} + \ldots + \hat{\zeta}_{p,n} \tilde{u}_{t-p+1,0} + \tilde{x}_{t-1,0},
\end{align*}
\] \(^{21}\)Recall that having correct size under both drifting and fixed parameters specifications is required for the uniform validity (Andrews et al., 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 \( \Sigma_n \) 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 \( \chi^2_2 \) distribution.

**Proposition 5.1.** Suppose that \( \{y_t\} \) is generated according to \( (1 - \phi_1 L - \phi_2 L^2) y_t = u_t \) with the coefficients \( \phi_1 \) and \( \phi_2 \) fixed in the stationary range, and \( \{u_t\} \) satisfying (5.4) with the coefficients \( \rho_1, \ldots, \rho_p \) in the stationary range and \( \varepsilon_t \sim \text{i.i.d.} (0, \sigma^2_\varepsilon) \). 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 (5.1).

**Proposition 5.2.** Suppose that \( \{y_t\} \) is generated according to (4.1), where \( \{u_t\} \) satisfies (5.4) with the coefficients \( \rho_1, \ldots, \rho_p \) in the stationary range and \( \varepsilon_t \sim \text{i.i.d.} (0, \sigma^2_\varepsilon) \). Suppose further that Assumption 2.1 holds. Then,

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

where \( \mathcal{W}(c, d) \) is defined in (5.1) with \( \tilde{J}_{c,d} \) and \( \tilde{G}_{c,d} \) defined according to the specification of \( D_t \).

Using the results of Propositions 5.1 and 5.2, one can now construct confidence sets for \( (c, d) \) using the modified Wald statistic as

\[
CS_{n,p,1-\alpha} \equiv \left\{ (c, d) : W_{n,p}(\Phi_{1,n}(c, d), \Phi_{2,n}(c, d)) \leq W_{1-\alpha}(c, d) \right\}.
\]

Similarly to the construction in (5.2) and (5.3), the confidence set \( CS_{n,p,1-\alpha} \) can be projected to construct confidence intervals for \( \tau_\theta \) and \( \tau_\omega \).

6. Cyclical properties of macroeconomic and financial variables

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 a 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 et al. (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 et al., 2015; Comin and Gertler, 2006).
The cyclical properties of the 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 et al., 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 in Appendix A that the periodogram-based estimator is asymptotically biased in the case of long cycles.

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 2 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 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

\[
y_t = y^c_t + D_t,
\]

\[
y^c_t = \phi_{1,n} y^c_{t-1} + \phi_{2,n} y^c_{t-2} + u_t,
\]

where \( y^c_t \) 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 are not seasonally adjusted and, therefore, the 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 select the lag order \( p \) in the AR(\( p \)) specification for \( u_t \). To this end, suppose that \( p \leq M \) for some known positive integer \( M \). As discussed in Section F of the Supplement, one can choose \( p \in \{0, 1, \ldots, M\} \) that minimizes the BIC for the regression of \( y_t \) against \( y_{t-1}, \Delta y_{t-1}, \) the deterministic components, and the second-order differences \( \Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p} \), where \( \Delta^2 y_t \equiv \Delta y_t - \Delta y_{t-1} \).

In our empirical application, most of the time series do not exhibit autocorrelation in \( \{u_t\} \), that is \( p = 0 \), except for the credit-to-GDP ratio, as indicated by the BIC. Moreover, the credit-to-GDP ratio is the only series in which we have included seasonal dummies.

Table 1 presents our results. Note that the last three columns of the table describe the specifications selected by the BIC for \( D_t \) and the order of autocorrelation for \( \{u_t\} \). For example, the hours per capita contains a linear time trend and deterministic cycles of cosine and sine waves with \( k = 1 \), which corresponds to the 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
|                              | $n_{\theta}$ | $n_{\omega}$ | $c$   | $d$   | $n$ | Linear time trend | Deterministic cycles | Autocorr. $u_t$ |
|------------------------------|-------------|-------------|------|------|------|------------------|----------------------|------------------|
| **Macroeconomic variables**  |             |             |      |      |      |                  |                      |                  |
| Real GDP per capita          | (23, 264)   | (25, 512)   | (-108, 0) | (7, 81) | 294 | Yes              | No                  | No               |
| Unemployment rate            | (22, 260)   | (27, 185)   | (-104, -43) | (7, 84) | 290 | No               | No                  | No               |
| Hours per capita             | (18, 64)    | (18, 198)   | (-73, -10) | (29, 105) | 294 | Yes              | $k = 1$              | No               |
| **Financial variables**      |             |             |      |      |      |                  |                      |                  |
| VXO S&P 100 volatility index | —           | —           | —    | —    | —    | —                | —                    | —                |
| Credit risk premium BAA to 10Y | —         | —           | —    | —    | —    | —                | —                    | —                |
| Equity price index           | —           | —           | —    | —    | —    | —                | —                    | —                |
| Private non-financial sector credit % GDP | (48, 245) | (55, 308) | (-41, -13) | (7, 36) | 273 | Yes              | No                  | AR(1)            |
| Home price index             | (42, 120)   | (43, 234)   | (-16, -1) | (7, 20) | 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 duration of the cycles measured by the angular frequency $n_{\theta}$ and the spectrum-maximizing frequency $n_{\omega}$.
3. In columns 1 and 2, the numbers on the top indicate the point estimates of the cycle length. A dashed line "—" is used when the point estimate corresponds to an acyclical process and when the point estimate is unavailable in the case of autocorrelation. Enclosed in parentheses are the minimum and maximum cycle lengths implied by the 95% confidence intervals of $n_{\theta}$ and $n_{\omega}$. When the interval is empty, it is indicated by "∅". All numbers are in quarters.
4. The intercept is included in all specifications.
5. Credit to the private non-financial sector (% GDP) are seasonally adjusted by including seasonal dummies in the regression.

for $n_{\theta}$ and $n_{\omega}$ are constructed by backtracking the corresponding values of $c$, $d$ from the OLS estimates of $\phi_{1,n}$ and $\phi_{2,n}$, which can then be used to compute $T_{\theta}$ and $T_{\omega}$. The point estimates 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 lengths implied by the 95% confidence intervals of $n_{\theta}$ and $n_{\omega}$ are given in parentheses. For completeness, the table also reports the 95% confidence intervals for $c$ and $d$.

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 the real GDP per capita, the unemployment rate, and the hours per capita, contain stochastic cycles with periodicity of at least 5-6 years. Partly due to the projection-based construction of the $n_{\theta}$ and $n_{\omega}$ confidence intervals, the implied range of the cycle length is typically wide. The upper bound confidence estimates usually are large and differ considerably

22. While the OLS estimates of the autoregressive coefficients may correspond to an acyclical process (no complex roots or no spectrum peak) our confidence sets for $\phi_{1,n}$ and $\phi_{2,n}$ may nevertheless include cyclical values.
between the two measures. Nevertheless, our results point to the presence of cyclicity among macroeconomic variables, conforming to the view of endogenous business cycles (Beaudry et al., 2020). On the financial side, we find that credit to the private non-financial sector as a percent of GDP and the 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 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 macro variables and financial variables such as 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 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. However, it is different in nature from the stochastic cycles detected in the other variables.

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 \( CS_{n,1-\alpha} \). The dynamics shown in the figure resonate with the results in Table 1. Financial variables such as the credit to the private non-financial sector and the 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. Furthermore, 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 the credit and the home prices, and about 1.5 to 3 times for the unemployment rate and the hours per capita.

For real GDP per capita, the impulse responses are split into two parts. On the left, the axis corresponds to the set of impulse responses similar to those observed in the unemployment rate and the 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. Although the possibility of having much longer and highly persistent stochastic cycles cannot be rejected, the real GDP per capita shares a similar dynamic to the unemployment rate and the hours per capita.

In sum, our results suggest that business cycles (as marked by the expansions and contractions of 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 business cycles: at least 10 years in duration. Additionally, these financial cycles have more prominent oscillations with amplitudes much larger

\[^{23}\text{For the credit to the private non-financial sector, the standard deviation of the innovation is computed assuming no serial correlation.}\]

\[^{24}\text{Note also that the hours per capita is much more persistent than the unemployment rate and the real GDP per capita.}\]
than those of business cycles. Moreover, we find that equity prices, though commonly included in the characterization of financial cycles, do not exhibit stochastic cycles, and therefore merit separate consideration from credit and home prices. Lastly, our results suggest that asset market fluctuations are a different phenomenon from changes in real economic activities.
Appendix A. The periodogram of long-cycle processes

A.1. Asymptotic properties

Periodogram-based nonparametric estimators are commonly used for inference about the cyclical behavior 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

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

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 \(\omega\) (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 variance \(\sigma_u^2\). In this case, the spectral density of \(\{y_t\}\), denoted \(f_n(\omega)\), satisfies

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

We show in the following 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

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

The result can be extended to allow 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, \(\sigma_u^2\) 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 \(\tau_{\omega}\). The solid line in Figure 3 shows the limiting expression for the expected values of the periodogram of a long-cycle process at near-origin frequencies. Its maximizing frequency is shown by the solid vertical line. The dashed line

\[\text{See the proof of Proposition A.1.}\]
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 $\omega_n$ denotes frequencies.

displays the limit of the true spectral density. The vertical dashed line indicates the true frequency maximizing the spectrum $\sqrt{d^2 - c^2}$ 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 underestimate the spectrum maximizing frequency and, as a result, overestimate the length of the cycle. According to the true spectrum, the cycle length relative 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 length of a cycle.

