Testing for a Unit Root in a Near-Integrated Model with Skip-Sampled Data

Marcus J. Chambers
University of Essex

July 2014

Abstract

This paper examines tests for a unit root in skip-sampled data. A generalisation of the usual discrete time framework is proposed that allows for a continuous time detrending procedure prior to estimation of the resulting discrete time dynamic model that embodies exactly the restrictions imposed by the process of temporal aggregation. A simulation study reveals that taking these restrictions into account can yield improved size and power properties compared to a statistic based on a model that ignores the temporal aggregation, and an empirical illustration of the methods using monthly producer price data for the UK and US is provided. Further avenues for investigation in future work are also highlighted.

Address for Correspondence: Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England.
Tel: +44 1206 872756; fax: +44 1206 872724; e-mail: mchamb@essex.ac.uk.

1. INTRODUCTION

The research of John Nankervis was based firmly in the analysis of time series, and some of his most influential work was concerned with the study of tests for a unit root, a topic that established itself at the forefront of econometric research in the late 1980s and into the 1990s. Prominent examples of John's contributions to this literature include DeJong, Nankervis, Savin and Whiteman (1992a, 1992b).[1] One of the main findings to emerge from this work (and that of others) was that many unit root tests suffer from significant size distortions and low power, leading DeJong, Nankervis, Savin and Whiteman (1992a, p.432) to conclude that "inferences based exclusively on tests for integration may be fragile." A significant volume of subsequent research has attempted to develop tests with improved size and power properties, and the choice of tests available to applied researchers is now quite substantial.

A popular approach to testing for a unit root nowadays derives the test statistic from a sufficiently long autoregressive approximation based on data that have been detrended prior to estimation. Choice of lag length (k) in the autoregression is a critical part of the process, and Ng and Perron (2001) have proposed modified information criteria designed for the optimal choice of k in the construction of unit root test statistics that allow k to grow at an appropriate rate with sample size. A key advantage of employing an autoregression as the basis for the unit root test is that it is easy to estimate (repeatedly, for different values of k) with standard regression software. However, one reason why the choice of k is so critical is that the autoregression will typically be misspecified and so k needs to be large enough to capture the serial correlation in the detrended series. For example, if the series is in fact of autoregressive moving average (ARMA) form (a type of process often employed to evaluate the properties of tests in simulation studies) then k may have to be large to account for the serial correlation induced by the moving average (MA) component.

An area where moving average disturbances arise naturally is when the observed series is temporally aggregated.[2] For example, discrete time observations generated by an underlying continuous time autoregressive process of order p satisfy an ARMA($p, p-1$) process if the variable is a stock (observed at single points in time i.e. skip-sampled) or an ARMA(p, p) process for a flow variable (observed as an integral of the underlying rate of flow over the sampling interval); see, for example, Bergstrom (1984, pp.1201–1202 in particular). The same discrete time ARMA orders have also been shown to hold by Chambers and Thornton (2012) if the underlying continuous time process is ARMA(p, q) (with $q < p$). One advantage of using an underlying continuous time process is that the orders of the resulting ARMA process for the discrete time observations are invariant to the sampling frequency, an invariance property that is not universally true when the underlying process operates in discrete time on a finer timescale than the observations; see, for example, Weiss (1984). The observed temporally aggregated discrete time process also inherits its integration order from

---

[1]I based my presentation at John's Memorial Conference at the University of Essex in July 2013 on these two articles in a personal account of how I first came to know John's work as a young lecturer in the early 1990s. I would particularly like to thank Gene Savin for the additional insight and background that he was able to give me concerning these articles.

[2]We use the term 'temporal aggregation' in a generic sense to denote the discrete time sampling of both stock and flow variables but note that sometimes it is used in the literature purely for the latter while the former is referred to as skip-sampling or systematic sampling.

the underlying continuous time process, as demonstrated by Phillips (1991), who also shows that the property of cointegration of multiple series is preserved under temporal aggregation. For example, the discrete time process will have a unit root (and be integrated of order one) if the continuous time process has a zero root (integrated of order one in continuous time).

The above preservation of integration order under temporal aggregation suggests that standard unit root tests can be applied to temporally aggregated data to determine the order of integration of the underlying process. The presence of MA disturbances does suggest, however, that the number of lags in the autoregression used to carry out the test may have to be quite large in order to capture adequately the serial correlation properties. However, if the temporal aggregation is to be taken seriously, it is possible that there may be efficiency gains to be made by accounting for the restrictions on the discrete time data implied by the process of temporal aggregation. For example, a continuous time $AR(p)$ process contains p AR parameters plus the variance of the disturbance, which implies that, if the variable is a stock, the $p + (p - 1)$ discrete time ARMA parameters (plus the innovation variance) are functions of this smaller number of underlying parameters. It therefore seems apposite to investigate whether accounting for temporal aggregation yields improvements in testing the order of integration in a time series, and this investigation is the aim of the present paper.