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

$$f_n(\omega) = \frac{\sigma_n^2}{2\pi} \frac{1}{|1 - \lambda_{n,1} e^{i\omega}|^2 |1 - \lambda_{n,2} e^{i\omega}|^2},$$

(A.3)
By the results in Priestley (1981), equations (6.2.10)–(6.2.11),

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

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 \( f_n(x/n) \).

\[
1 - \lambda_{n,1} e^{i \pi/n} \\
= 1 - e^{c/n} \left( \cos \left( \frac{d + x}{n} \right) + i \sin \left( \frac{d + x}{n} \right) \right) \\
= 1 - \left( 1 + \frac{c}{n} + O \left( \frac{1}{n^2} \right) \right) \\
\times \left( 1 - \frac{1}{2} \left( \frac{d + x}{n} \right)^2 + O \left( \frac{d + x}{n} \right)^4 \right) + i \left( \frac{d + x}{n} \right) + O \left( \frac{d + x}{n}^3 \right) \\
= -\frac{c}{n} + O \left( \frac{(d + x)^2}{n^2} + \frac{(d + x)^4}{n^2} \right) + i \left( \frac{d + x}{n} + O \left( \frac{(d + x)^3}{n^2} \right) \right). 
\]

Hence,

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

Similarly,

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

By (A.3) and (A.6)–(A.7),

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

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

\[ \square \]

**Proof of Corollary A.1.** Since

\[
\int_{t/n}^{(t+1)/n} e^{-i hs} ds = \frac{e^{-iht/n}}{n} (1 + O(n^{-1})),
\]

29
we have
\[ n^{-5/2} \sum_{t=1}^{n} y_t e^{-iht/n} = \frac{1}{1 + O(n^{-1})} \sum_{t=1}^{n} n^{-3/2} y_t \int_{t/n}^{(t+1)/n} e^{-ihs} \, ds \]
\[ = \frac{1}{1 + O(n^{-1})} \sum_{t=1}^{n} \int_{t/n}^{(t+1)/n} n^{-3/2} y_{[ns]} e^{-ihs} \, ds \]
\[ = \frac{1}{1 + O(n^{-1})} \int_{0}^{1} n^{-3/2} y_{[ns]} e^{-ihs} \, ds + O_p(n^{-1}) \]
\[ \Rightarrow \int_{0}^{1} J_{c,d} e^{-ihs} \, ds, \] (A.9)
where the equality in the third line holds because \( y_{[ns]} = y_t \) 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 6

Table 2 in this appendix provides a description of the variables used in the empirical application in Section 6. The description includes the source with exact identifiers for each variable, any transformations applied to the raw data, and the sample periods.

| Source Identifier Construction Sample |
|---------------------------------------|
| Real GDP per capita A939RXOQ048SBEA Natural logarithm 1947Q1 to 2020Q2 |
| Unemployment rate UNRATE Natural logarithm 1948Q1 to 2020Q2 |
| Hours per capita B230RC0Q173SBEA Ratio of non-farm business hours to population, Natural logarithm 1947Q1 to 2020Q2 |
| VXO V10CLS Natural logarithm 1986Q1 to 2020Q3 |
| Credit risk premium BAA10YM — 1953Q2 to 2020Q2 |
| Credit to non-financial sector % GDP QUSPAM770A — 1952Q1 to 2020Q1 |
| Home price index CSUSHPI7A S&P/Case-Shiller U.S. National Home Price Index, Natural logarithm 1987Q1 to 2020Q2 |
| Equity price index 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),
\[ -\frac{\phi_{1,n}(1 - \phi_{2,n})}{4\phi_{2,n}} = 0.5 \cos\left(\frac{d n^{-1}}{c n^{-1}}\right) + 0.5 \cos\left(\frac{-d n^{-1}}{c n^{-1}}\right) \]
Lemma C.2.
The diffusion processes in Lemmas C.1 and C.2 below present auxiliary results needed for the proof of Proposition 3.2 and Lemma 3.1. In particular, Lemma C.2 establishes the properties of the diffusion processes that appear in the limiting expressions for the estimators and the test statistics.

\[ \omega_n^* = \cos^{-1}\left(1 - 0.5(d^2 - c^2)n^{-2} + O(n^{-4})\right) \]

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

\[ \omega_n^* = \cos^{-1}\left(1 - 0.5(d^2 - c^2)n^{-2} + O(n^{-4})\right) = o(1). \quad \text{(C.1)} \]

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

\[
t = \sqrt{2s(1 + O(t^2))} \quad = \sqrt{2s(1 + O(2s(1 + O(t^2))))} \quad = \sqrt{2s} + O(s^{3/2}).
\]

Therefore,

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

\[ \square \]

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

\[
\begin{align*}
(a) \quad & \phi_{1,n} = 2 + 2c + \frac{c^2 - d^2}{n^2} + O(n^{-3}), \\
(b) \quad & -\phi_{2,n} = 1 + \frac{2c}{n} + \frac{c^2 - d^2}{n^2} + O(n^{-3}), \\
(c) \quad & \frac{\phi_{1,n}}{1 - \phi_{2,n}} = 1 - \frac{d^2 + c^2}{2n^2} + O(n^{-3}), \\
(d) \quad & \frac{1}{-\phi_{2,n}} = 1 - \frac{c}{n} - \frac{c^2}{n^2} + O(n^{-3}), \\
(e) \quad & \phi_{1,n} + \phi_{2,n} = 1 - \frac{c^2 + d^2}{n^2} + O(n^{-3}), \\
(f) \quad & - (\phi_{1,n} + \phi_{2,n})\phi_{2,n} = 1 + \frac{2c}{n} + \frac{c^2 - d^2}{n^2} + O(n^{-3}), \\
(g) \quad & 1 - (\phi_{1,n} + \phi_{2,n})^2 = \frac{2(c^2 + d^2)}{n^2} + O(n^{-3}). \\
(h) \quad & \phi_{2,n}^2 = 1 + \frac{4c}{n} + \frac{8c^2}{n^2} + O(n^{-3}).
\end{align*}
\]

**Lemma C.2.** The diffusion processes \( J_{c,d}(\cdot) \), \( K_{c,d}(\cdot) \), and \( G_{c,d}(\cdot) \) have the following properties:

\[
\begin{align*}
(a) \quad & dJ_{c,d}(r) = c \cdot J_{c,d}(r)\,dr + d \cdot K_{c,d}(r)\,dr = G_{c,d}(r)\,dr, \\
(b) \quad & dK_{c,d}(r) = c \cdot J_{c,d}(r)\,dr - d \cdot J_{c,d}(r)\,dr + \frac{1}{2}dW(r), \\
(c) \quad & \int_0^1 e^{2c(r-s)}J_{c,d}(r)\,ds = \frac{1}{c + d^2} \left\{ \int_0^r e^{2c(r-s)}dW(s) - \left(c \cdot J_{c,d}(r) + d \cdot K_{c,d}(r)\right) \right\}, \\
(d) \quad & d\left(J_{c,d}(r) \cdot K_{c,d}(r)\right) = 2c \cdot J_{c,d}(r) \cdot K_{c,d}(r)\,dr + d \cdot (K_{c,d}^2(r) - J_{c,d}^2(r))\,dr + \frac{1}{2}J_{c,d}(r)dW(r), \\
(e) \quad & \int_0^1 G_{c,d}^2(r)\,dr = \left(c^2 + d^2\right) \int_0^1 J_{c,d}^2(r)\,dr + J_{c,d}(1)G_{c,d}(1) - \int_0^1 J_{c,d}(r)dW(r) - c \cdot J_{c,d}^2(1), \\
(f) \quad & J_{c,d}^2(1) = 2 \int_0^1 J_{c,d}(r)G_{c,d}(r)\,dr, \\
(g) \quad & (G_{c,d}^2(1) - 1)/2 = c \int_0^1 G_{c,d}^2(r)\,dr + cd \int_0^1 K_{c,d}(r)G_{c,d}(r)\,dr - d^2 \int_0^1 J_{c,d}(r)G_{c,d}(r)\,dr + \int_0^1 G_{c,d}(r)dW(r).
\end{align*}
\]