The paper is organised as follows. Section 2 summarises a typical approach to testing for the presence of a unit root in a discrete time series, including a discussion of generalised least squares (GLS) data detrending. The framework is then modified to show how the procedure can be carried out in a continuous time setting with skip-sampled data, allowing for an arbitrary sampling interval and data span. Theorem 1 establishes the appropriate continuous time GLS detrending procedure based on skip-sampled discrete time data, and two unit root tests are discussed – one is the normalised estimator of the continuous time (zero root) parameter, the other being a likelihood ratio statistic. Section 3 provides an illustrative example of the methods at work in a model driven by an AR(1) disturbance in continuous time. The discrete time skip-sampled data are shown in Theorem 2 to satisfy a discrete time ARMA(2,1) process and an explicit representation for the MA(1) disturbance is shown, in Proposition 1, to have a strictly positive moving average parameter. This representation provides a particularly convenient basis for computing the Gaussian likelihood function. In order to investigate the effectiveness of the two tests and to compare them with a test based on an unrestricted discrete time ARMA(2,1) model that ignores the temporal aggregation, a simulation study is conducted in section 4 using the model in section 3 as a basis for data generation. The test results obtained from 10,000 replications are presented in four tables, allowing for two types of detrending, and both raw and size-adjusted power results are reported. The main conclusion is that the tests that allow for temporal aggregation generally outperform the test that does not, implying that in circumstances where temporal aggregation is an issue, then it should be taken seriously in the conduct of unit root tests. An empirical illustration of the methods is provided in section 5 using monthly data on producer price indices for the UK and the US. An appendix contains the proofs of the two theorems and proposition in the main text, as well as an additional lemma that is utilised in the proof of Proposition 1.

The following notation will be used throughout: L denotes the lag operator such that

$L^j x_t = x_{t-j}$; D denotes the mean square differential operator which satisfies

$$\lim_{h \to \infty} E \left[\frac{x(t+h) - x(t)}{h} - Dx(t) \right]^2 = 0$$

for some continuous time process $x(t)$ that is mean square continuous; $x_t \sim \text{iid}(0, \sigma_x^2)$ means that x_t is independently and identically distributed with mean zero and variance σ_x^2; $W(r)$ denotes a standard Wiener process; $W_c(r)$ denotes an Ornstein Uhlenbeck process with parameter c that satisfies $dW_c(r) = cW_c(r)dr + dW(r)$ (so that $W_0(r) = W(r)$); $\Rightarrow$ denotes weak convergence; and integrals such as $\int_0^1 W_c(r)dW(r)$ will be denoted more simply as $\int_0^1 W_c dW$ for notational convenience.

2. THE TESTING FRAMEWORK

A common framework for testing a scalar random process, y_t, for the presence of a unit root in discrete time assumes that y_t satisfies

$$y_t = d_t + u_t, \quad u_t = \rho u_{t-1} + v_t, \quad v_t = \delta(L)\epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where d_t is a deterministic time trend, $\rho = 1 + c/T$ for some finite constant c, $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$, $\delta(z) = \sum_{j=0}^{\infty} \delta_j z^j$, $\delta_0 = 1$, $\sum_{j=0}^{\infty} j|\delta_j| < \infty$, and T denotes sample size. The constant c is often referred to as a local-to-unity parameter because it measures the extent of deviations from a unit root ($\rho = 1$). The deterministic term, d_t, is usually assumed to be a low-order polynomial in t of the form $d_t = \psi' z_t$, where $z_t = [1, t, t^2, \ldots, t^\tau]$ and ψ is a $(\tau + 1)$-vector of coefficients, the cases $\tau = 0$ and $\tau = 1$ receiving most attention; we will also use $\tau = -1$ to denote the case where no deterministic trend is present.

In practice, the coefficients of ψ are unknown and have to be estimated from the data, and a form of quasi-differencing – usually referred to as GLS detrending – is often employed. For a variable x_t the quasi-differenced variable is denoted $x_t^{\bar{\rho}}$ and is defined for $t = 1, \ldots, T$ as $x_t^{\bar{\rho}} = x_t - \bar{\rho} x_{t-1}$ (with $x_0^{\bar{\rho}} = x_0$) for some appropriate choice of $\bar{\rho} = 1 + \bar{c}/T$. Elliott, Rothenberg and Stock (1996) recommend that when $\tau = 0$, $\bar{c} = -7$ and when $\tau = 1$, $\bar{c} = -13.5$, these values being chosen so as to make the asymptotic local power function of tests tangent to the asymptotic Gaussian power envelope at the point where power equals one half. The detrended series is obtained as $y_t^d = y_t - \hat{\psi}' z_t$ where $\hat{\psi}$ is obtained from a regression of $y_t^{\bar{\rho}}$ on $z_t^{\bar{\rho}}$.