31
Proof of Lemma C.2. To prove part (a) and (b), note that by applying trigonometric identities, we have
\[
J_{c,d}(r) = \frac{1}{d} \int_0^r e^{c(r-s)} \{ \sin(dr) \cos(ds) - \cos(dr) \sin(ds) \} \, dW(s),
\]
\[
K_{c,d}(r) = \frac{1}{d} \int_0^r e^{c(r-s)} \{ \cos(dr) \cos(ds) + \sin(dr) \sin(ds) \} \, dW(s).
\]
By applying stochastic differentiation of \( J_{c,d}(r) \) and \( K_{c,d}(r) \),
\[
d \cdot dJ_{c,d}(r) = (c \cdot e^{cr} \sin(dr) + d \cdot e^{cr} \cos(dr)) \int_0^r e^{-cs} \cos(ds) \, dW(s) \cdot dr
\]
\[
+ e^{cr} \sin(dr) e^{-cr} \cos(dr) \, dW(r)
\]
\[
-(c \cdot e^{cr} \cos(dr) - d \cdot e^{cr} \sin(dr)) \int_0^r e^{-cs} \sin(ds) \, dW(s) \cdot dr
\]
\[
-e^{cr} \sin(dr) e^{-cr} \cos(dr) \, dW(r)
\]
\[
= c \int_0^r e^{c(r-s)} \{ \sin(dr) \cos(ds) - \cos(dr) \sin(ds) \} \, dW(s) \cdot dr
\]
\[
+ d \int_0^r e^{c(r-s)} \{ \cos(dr) \cos(ds) + \sin(dr) \sin(ds) \} \, dW(s) \cdot dr,
\]
\[
d \cdot dK_{c,d}(r) = (c \cdot e^{cr} \cos(dr) - d \cdot e^{cr} \sin(dr)) \int_0^r e^{-cs} \cos(ds) \, dW(s) \cdot dr
\]
\[
+ e^{cr} \cos(dr) e^{-cr} \cos(dr) \, dW(r)
\]
\[
+(c \cdot e^{cr} \sin(dr) + d \cdot e^{cr} \cos(dr)) \int_0^r e^{-cs} \sin(ds) \, dW(s) \cdot dr
\]
\[
+ e^{cr} \sin(dr) e^{-cr} \sin(dr) \, dW(r)
\]
\[
= c \int_0^r e^{c(r-s)} \{ \cos(dr) \cos(ds) + \sin(dr) \sin(ds) \} \, dW(s) \cdot dr
\]
\[
+ d \int_0^r e^{c(r-s)} \{ \sin(dr) \cos(ds) - \cos(dr) \sin(ds) \} \, dW(s) \cdot dr
\]
\[
+ dW(r).
\]
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:
\[
\int_0^r e^{2c(r-s)} J_{c,d}(s) \, ds = \frac{d}{c} \int_0^r e^{2c(r-s)} K_{c,d}(s) \, ds - \frac{1}{c} J_{c,d}(r),
\]
\[
\int_0^r e^{2c(r-s)} K_{c,d}(s) \, ds = \frac{1}{cd} \int_0^r e^{2c(r-s)} \, dW(s) - \frac{d}{c} \int_0^r e^{2c(r-s)} J_{c,d}(s) \, ds - \frac{1}{c} K_{c,d}(r).
\]
With some algebraic manipulations, we obtain part (c).
By Itô’s lemma,
\[
d\left( J_{c,d}(r) \cdot K_{c,d}(r) \right) = dJ_{c,d}(r) \cdot K_{c,d}(r) + J_{c,d}(r) \cdot dK_{c,d}(r).
\]
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
\[ d \cdot J_{c,d}(1)K_{c,d}(1) = 2cd \int_0^1 J_{c,d}(r)K_{c,d}(r)dr + d^2 \int_0^1 (K_{c,d}^2(r) - J_{c,d}^2(r))dr + \int_0^1 J_{c,d}(r)dW(r), \]
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):
\[ \int_0^1 G_{c,d}^2(r)dr = c^2 \int_0^1 J_{c,d}^2(r)dr + 2cd \int_0^1 J_{c,d}(r)K_{c,d}(r)dr + d^2 \int_0^1 K_{c,d}^2(r)dr \]
\[ = (c^2 + d^2) \int_0^1 J_{c,d}^2(r)dr + d \cdot J_{c,d}(1)K_{c,d}(1) - \int_0^1 J_{c,d}(r)dW(r) \]
\[ = (c^2 + d^2) \int_0^1 J_{c,d}(r)dr + \cdot J_{c,d}(1)G_{c,d}(1) - c \cdot J_{c,d}(1) - \int_0^1 J_{c,d}(r)dW(r). \]
To prove (f) and (g), we use stochastic differentiation of \( J_{c,d}^2(r) \) and \( G_{c,d}^2(r) \), respectively:
\[ dJ_{c,d}^2(r) = 2J_{c,d}(r)dJ_{c,d}(r) + 2d \cdot J_{c,d}(r)K_{c,d}(r)dr, \]
\[ dG_{c,d}^2(r) = 2G_{c,d}(r)dG_{c,d}(r) + (dG_{c,d}(r))^2 \]
\[ = 2G_{c,d}(r)(c \cdot dJ_{c,d}(r) + d \cdot dK_{c,d}(r)) + dr \]
\[ = 2c \cdot G_{c,d}(r)G_{c,d}(r)dr + 2cd \cdot G_{c,d}(r)K_{c,d}(r)dr - 2d^2 \cdot G_{c,d}(r)J_{c,d}(r) \]
\[ + 2G_{c,d}(r)dW(r) + dr. \]
The results in (f) and (g) follow 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),
\[ y_t = \left(2 + \frac{2c}{n} + \frac{c^2 - d^2}{n^2} + O(n^{-3})\right)y_{t-1} - \left(1 + \frac{2c}{n} + \frac{2c^2}{n^2} + O(n^{-3})\right)y_{t-2} + u_t, \]
\[ \Delta y_t = \left(1 + \frac{2c}{n}\right)\Delta y_{t-1} + \left(\frac{c^2 - d^2}{n^2} + O(n^{-3})\right)y_{t-1} - \left(\frac{2c^2}{n^2} + O(n^{-3})\right)y_{t-2} + u_t \]
\[ = \sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} u_j + \left(\frac{c^2 - d^2}{n^2} + O(n^{-3})\right)\sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} y_{j-1} \]
\[ - \left(\frac{2c^2}{n^2} + O(n^{-3})\right)\sum_{j=0}^{t} \left(1 + \frac{2c}{n}\right)^{t-j} y_{j-2}. \]

Define \( S_n(r) = \sum_{i=1}^{\lfloor nr \rfloor} u_i \). We have:
\[ \Delta y_{\lfloor nr \rfloor} = \sum_{j=0}^{\lfloor nr \rfloor} \left(1 + \frac{2c}{n}\right)^{\lfloor nr \rfloor-j} \int_{j-1}^{j} dS_n(s) \]

33
\[
\begin{aligned}
&\frac{c^2 - d^2}{n} + O(n^{-2}) \sum_{j=0}^{\lfloor n \rfloor - 1} \left( 1 + \frac{2c}{n} \right) y_{\lfloor n \rfloor - j}^{|\lfloor n \rfloor| - 1} ds \\
&- \left( \frac{2c^2}{n} + O(n^{-2}) \right) \sum_{j=0}^{\lfloor n \rfloor - 1} \left( 1 + \frac{2c}{n} \right) y_{\lfloor n \rfloor - 2j}^{|\lfloor n \rfloor| - 1} ds.
\end{aligned}
\]

By the CMT and Proposition 3.1,
\[
n^{-1/2} \Delta y_{n} \Rightarrow \sigma \int_0^r e^{2c(r-s)} dW(s) - \sigma(c^2 + d^2) \int_0^r e^{2c(r-s)} J_{c,d}(s) ds,
\]
where the result in the last line follows by Lemma C.2(c). The result of the proposition now follows by the definition of \( G_{c,d}(r) \) in (3.6).

\begin{proof}[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:
\[
\sum y_t^2 = (\phi_1 + \phi_2)^2 \sum y_{t-1}^2 + \phi_2^2 \sum (\Delta y_{t-1})^2 + \sum u_t^2 - 2(\phi_1 + \phi_2) \phi_2 \sum y_{t-1} \Delta y_{t-1} + 2(\phi_1 + \phi_2) \sum y_{t-1} y_{t} - 2\phi_2 \sum \Delta y_{t-1} u_t.
\]

After rearranging and applying the results of Lemmas C.1 and above identities, we have:
\[
\sum y_{t-1} u_t = \frac{c^2 + d^2}{n^2} \sum y_{t-1} + y_{n} \Delta y_{n} - \sum (\Delta y_{t-1})^2 - \frac{2c}{n} \sum y_{t-1} \Delta y_{t-1} + O_p(n).
\]

By the results in parts (a)–(c) of the lemma, and using the shortened notation as explained on page 12,
\[
n^{-2} \sum y_{t-1} u_t \Rightarrow \sigma^2 \left( (c^2 + d^2) \int J_{c,d}^2 + J_{c,d}(1) G_{c,d}(1) - \int G_{c,d}^2 - 2c \int J_{c,d} G_{c,d} \right)
\]
where the result in the last line is by Lemma C.2(e) and (f).

To prove part (e), we follow the same steps as in part (d) using (3.4) to obtain
\[
\begin{aligned}
\frac{1}{n} \sum \Delta y_{t-1} u_t &= \frac{c^2 + d^2}{n^3} \sum y_{t-1} \Delta y_{t-1} - \frac{2c}{n^2} \sum (\Delta y_{t-1})^2 - \frac{1}{2n} \sum u_t^2 + \frac{1}{2n} (\Delta y_{n})^2 + O(n^{-1}) \\
&\Rightarrow \sigma^2(c^2 + d^2) \int J_{c,d} G_{c,d} - 2\sigma^2 \int G_{c,d}^2 - \frac{1}{2} \sigma_u^2 + \frac{1}{2} \sigma^2 G_{c,d}^2(1)
\end{aligned}
\]
\[
= \sigma^2 \int G_{c,d} dW + \frac{1}{2}(\sigma^2 - \sigma_u^2),
\]
where the equality in the last line is by part (g) of Lemma C.2 and the definition of \( G_{c,d} \).\end{proof}
Proof of Proposition 3.3. By (3.5),

\[
\left(\frac{\hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}}{\hat{\phi}_{2,n} - \phi_{2,n}} \right) = \frac{1}{\sum y_{t-1}^2 \sum (\Delta y_{t-1})^2 - (\sum y_{t-1} \Delta y_{t-1})^2} \times \left( \frac{\sum (\Delta y_{t-1})^2}{\sum y_{t-1} \Delta y_{t-1}} \frac{\sum y_{t-1} \Delta y_{t-1}}{\sum y_{t-1}^2} \right) \left( - \sum y_{t-1} u_t \right).
\]

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

\[
n(\hat{\phi}_{1,n} - \phi_{1,n}) = n(\hat{\phi}_{1,n} + \hat{\phi}_{2,n} - \phi_{1,n} - \phi_{2,n}) - n(\hat{\phi}_{2,n} - \phi_{2,n}) = O_p(n^{-1}) - n(\hat{\phi}_{2,n} - \phi_{2,n}),
\]

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

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

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

Denote \( \phi_{12,n} \equiv \phi_{1,n} + \phi_{2,n} \) and \( \hat{\phi}_{12,n} \equiv \hat{\phi}_{1,n} + \hat{\phi}_{2,n} \). We have:

\[
\hat{\sigma}^2_{u,n} = \frac{1}{n} \sum u_t^2 - (\hat{\phi}_{12,n} - \phi_{12,n}) \frac{2}{n} \sum y_{t-1} u_t + (\hat{\phi}_{2,n} - \phi_{2,n}) \frac{2}{n} \sum \Delta y_{t-1} u_t + \frac{1}{n} \sum \left( (\hat{\phi}_{12,n} - \phi_{12,n}) y_{t-1} + (\hat{\phi}_{2,n} - \phi_{2,n}) \Delta y_{t-1} \right)^2 \rightarrow_p \sigma^2_{u,n},
\]

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,

\[
\hat{\sigma}^2_n = \hat{\sigma}^2_{u,n} + O_p(m_n/n),
\]

and the result follows by Assumption 3.3. \( \square \)

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

\[
n^{-3/2} \bar{y}_n / \sigma \Rightarrow \int_0^1 J_{c,d}(s) ds, \quad n^{-1/2} \Delta \bar{y}_n / \sigma \Rightarrow \int_0^1 G_{c,d}(s) ds.
\]
For the general case, the proof is analogous, but requires more a complicated notation. Consider Proof of Proposition 5.2. The result follows since under the null, hence,
\[ n^{-3/2}(\bar{y}_{[nr]} - \bar{y}_n)/\sigma \Rightarrow J_{c,d}(s) - \int_0^1 J_{c,d}(r) \, ds = \widetilde{J}_{c,d}(r), \]
\[ n^{-1/2}(\Delta y_{[nr]} - \Delta \bar{y}_n)/\sigma \Rightarrow G_{c,d}(r) - \int_0^1 G_{c,d}(s) \, ds = \widetilde{G}_{c,d}(r) ds. \]

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

\[ \Box \]

**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 \( \int_0^1 \cos^2(2\pi ks) \, ds = \int_0^1 \sin^2(2\pi ks) \, ds = 1/2. \)

**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
\[ \left( \begin{array}{c} \int_0^1 s \, ds \\ \int_0^1 s^2 \, ds \end{array} \right)^{-1} = \left( \begin{array}{cc} 4 & -6 \\ -6 & 12 \end{array} \right). \]

**Proof of Proposition 5.1.** 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 \( H_0, \bar{u}_{t,0} = \tilde{u}_t, \) where \( \{\tilde{u}_t\} \) are the residuals from the projection of \( \{u_t\} \) against the components of \( D_t. \) Since
\[ (1 - \phi_1 L - \phi_2 L^2)\tilde{x}_{t,0} = \tilde{e}_t - (\tilde{\rho}_{1,0} - \rho_1)\tilde{u}_{t-1}, \]
\[ \tilde{\rho}_{1,0} - \rho_1 = \sum \tilde{u}_{t-1} \tilde{e}_t, \]
the estimators of \( \phi_1 \) and \( \phi_2 \) satisfy:
\[ \left( \begin{array}{c} \hat{\phi}_{1,n} - \phi_1 \\ \hat{\phi}_{2,n} - \phi_2 \end{array} \right) = \left( \begin{array}{cc} \sum \tilde{x}_{t-1,0} & \sum \tilde{x}_{t-1,0} \tilde{x}_{t-2,0} \\ \sum \tilde{x}_{t-1,0} \tilde{x}_{t-2,0} & \sum \tilde{x}_{t-2,0} \tilde{x}_{t-2,0} \end{array} \right)^{-1} \left( \begin{array}{c} \sum \tilde{x}_{t-1,0} (\tilde{e}_t - (\tilde{\rho}_{1,0} - \rho_1)\tilde{u}_{t-1}) \\ \sum \tilde{x}_{t-2,0} (\tilde{e}_t - (\tilde{\rho}_{1,0} - \rho_1)\tilde{u}_{t-1}) \end{array} \right), \]
with
\[ \left( \begin{array}{c} \sum \tilde{x}_{t-1,0} (\tilde{e}_t - (\tilde{\rho}_{1,0} - \rho_1)\tilde{u}_{t-1}) \\ \sum \tilde{x}_{t-2,0} (\tilde{e}_t - (\tilde{\rho}_{1,0} - \rho_1)\tilde{u}_{t-1}) \end{array} \right) = \left( \begin{array}{c} \sum (\tilde{x}_{t-1,0} - \sum \tilde{x}_{t-1,0} \tilde{u}_{t-1} \tilde{u}_{t-1}) \tilde{e}_t \\ \sum (\tilde{x}_{t-2,0} - \sum \tilde{x}_{t-2,0} \tilde{u}_{t-1} \tilde{u}_{t-1}) \tilde{e}_t \end{array} \right). \]
The result follows since under the null, \( \tilde{\rho}_{1,0} - \rho_1 = O_p(n^{-1/2}) \) and \( \tilde{x}_{t,0} = \tilde{x}_t - (\tilde{\rho}_{1,0} - \rho_1)\bar{y}_{t-1}. \)

**Proof of Proposition 5.2.** Similarly to the proof of Proposition 5.1, we prove the result for \( p = 1. \) For the general case, the proof is analogous, but requires more a complicated notation. Consider \( \hat{\zeta}_{1,n} \) in (5.5):
\[ \hat{\zeta}_{1,n} = \sum \tilde{u}_t (\tilde{x}_t - (\tilde{\rho}_{1,0} - \rho_1)\bar{y}_{t-1}) \sum \tilde{u}_t^2 = O_p(n), \]

36
where the second equality holds by the lemmas in Section 4. Next, consider the elements of the matrix $\Sigma_n$:

$$n^{-4}\sum_{t} \tilde{x}_{t-1,0}^2 = n^{-4}\sum_{t} (\tilde{x}_{t,0} - \tilde{\zeta}_{1,n} \tilde{u}_{t,0})^2$$

$$= n^{-4}\sum_{t} \tilde{x}_{t}^2 + o_p(1)$$

$$\Rightarrow \sigma^2 \int J_{c,d}^2,$$

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 $\Sigma_n$, the elements of $M_n$, and the expressions on the right-hand side of (C.3), the result of the proposition follows by the CMT.

□

Appendix D. Long cycles and the local-to-unity model

In this appendix, we illustrate that one can discriminate between long cycles and the local-to-unity DGP, and in sufficiently large samples the confidence sets for $c, d$ and the local-to-unity parameter can not be non-empty at the same time.\footnote{We thank the associate editor for bringing up this issue.}

For the purpose of the illustration, we assume below that $\{u_t\}$ is serially uncorrelated, $\sigma = 1$ and is known. Suppose $\{x_t\}$ is a local-to-unity process as in Section 3 and the result in (3.1) holds. Consider the $t$-statistic $t_n(a_n) = (\hat{a}_n - a_n)(\sum_t x_{t-1})^{1/2}$, where $\hat{a}_n = \sum_t x_{t} x_{t-1}/\sum_t x_{t-1}$. In this case, $t_n(a_n) = O_p(1)$ as $t_n(a_n) \Rightarrow c J_{c,d}^2(\int J_{c,d}^2 G_{c,d} \int J_{c,d}^2)^{1/2}$ (Phillips, 1987).

Next, consider the above $t$-statistic computed using a long-cycle process $\{y_t\}$ satisfying the assumptions of Proposition 3.1:

$$t_n(e^{b/n}) = \left( \sum_t y_{t-1}^2 \right)^{1/2} \left( \frac{\sum_t y_t y_{t-1}}{\sum_t y_{t-1}^2} - e^{b/n} \right)$$

$$= \left( \sum_t y_{t-1}^2 \right)^{1/2} \left( \phi_{1,n} + \phi_{2,n} - e^{b/n} \right) + e^{2c/n} \sum_t y_{t-1} \Delta y_{t-1} \left( \sum_t y_{t-1}^2 \right)^{1/2} + \sum_t y_{t-1} u_t \left( \sum_t y_{t-1}^2 \right)^{1/2},$$

where $\phi_{1,n} + \phi_{2,n} - e^{b/n} = 2e^{c/n} \cos(d/n) - e^{2c/n} - e^{b/n} = -b/n + o(1/n)$. By Lemma 3.1(a),(b), and (d),

$$n^{-1}t_n(e^{b/n}) \Rightarrow -b \left( \int J_{c,d}^2 \right)^{1/2} + \int J_{c,d} G_{c,d} \left( \int J_{c,d}^2 \right)^{1/2}.$$

Thus, if data are generated from a local-to-unity model, the $t$-statistic is $O_p(1)$. However, if data are generated as a long-cycle process, the $t$-statistic is $O_p(n)$. Hence, in the case of a long-cycle process, a grid-based confidence set for a local-to-unity parameter such as in Hansen (1999) would be empty in sufficiently large samples. Note that in the case of local-to-unity DGPs, our confidence set for $c, d$ would be empty in sufficiently large samples as discussed above.
Appendix E. Size distortions from conventional critical values

In the AR(2) model with complex roots, the cycle frequency can be directly inferred from the autoregressive coefficients. In this section, we discuss size distortions one may see when using conventional $\chi^2$ critical values in place of the quantiles of the distributions derived in equation (4.7) in Section 4. For the purpose of this exercise, we assume that there is no serial correlation in $\{u_t\}$ and, as a result, the noncentrality term $0.5(1 - \sigma^2_u/\sigma^2)$ is zero. Note that if $\{u_t\}$ are serially correlated, one can expect more substantial size distortions due to the presence of the noncentrality 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 critical value $\chi^2_{\alpha}$, where $\chi^2_{\alpha}$ is the $1 - \alpha$ quantile of the $\chi^2$ distribution. The asymptotic size of this test is $1 - F_{c,d}(\chi^2_{\alpha})$, and size distortion are given by the difference between the asymptotic size and the nominal size $\alpha$.

Next, we examine the extent of the 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 3 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 $G_{c,d}$ processes 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 the value of $d$, the lower the oscillation frequency and the longer the cycle length relative to the sample size.

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

For example, in the 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 critical value $\chi^2_{\alpha}$ over-rejects the null by 0.066. As we move down the rows and across the columns of Table 3, the process becomes less persistent and with a shorter cycle period, and as a result the size distortions become negligible. Note, however, that the relationship can be non-monotone.

Although in the case of the constant mean model, the size distortions are relatively minor, they are much more prominent in the case of the specifications with deterministic cycles and linear trends. In particular, the usage of conventional $\chi^2$ 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. Although $d = 5$ corresponds to very long cycles as measured by $\tau_\theta$, the size distortions remain substantial even for shorter cycles. For example, in the specifications 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 $\tau_\omega = 0.56$. 
Table 3. 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.

| $c$ | 5   | 15  | 25  | 35  | 45  | 55  |
|-----|-----|-----|-----|-----|-----|-----|
| constant mean |
| -1  | 0.116 | 0.059 | 0.070 | 0.054 | 0.051 | 0.050 |
| -5  | 0.110 | 0.060 | 0.052 | 0.062 | 0.054 | 0.050 |
| -10 | 0.089 | 0.064 | 0.053 | 0.049 | 0.056 | 0.051 |
| -15 | 0.077 | 0.064 | 0.055 | 0.051 | 0.049 | 0.061 |
| -20 | 0.070 | 0.062 | 0.056 | 0.051 | 0.049 | 0.049 |
| -70 | 0.051 | 0.051 | 0.050 | 0.050 | 0.049 | 0.048 |
| deterministic cycle: $k = 1$ |
| -1  | 0.746 | 0.103 | 0.071 | 0.054 | 0.051 | 0.050 |
| -5  | 0.641 | 0.161 | 0.070 | 0.063 | 0.054 | 0.050 |
| -10 | 0.441 | 0.192 | 0.090 | 0.059 | 0.056 | 0.051 |
| -15 | 0.317 | 0.186 | 0.104 | 0.069 | 0.054 | 0.059 |
| -20 | 0.242 | 0.170 | 0.109 | 0.075 | 0.058 | 0.050 |
| -130| 0.052 | 0.052 | 0.051 | 0.051 | 0.050 | 0.048 |
| linear time trend |
| -1  | 0.317 | 0.071 | 0.072 | 0.053 | 0.051 | 0.050 |
| -5  | 0.257 | 0.089 | 0.058 | 0.062 | 0.054 | 0.051 |
| -10 | 0.188 | 0.101 | 0.066 | 0.053 | 0.056 | 0.051 |
| -15 | 0.146 | 0.101 | 0.072 | 0.057 | 0.051 | 0.060 |
| -20 | 0.121 | 0.096 | 0.074 | 0.060 | 0.052 | 0.049 |
| -100| 0.053 | 0.052 | 0.052 | 0.052 | 0.050 | 0.049 |

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

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 $c \to -\infty$.

In conclusion, 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 $\chi^2_2$ distribution for the values of $c, d$ sufficiently close to zero. Such specifications correspond to longer cycles. For the values of $c, d$ sufficiently far from zero, which correspond to shorter cycles, the distributions converge to the $\chi^2_2$ distribution. In particular, across all specifications of the deterministic component, the size distortions from using $\chi^2_2$ critical values become negligible for $\tau_\theta \leq 0.14$. However, when the length of the cycle as measured by $\tau_\theta$ exceeds 14% of the sample size, the use of $\chi^2_2$ critical values leads to size distortions. The distortions are typically more pronounced for longer cycles.
Appendix F. Consistency of the BIC under long-cycles specifications

There is a rich literature in econometrics and statistics on the use of information criteria (IC) for the specification of time series models. For example, Hannan and Quinn (1979) show the consistency of the BIC for the selection of lags in I(0) autoregressive models; Ng and Perron (2001) propose an IC procedure for the selection of an autoregressive lag truncation parameter in unit root tests; Marcellino et al. (2006) consider the AIC and BIC for data-dependent lag order choices in a large-scale forecasting study with macroeconomic time series data. In this section, we show that the BIC can be used to specify long-cycle models consistently.

First, consider a model with serially correlated \( \{ u_t \} \) but without deterministic components. That is, data are generated according to (2.4) and (5.4):

\[
(1 - \phi_{1,n}L - \phi_{2,n}L^2)y_t = u_t = (1 - \rho_1L - \ldots - \rho_pL^p)^{-1}\varepsilon_t,
\]

where \( \varepsilon_t \) are i.i.d. \( (0, \sigma_e^2) \), and the roots of the polynomial \( 1 - \rho_1L - \ldots - \rho_pL^p \) are real and bounded away from unity. We assume that the number of lags \( p \) is unknown but bounded from above by a known integer constant \( M > 0 \). Under (F.1), \( \{ y_t \} \) is an AR(\( p + 2 \)) process:

\[
y_t = \zeta_{1,n} y_{t-1} + \ldots + \zeta_{p+2,n} y_{t-p-2} + \varepsilon_t,
\]

where assuming \( p > 0 \), \( \zeta_{1,n} \equiv \phi_{1,n} + \rho_1, \zeta_{2,n} = \phi_{2,n} + \rho_2 - \phi_{1,n}\rho_1, \zeta_{p+1,n} \equiv -\phi_{1,n}\rho_p - \phi_{2,n}\rho_{p-1}, \zeta_{p+2,n} \equiv -\phi_{2,n}\rho_p, \) and for \( 3 \leq j \leq p, \zeta_{j,n} \equiv \rho_j - \phi_{1,n}\rho_{j-1} - \phi_{2,n}\rho_{j-2} \). Similarly to (3.4), we transform the model into first and second differences to avoid singularities in the asymptotic distributions. Recall that \( \Delta y_t \equiv y_t - y_{t-1} \), and let \( \Delta^2 y_t \equiv \Delta y_t - \Delta y_{t-1} \). We have:

\[
y_t = \zeta_{1,n} y_{t-1} + \ldots + (\zeta_{p+1,n} + \zeta_{p+2,n}) y_{t-p-1} - \zeta_{p+2,n} \Delta y_{t-p-1} + \varepsilon_t
\]

\[
= \left( \sum_{j=1}^{p+2} \zeta_{j,n} \right) y_{t-1} - \left( \sum_{j=2}^{p+2} \zeta_{j,n} \right) \Delta y_{t-1} - \ldots - (\zeta_{p+1,n} + \zeta_{p+2,n}) \Delta y_{t-p} - \zeta_{p+2,n} \Delta y_{t-p-1} + \varepsilon_t
\]

\[
= \left( \sum_{j=1}^{p+2} \zeta_{j,n} \right) y_{t-1} - \left( \sum_{j=2}^{p+2} \zeta_{j,n} \right) \Delta y_{t-1} - \ldots - (\zeta_{p+1,n} + 2\zeta_{p+2,n}) \Delta y_{t-p} + \zeta_{p+2,n} \Delta^2 y_{t-p} + \varepsilon_t
\]

\[
= \nu_{1,n} y_{t-1} + \nu_{2,n} \Delta y_{t-1} + \nu_{1,n} \Delta^2 y_{t-1} + \ldots + \nu_{p,n} \Delta^2 y_{t-p} + \varepsilon_t,
\]

where \( \nu_{1,n} \equiv \sum_{j=1}^{p+2} j \zeta_{j,n}, \nu_{2,n} \equiv -\sum_{j=2}^{p+2} (j - 1) \zeta_{j,n}, \nu_{1,n} \equiv \sum_{j=1}^{p} j \zeta_{j+2,n}, \nu_{2,n} \equiv \sum_{j=2}^{p} (j - 1) \zeta_{j+2,n}, \) etc. Thus, for \( 1 \leq l \leq p, \nu_{l,n} \equiv \sum_{j=1}^{p} (j - l + 1) \zeta_{j+2,n} \).

The results of Lemma 3.1 are extended below to cover expressions with \( \Delta^2 y_{t-l} \) terms.

**Lemma F.1.** Suppose that \( \{ y_t \} \) are generated according to Assumption 2.1 and equation (F.1), \( \{ \varepsilon_t \} \) are i.i.d. and have a zero mean and a finite variance, and the roots of the polynomial \( 1 - \rho_1L - \ldots - \rho_pL^p \) are real and bounded away from one. Then,

(a) \( \Delta^2 y_t = u_t + O_p(n^{-1/2}) \).

Moreover, for \( l, s \geq 1 \),

(b) \( \sum y_{t-1} \Delta^2 y_{t-l} = O_p(n^2) \).

(c) \( \sum \Delta y_{t-1} \Delta^2 y_{t-l} = O_p(n) \).
(d) \( \sum \Delta^2 y_{t-i} \Delta^2 y_{t-s} = O_p(n) \).
(e) \( \sum \Delta^2 y_{t-i} \varepsilon_t = O_p(n^{1/2}) \).

Combining (F.2) with (4.1), we obtain the following specification:

\[
y_{t-1} = \alpha_n + \beta_n \cdot (t/n) + \sum_{k \in \kappa} \left\{ \gamma_{1k,n} \cos \left( \frac{2\pi kt}{n} \right) + \gamma_{2k,n} \sin \left( \frac{2\pi kt}{n} \right) \right\} \\
+ \varrho_{1,n} y_{t-1} + \varrho_{2,n} \Delta y_{t-1} + \nu_{1,n} \Delta^2 y_{t-1} + \ldots + \nu_{p,n} \Delta^2 y_{t-p} + \varepsilon_t \\
= \alpha_n + \beta_n \cdot (t/n) + C_n(t/n) \gamma_n + \varrho_{1,n} y_{t-1} + \varrho_{2,n} \Delta y_{t-1} + z_{t-1,p} \nu_n + \varepsilon_t,
\]

where \( \kappa \) is a set of positive integers that determine the deterministic cyclical components, \( C_n(x) = (\cos(2\pi kx), \sin(2\pi kx) : k \in \kappa) \) and \( \gamma_n = (\gamma_{1k,n}, \gamma_{2k,n} : k \in \kappa) \) is the vector of corresponding coefficients, \( z_{t-1,p} = (\Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p}) \) and \( \nu_n = (\nu_{1,n}, \ldots, \nu_{p,n}) \) is the vector of corresponding coefficients. We use the BIC to determine \( \kappa, p \), and whether the linear trend should be included in the regression. Note that while \( \kappa \) is unknown, we assume that its elements are selected from a small known set of positive integers. For example, we consider \( k = 1, 2, 3 \) in the main paper application section and allow for \( \kappa = \emptyset \).

To determine \( \kappa, p \), and whether the linear trend should be included, the econometrician computes the BIC:

\[
BIC_n(T, \kappa, p) \equiv n \log(n^{-1} SSR_n(T, \kappa, p)) + (T + |\kappa| + p) \log(n),
\]

where \( T \in \{0, 1\} \) with zero indicating no linear trend and one indicating that the linear trend is included, \( \kappa \) is a set of integers for \( C_n(t/n) \) considered in a specification, \( |\kappa| \) is the number of elements in \( \kappa \), \( p \in \{0, 1, \ldots, M\} \) indicates the lag length in a specification: \( z_{t-1,p} = (\Delta^2 y_{t-1}, \ldots, \Delta^2 y_{t-p}) \),

\[
SSR_n(1, \kappa, p) \equiv \sum \left( y_t - \hat{\alpha}_n - \hat{\beta}_n \cdot (t/n) - C_n(t/n) \gamma_n - \hat{\varrho}_{1,n} y_{t-1} - \hat{\varrho}_{2,n} \Delta y_{t-1} - z_{t-1,p} \nu_n \right)^2,
\]

\[
SSR_n(0, \kappa, p) \equiv \sum \left( y_t - \hat{\alpha}_n - C_n(t/n) \gamma_n - \hat{\varrho}_{1,n} y_{t-1} - \hat{\varrho}_{2,n} \Delta y_{t-1} - z_{t-1,p} \nu_n \right)^2,
\]

where \( \hat{\beta}_n, \hat{\gamma}_n, \hat{\nu}_n, \hat{\varrho}_{1,n}, \hat{\varrho}_{2,n} \) denote the OLS estimators of the corresponding coefficients in a specification. Let \( \mathcal{K} \) denote the set of possible values for \( \kappa \). The BIC estimators of \( \kappa, p \), and the choice for the linear trend are given by

\[
(\hat{T}_n, \hat{\kappa}_n, \hat{\nu}_n) \equiv \arg \min_{T \in \{0, 1\}, \kappa \in \mathcal{K}, p \leq M} BIC_n(T, \kappa, p).
\]

The consistency of \( \hat{T}_n, \hat{\kappa}_n, \hat{\nu}_n \) follows by standard arguments from the following lemma. Let \( T_0, \kappa_0, p_0 \) denote the true values of \( T, \kappa, p \) respectively.

**Proposition F.1.** Suppose that \( \{y_t\} \) is generated according to (F.3) and Assumption 2.1, and \( \{\varepsilon_t\} \) are i.i.d. with a zero mean and a finite variance \( \sigma^2 \).

(a) If \( T_0 \leq T, \kappa_0 \subseteq \kappa \), and \( p_0 \leq p \), then \( SSR_n(T, \kappa, p) = \sum \varepsilon^2_t + O_p(1) \).

(b) If \( T_0 > T \), or \( p_0 > p \), or \( \kappa_0 \cap \kappa \neq \emptyset \), then \( n^{-1} SSR_n(T, \kappa, p) \to_d \sigma^2 + B \), where the random variable \( B \) is positive with probability one.
Suppose that the model is over-specified: \( T_0 \leq T, \kappa_0 \subseteq \kappa, \) and \( p_0 \leq p \) with at least one relation held strictly. In that case,
\[
P \left( BIC_n(T_0, \kappa_0, p_0) > BIC_n(T, \kappa, p) \right)
\]
\[
= P \left( n \log \left( \frac{n^{-1} SSR_n(T_0, \kappa_0, p_0)}{n^{-1} SSR_n(T, \kappa, p)} \right) > (T - T_0 + |\kappa| - |\kappa_0| + p - p_0) \log(n) \right)
\]
\[
= P \left( n \log \left( \frac{\sigma^2 + O_p(n^{-1})}{\sigma^2 + O_p(n^{-1})} \right) > (T - T_0 + |\kappa| - |\kappa_0| + p - p_0) \log(n) \right)
\]
\[
= P \left( n \log \left( 1 + O_p(n^{-1}) \right) > (T - T_0 + |\kappa| - |\kappa_0| + p - p_0) \log(n) \right)
\]
\[
= P \left( O_p(1) > (T - T_0 + |\kappa| - |\kappa_0| + p - p_0) \log(n) \right)
\]
\[\rightarrow 0,\]
where the second equality holds by Proposition F.1(a).

If the model is under-specified, i.e. there are missing terms, by Proposition F.1(b),
\[
P \left( n^{-1} BIC_n(T_0, \kappa_0, p_0) > n^{-1} BIC_n(T, \kappa, p) \right) \rightarrow_d P \left( \log \left( \frac{\sigma^2}{\sigma^2 + B} \right) > 0 \right) = 0.
\]
This establishes the consistency of the BIC selection procedure for the deterministic components and the lag length \( p \) in the AR(\( p \)) specification for \( \{u_t\} \).

Appendix G. Monte Carlo simulations

In this section, we use Monte Carlo simulations to study the effect of using the BIC to select a specification, that is, the time trend, deterministic cycles, and autoregressive lags for \( \{u_t\} \), on the size and power of our inferential procedure. For this purpose, we generate data from models as in (4.1) and (5.4):
\[
y_t = (1 - \phi_{1,n}L - \phi_{2,n}L^2)(1 - \rho_1 L) \left( y_t - \mu - \beta(t/n) - \eta_1 \cos(2\pi kt/n) - \eta_2 \sin(2\pi kt/n) \right) = \varepsilon_t,
\]
where \( \{\varepsilon_t\} \) are i.i.d. \( N(0, 0.1^2) \).

The coefficients \( \phi_{1,n} \) and \( \phi_{2,n} \) are generated according to (2.6) and (2.7) respectively. We consider two sets of values for \( c \) and \( d \): i) \( c = -10, d = 26 \), and ii) \( c = -150, d = 80 \). The first set corresponds to \( \tau_\omega = 0.26 \) and \( \tau_\theta = 0.24 \). The latter corresponds to a shorter cycle with \( \tau_\theta = 0.08 \) (\( \tau_\omega \) is not defined in that case).

For the autoregressive specifications for \( \{u_t\} \), we consider \( \rho = 0 \) (no serial correlation) and \( \rho = 0.5 \) (serial correlation with an AR(1) specification).

For the deterministic component, we set \( \mu = 0 \), however, the regressions include the intercept, and the critical values are corrected accordingly. For the linear trend, we use \( \beta = 0 \) (no trend)
Table 4. Simulated size of the nominal 0.05-size test of $H_0 : c = c_0, d = d_0$ vs. $H_1 : c \neq c_0$ or $d \neq d_0$ for different choices of $c$ and $d$, different specifications of the deterministic component (trend, cycle) and serial correlation in $\{u_t\}$, using the true or BIC-selected specifications.

|                  | $c = -10, d = 26$ | $c = -150, d = 80$ |
|------------------|-------------------|-------------------|
|                  | true model | BIC-selected | true model | BIC-selected |
| no trend         | 0.046       | 0.054       | 0.048       | 0.049       |
| no deterministic cycles no serial correlation | 0.044       | 0.058       | 0.071       | 0.057       |
| no trend         | 0.040       | 0.052       | 0.053       | 0.055       |
| deterministic cycles no serial correlation | 0.031       | 0.016       | 0.032       | 0.015       |
| trend            | 0.042       | 0.025       | 0.069       | 0.027       |
| deterministic cycles serial correlation | 0.042       | 0.025       | 0.069       | 0.027       |

and $\beta = \sigma$ (trend), where $\sigma$ is the square root of the long-run variance of the corresponding specification. For the deterministic cyclical component, we use $k = 1$, and $\eta_1 = \eta_2 = 0$ (no deterministic cycles) or $\eta_1 = \eta_2 = \sigma$ (deterministic cycles).

We generate $n = 500$ observations and test the hypothesis $H_0 : c = c_0, d = d_0$ vs. $H_1 : c \neq c_0$ or $d \neq d_0$ over a grid with $c_0 = -200, -199, \ldots, 0$ and $d_0 = 10, 12, \ldots, 100$. The deterministic components and the serial correlation in $\{u_t\}$ are removed according to the selected specification as described in Section 5, and the corresponding critical values are used.

When selecting a model using the BIC, we search across the specifications with or without linear trend, no deterministic cycles or $k = 1, 2$, AR($p$) with $p = 0, 1, 2$, and their combinations.

Table 4 reports the simulated rejection probabilities based on 1,000 Monte Carlo repetitions for each specification. Columns 1 and 3 are computed using the true specifications of the deterministic component and the serial correlation in $\{u_t\}$, and the tests in columns 2 and 4 are performed using the specifications selected by the BIC. According to our results, for the DGPs with $c = -10, d = 25$, our procedure provides reliable control of the size with all deviations above the nominal 0.05 within the expected simulation errors. Only minor size distortions are observed in two cases for the DGPs with $c = -150, d = 80$. The distortions disappear with larger sample.
sizes (e.g., \( n = 1,000 \)).\(^{27}\) We conclude that the BIC selection procedure performs well and as expected, given the results of Section F.

Figure 4 reports the simulated rejection probabilities at the points of the \( c_0, d_0 \)-grid and various specifications when data are generated with \( c = -10, d = 26 \). The tests in the figure are conducted according to the true specifications. The power quickly reaches a rejection probability of one for sufficiently distant from the truth \( c_0, d_0 \) combinations. Compared to the DGPs without serial correlation in \( \{ u_t \} \), there is some loss of power when \( \{ u_t \} \) are serially correlated.

Figure 5 reports the power of the tests performed using the specifications chosen by the BIC. The results are extremely close qualitatively and numerically to those where the true specifications were used.

Figure 6 reports the corresponding results for DGPs with \( c = -150, d = 80 \), where the tests are performed using the true specifications. The tests are substantially less powerful than those when \( c = -10, d = 26 \). This can be explained by the fact that under \( c = -150, d = 80 \), the DGPs are near \( I(0) \) and the estimators of the \( \phi 's \) lose their superconsistency properties. One can see that the deviations of \( c_0 \) from the true \( c \) matter more for power than the deviations of \( d_0 \) from \( d \): the test has little power to detect incorrect \( d 's \) when the true \( c \) is far from zero. However, note that, as we discuss in Section 2, for such values of \( c \), data are unlikely to display any oscillating behavior. The power increases faster with distance between \( c_0 \) and \( c \) and reaches a rejection probability of one for sufficiently distant values of \( c_0 \). Similarly to the more persistent specifications, serial correlation in \( \{ u_t \} \) reduces the power of the test.

Figure 7 reports the corresponding results for the tests with the BIC-selected specifications for the deterministic component and the serial correlation in \( \{ u_t \} \). Again, one can see that the results are numerically and qualitatively close to those under the true specifications.

From Table 4, and the comparisons between Figures 4 and 5, and 6 and 7, we conclude that the BIC is capable of consistently selecting the correct specifications without affecting the size or power properties of our tests.

Appendix H. Proofs of the results in Appendix F

**Proof of Lemma F.1.** For part (a), using the expansion in \( (C.2) \) in Proposition 3.2,

\[
\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}
\]

\[
= u_t + \frac{2c}{n} \sum_{j=0}^{t-1} \left( 1 + \frac{2c}{n} \right)^{t-1-j} u_j + \left( \frac{c^2 - d^2}{n^2} + O(n^{-3}) \right) \left( y_{t-1} + \frac{2c}{n} \sum_{j=0}^{t-1} \left( 1 + \frac{2c}{n} \right)^{t-1-j} y_{j-1} \right) - \left( \frac{2c^2}{n^2} + O(n^{-3}) \right) \left( y_{t-2} + \frac{2c}{n} \sum_{j=0}^{t-1} \left( 1 + \frac{2c}{n} \right)^{t-1-j} y_{j-2} \right)
\]

\[
= u_t + \left( \frac{c^2 - d^2}{n^2} + O(n^{-3}) \right) y_{t-1} - \left( \frac{2c^2}{n^2} + O(n^{-3}) \right) y_{t-2} + \frac{2c}{n} \Delta y_{t-1}
\]

\(^{27}\)The results for larger sample sizes are not reported here as they are similar to those with \( n = 500 \) except for the two discussed cases with \( c = -150, d = 80 \).
Figure 4. Heatmaps of the simulated rejection probabilities of the nominal 0.05-
size test of $H_0: c = c_0, d = d_0$ vs. $H_1: c \neq c_0$ or $d \neq d_0$ for different specifications
of the deterministic component (trend, cycle) and serial correlation in \{u_t\}, and
different values of $c_0, d_0$. Data are generated with $c = -10, d = 26$ (marked by \times
in the graphs). The tests are performed using the true specifications.
Figure 5. Heatmaps of the simulated rejection probabilities of the nominal 0.05-size test of $H_0: c = c_0, d = d_0$ vs. $H_1: c \neq c_0$ or $d \neq d_0$ for different specifications of the deterministic component (trend, cycle) and serial correlation in $\{u_t\}$, and different values of $c_0, d_0$. Data are generated with $c = -10, d = 26$ (marked by $\times$ in the graphs). The tests are performed using the BIC-selected specifications.
Figure 6. Heatmaps of the simulated rejection probabilities of the nominal 0.05-
size test of $H_0 : c = c_0, d = d_0$ vs. $H_1 : c \neq c_0$ or $d \neq d_0$ for different specifications
of the deterministic component (trend, cycle) and serial correlation in $\{u_t\}$, and
different values of $c_0, d_0$. Data are generated with $c = -150, d = 80$ (marked by ×
in the graphs). The tests are performed using the true specifications.
Figure 7. Heatmaps of the simulated rejection probabilities of the nominal 0.05-
size test of $H_0 : c = c_0, d = d_0$ vs. $H_1 : c \neq c_0$ or $d \neq d_0$ for different specifications of the deterministic component (trend, cycle) and serial correlation in $\{u_t\}$, and different values of $c_0, d_0$. Data are generated with $c = -150, d = 80$ (marked by × in the graphs). The tests are performed using the BIC-selected specifications.
\[ \sum \varepsilon_t^2 = u_t + O_p(n^{-1/2}), \]

where the second and third equalities hold by the expansion in (C.2), and the equality in the last line holds by Propositions 3.1 and 3.2. The remaining lemma results follow those in Part (a), Lemma 3.1, and because \( \{u_t\} \) is an I(0) and stationary AR\( (p) \) process.

\[ \nabla \]

**Proof of Proposition F.1.** In part (a), suppose that \( \kappa \neq \emptyset \) and \( p > 0 \).

\[ BIC_n(1, \kappa, p) - \sum \varepsilon_t^2 = -v_n A_n^{-1} v_n \]

where

\[ v_n \equiv \left( \sum \varepsilon_t, \sum \varepsilon_t \cdot (t/n), \sum \varepsilon_t C_\kappa(t/n)', \sum \varepsilon_t y_{t-1}, \sum \varepsilon_t \Delta y_{t-1}, \sum \varepsilon_t z_{t-1,p}' \right)', \]

and the \( A_n \) matrix is given by

\[
\begin{pmatrix}
\sum t/n & \sum C_\kappa(t/n)' & \sum y_{t-1} & \sum \Delta y_{t-1} & \sum z_{t-1,p}' \\
\sum (t/n)^2 & \sum (t/n)C_\kappa(t/n)' & \sum (t/n)y_{t-1} & \sum (t/n)\Delta y_{t-1} & \sum (t/n)z_{t-1,p}' \\
\sum C_\kappa(t/n)C_\kappa(t/n)' & \sum C_\kappa(t/n)y_{t-1} & \sum C_\kappa(t/n)\Delta y_{t-1} & \sum C_\kappa(t/n)z_{t-1,p}' \\
\sum y_{t-1}^2 & \sum y_{t-1} \Delta y_{t-1} & \sum (\Delta y_{t-1})^2 & \sum \Delta y_{t-1} z_{t-1,p}' & \sum z_{t-1,p} z_{t-1,p}' \\
\end{pmatrix}
\]

Define a diagonal scaling matrix

\[
H_n = \begin{pmatrix}
n^{-1/2} & 0 & \ldots & 0 \\
0 & n^{-1/2} & \ldots & 0 \\
n^{-1/2} I_{|\kappa|} & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 \\
n^{-1} & 0 & \ldots & 0 \\
n^{-1} I_p
\end{pmatrix}
\]

By Lemmas 3.1 and F.1,

\[
H_n v_n = \left( n^{-1/2} \sum \varepsilon_t, n^{-1/2} \sum \varepsilon_t \cdot (t/n), n^{-1/2} \sum \varepsilon_t C_\kappa(t/n)', n^{-2} \sum \varepsilon_t y_{t-1}, n^{-1} \sum \varepsilon_t \Delta y_{t-1}, n^{-2} \sum \varepsilon_t z_{t-1,p}' \right)'
\]

\[ = O_p(1). \]
By the same arguments, $H_n A_n H_n$ is given by

$$
\begin{pmatrix}
1 & \sum_{n}^{t/n} & \sum_{n}^{C(t/n)'} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{(t/n)c(t/n)'} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{c(t/n)c(t/n)'} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{C(t/n)'} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\frac{t/n}{n} & \sum_{n}^{y_{t-1}} & \sum_{n}^{\Delta y_{t-1}} & \sum_{n}^{z_{t-1,p}}
\end{pmatrix}
$$

where $\sigma^2 = \frac{\sigma^2}{1 - \rho_1 - \ldots - \rho_{p_0}}$ is the long-run variance of $\{u_t\}$, and $\Gamma_p$ is the $p \times p$ matrix of the variances and autocovariances of orders up to $p$ of $\{u_t\}$: $\Gamma_p \equiv E z_{t-1,p} z_{t-1,p}'$. Therefore, $(H_n A_n H_n)^{-1} = O_p(1)$, and the result in part (a) follows for the case of $T = 1, \kappa \neq \emptyset$, and $p > 0$. The cases where $T_0 = T = 0, \kappa_0 = \kappa = \emptyset$, and $p = p_0 = 0$ can be handled similarly.

For part (b), suppose that the terms in $\kappa \subset \kappa_0$ are omitted from $\kappa$. Let $C_{\kappa^*}(t/n)$ denote the vector of the corresponding deterministic cyclical components. Furthermore, suppose that $p_0 > p$, and let $z_{t-1,p} = (\Delta^2 y_{t-p_1}, \ldots, \Delta^2 y_{t-p_0})'$ denote the vector of $\Delta^2 y_t$ terms omitted from $z_{t-1,p}$. Let $\gamma_n^*$ and $\nu_n^*$ denote the coefficients of $C_{\kappa^*}(t/n)$ and $z_{t-1,p}^*$ respectively. Now,

$$
n^{-1} SSR_n(1, \kappa, p) = n^{-1} \sum (\varepsilon_t + C_{\kappa^*}(t/n)'^* + z_{t-1,p}^* p_{n}^*)^2 - n^{-1} \nu_n^* A_n^{-1} \nu_n^*,
$$

where $A_n$ is as defined in the proof of part (a), and

$$
\nu_n^* \equiv \left( \begin{array}{c}
\sum \varepsilon_t + \gamma_n^* \sum C_{\kappa^*}(t/n) + \nu_n^* \sum z_{t-1,p}^* \\
\sum \varepsilon_t \cdot (t/n) + \gamma_n^* \sum C_{\kappa^*}(t/n) \cdot (t/n) + \nu_n^* \sum z_{t-1,p}^* \cdot (t/n) \\
\sum \varepsilon_t C_{\kappa^*}(t/n) + \gamma_n^* \sum C_{\kappa^*}(t/n) C_{\kappa^*}(t/n) + \nu_n^* \sum z_{t-1,p}^* C_{\kappa^*}(t/n) \\
\sum \varepsilon_t y_{t-1} + \gamma_n^* \sum C_{\kappa^*}(t/n) y_{t-1} + \nu_n^* \sum z_{t-1,p}^* y_{t-1} \\
\sum \varepsilon_t \Delta y_{t-1} + \gamma_n^* \sum C_{\kappa^*}(t/n) \Delta y_{t-1} + \nu_n^* \sum z_{t-1,p}^* \Delta y_{t-1} \\
\sum \varepsilon_t z_{t-1,p} + \gamma_n^* \sum C_{\kappa^*}(t/n) z_{t-1,p} + \nu_n^* \sum z_{t-1,p}^* z_{t-1,p} \\
\end{array} \right).
$$

Next,

$$
n^{-1/2} H_n \nu_n^* = \left( \begin{array}{c}
n^{-1} \sum \varepsilon_t + \gamma_n^* n^{-1} \sum \sum C_{\kappa^*}(t/n) + \nu_n^* n^{-1} \sum z_{t-1,p}^* \\
n^{-1} \sum \varepsilon_t \cdot (t/n) + \gamma_n^* n^{-1} \sum C_{\kappa^*}(t/n) \cdot (t/n) + \nu_n^* n^{-1} \sum \sum z_{t-1,p}^* \cdot (t/n) \\
n^{-1} \sum \varepsilon_t C_{\kappa^*}(t/n) + \gamma_n^* n^{-1} \sum C_{\kappa^*}(t/n) C_{\kappa^*}(t/n) + \nu_n^* n^{-1} \sum \sum z_{t-1,p}^* C_{\kappa^*}(t/n) \\
n^{-1/2} \sum \varepsilon_t y_{t-1} + \gamma_n^* n^{-1/2} \sum C_{\kappa^*}(t/n) y_{t-1} + \nu_n^* n^{-1/2} \sum \sum z_{t-1,p}^* y_{t-1} \\
n^{-1/2} \sum \varepsilon_t \Delta y_{t-1} + \gamma_n^* n^{-1/2} \sum C_{\kappa^*}(t/n) \Delta y_{t-1} + \nu_n^* n^{-1/2} \sum \sum z_{t-1,p}^* \Delta y_{t-1} \\
n^{-1} \sum \varepsilon_t z_{t-1,p} + \gamma_n^* n^{-1} \sum C_{\kappa^*}(t/n) z_{t-1,p} + \nu_n^* n^{-1} \sum \sum z_{t-1,p}^* z_{t-1,p} \\
\end{array} \right).\quad (50)
$$
\[
\lim_{d} \left( \begin{array}{c}
\int_{0}^{1} \gamma^* C_{\kappa^*}(s) ds \\
\int_{0}^{1} s \gamma^* C_{\kappa^*}(s) ds \\
\int_{0}^{1} \gamma^* C_{\kappa^*}(s) C_{\kappa}(s) ds \\
\sigma \int_{0}^{1} \gamma^* C_{\kappa^*}(s) J_{s,d}(s) ds \\
\sigma \int_{0}^{1} \gamma^* C_{\kappa^*}(s) G_{s,d}(s) ds \\
\Gamma_{p,*}\nu^*
\end{array} \right),
\]

where \( \gamma^* \equiv \lim_{n \to \infty} \gamma_n = (1 - \rho_1 - \ldots - \rho_p)(\eta_{1k}, \eta_{2k} : k \in \kappa^*'), \Gamma_{p,*} \equiv E_{z_{t-1,p} \zeta_{t-1,p}}^* \), and \( \nu^* \equiv \lim_{n \to \infty} \nu^*_n \). It follows that

\[
n^{-1} SSR_n(1, \kappa, p) \to_d \sigma^2 + \int_{0}^{1} \left( \gamma^* C_{\kappa^*}(s) - X(s) \zeta \right)^2 ds + \nu^* \left( \Gamma_{*} - \Gamma_{p,*} \Gamma_{p}^{-1} \Gamma_{p,*} \right) \nu^*,
\]

where

\[
X(s) \equiv \left( 1, s, C_{\kappa}(s), J_{c,d}(s), G_{c,d}(s) \right)', \zeta \equiv \left( \int_{0}^{1} X(s)X(s)'ds \right)^{-1} \int_{0}^{1} X(s) \gamma^* C_{\kappa^*}(s) ds,
\]

\[
\Gamma_{*} \equiv E_{z_{t-1,p} \zeta_{t-1,p}}^*.
\]

Similar arguments can be used to show that when \( T_0 = 1 \),

\[
n^{-1} SSR_n(0, \kappa, p) \to_d \sigma^2 + \int_{0}^{1} \left( \gamma^* C_{\kappa^*}(s) + \beta \cdot s - X(s) \zeta \right)^2 ds + \nu^* \left( \Gamma_{*} - \Gamma_{p,*} \Gamma_{p}^{-1} \Gamma_{p,*} \right) \nu^*,
\]

where now

\[
X(s) \equiv \left( 1, C_{\kappa}(s), J_{c,d}(s), G_{c,d}(s) \right)', \zeta \equiv \left( \int_{0}^{1} X(s)X(s)'ds \right)^{-1} \int_{0}^{1} X(s) \left( \gamma^* C_{\kappa^*}(s) + \beta \cdot s \right) ds,
\]

and \( \beta \equiv \lim_{n \to \infty} \beta_n \). \[ \square \]

References

A’Hearn, B., Woitek, U., 2001. More international evidence on the historical properties of business cycles. Journal of Monetary Economics 47 (2), 321–346.

Aikman, D., Haldane, A. G., Nelson, B. D., 2015. Curbing the credit cycle. Economic Journal 125 (585), 1072–1109.

Andrews, D. W. K., May 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D. W. K., 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica , 139–165.

Andrews, D. W. K., Cheng, X., Guggenberger, P., 2020. Generic results for establishing the asymptotic size of confidence sets and tests. Journal of Econometrics 218, 496–531.
Beaudry, P., Galizia, D., Portier, F., 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, 962–983.

Bry, G., Boschan, C., 1971. Cyclic 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., Gertler, M., 2006. Medium-term business cycles. American Economic Review 96 (3), 523–551.

Dou, L., Müller, U. K., 2021. Generalized local-to-unity models. Econometrica forthcoming.

Drehmann, M., Borio, C. E., Tsatsaronis, K., 2012. Characterising the financial cycle: don't lose sight of the medium term!, bIS working paper.

Elliott, G., Stock, J. H., 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.

Hannan, E. J., Quinn, B. G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society: Series B (Methodological) 41 (2), 190–195.

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

Harding, D., Pagan, A., 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.

Kaiser, R., Maravall, A., 2001. Measuring business cycles in economic time series. Vol. 154. Springer Science & Business Media.

Marcellino, M., Stock, J. H., Watson, M. W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135 (1-2), 499–526.

Mikusheva, A., 2007. Uniform inference in autoregressive models. Econometrica 75 (5), 1411–1452.

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

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariane matrix. Econometrica 55, 703–708.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69 (6), 1519–1554.

Perron, P., Ng, S., 1996. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Review of Economic Studies 63 (3), 435–463.
Phillips, P. C. B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74 (3), 535–547.
Phillips, P. C. B., 1988. Regression theory for near-integrated time series. Econometrica 56 (5), 1021–1043.
Phillips, P. C. B., Jin, S., 2002. The KPSS test with seasonal dummies. Economics Letters 77 (2), 239–243.
Phillips, P. C. B., Solo, V., 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., Vlekke, M., 2018. Business, housing, and credit cycles. Journal of Applied Econometrics 33 (2), 212–226.
Sargent, T. J., 1987. Macroeconomic Theory, 2nd Edition. Economic Theory, Econometrics, and Mathematical Economics Series. Emerald Group Publishing Limited, Bingley, UK.
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., Proaño, C. R., Wolters, J., 2019. Characterizing the financial cycle: Evidence from a frequency domain analysis. Journal of Banking & Finance 106, 568–591.