A large number of tests exist for assessing whether $\rho = 1$ based on the detrended variable y_t^d, and many attempt to replicate the asymptotically pivotal distribution that is obtained under the ideal (but typically unrealistic) setting where $v_t = \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$. Often a test is based on an estimator of ρ, say $\hat{\rho}$, or on its associated t-ratio for testing $\rho = 1$. In the former case, the limiting distributions of the normalised estimator of ρ are as follows when $T \to \infty$:

$$T(\hat{\rho} - \rho) \Rightarrow \begin{cases} \dfrac{\int_0^1 W_c dW}{\int_0^1 W_c^2}, & \tau = -1, 0, \\[4mm] \dfrac{\int_0^1 V_c dW}{\int_0^1 V_c^2}, & \tau = 1, \end{cases} \tag{2}$$

3

where the process $V_c(r)$ arising when $\tau = 1$ is given by

$$V_c(r) = W_c(r) - r\left(\lambda W_c(1) + 3(1-\lambda)\int_0^1 sW_c(s)ds\right)$$

and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$. Under the null hypothesis ($\rho = c = 0$) critical values for these distributions are widely available. In order to use these distributions for inference in practice it is necessary to deal in an appropriate way with the serial correlation that manifests itself in v_t. Some approaches to this include the nonparametric adjustments proposed by Phillips and Perron (1988) as well as the fully parametric autoregressive approximations as in Said and Dickey (1984) and the M-type tests of Ng and Perron (2001).

To analyse situations in which the variable of interest evolves at a finer timescale than the sampling frequency, let N denote the span of time over which the variable is observed, and let h denote the sampling interval (i.e. the length of time between observations – assumed constant), so that the number of observations is given by $T = N/h$. The continuous time variable $y(t)$ is then assumed to satisfy

$$y(t) = d(t) + u(t), \quad Du(t) = \alpha u(t) + v(t), \quad \phi(D)v(t) = \theta(D)\epsilon(t), \quad t \geq 0, \qquad (3)$$

where $d(t)$ is a deterministic trend function, $\alpha = c/N$, $\phi(z) = z^p + \sum_{j=0}^{p-1}\phi_j z^j$, $\theta(z) = 1 + \sum_{j=1}^{q}\theta_j z^j$, $\epsilon(t)$ is a continuous time white noise process with mean zero and variance σ_ϵ^2, and $q < p$ to ensure that the spectral density of $v(t)$ is integrable (and, hence, that $v(t)$ has finite variance). It is also assumed that the roots of the equation $\phi(z) = 0$ have negative real parts to ensure that $v(t)$ is covariance stationary. The deterministic trend function is assumed to be of the form $d(t) = \psi'z(t)$, where $z(t) = [1, t, t^2, \ldots, t^\tau]$ and ψ is a $(\tau + 1)$-vector of coefficients, as in the discrete time setup except that t is now continuous rather than discrete. The observations will be assumed to be of the form $y_{th} = y(th)$ $(t = 1, \ldots, T)$ so that the observed sequence is

$$y_0, y_h, y_{2h}, \ldots, y_{Th}, \qquad (4)$$

where $y_{Th} = y(N)$.[3] Note that the objective is to test the null hypothesis that $\alpha = 0$ (equivalently, that $c = 0$) and that stationary alternatives correspond to $\alpha < 0$ (equivalently, $c < 0$).

According to (3), the trend-adjusted continuous time variable $u(t) = y(t) - d(t)$ satisfies the stochastic differential equation

$$\tilde{\phi}(D)u(t) = \theta(D)\epsilon(t), \quad \tilde{\phi}(z) = (z - \alpha)\phi(z), \qquad (5)$$

which is a continuous time ARMA($p + 1, q$) process. When $\alpha < 0$ the process $u(t)$ is covariance stationary but when $\alpha = 0$ it possesses a zero root and is, therefore, integrated, in which case the variable $Du(t)$ is ARMA(p, q). A discrete time sequence generated by this specification is shown in Chambers and Thornton (2012) to satisfy the discrete time

---

[3]It is also possible to consider the case of flow variables where the observations are instead of the form $y_{th} = h^{-1}\int_{th-h}^{th} y(r)dr$, which induces an additional moving average order into the disturbances of the discrete time representations; see Working (1960).

$\text{ARMA}(p+1, p)$ model

$$u_{th} = f_1 u_{th-h} + \ldots + f_{p+1} u_{th-(p+1)h} + w_{th}, \tag{6}$$

where w_{th} is $\text{MA}(p)$ and the coefficients $f_1, \ldots, f_{p+1}$ depend on α as well as the coefficients of the autoregressive polynomial $\phi(z)$. The zero root in continuous time translates into a unit root in discrete time so that, when $\alpha = 0$, $\Delta_h u_{th} = u_{th} - u_{th-h}$ satisfies the $\text{ARMA}(p, p)$ model

$$\Delta_h u_{th} = f_{10} \Delta_h u_{th-h} + \ldots + f_{p0} \Delta_h u_{th-ph} + w_{th}, \tag{7}$$

where w_{th} is also $\text{MA}(p)$ and $f_{10}, \ldots, f_{p0}$ depend only on the coefficients of $\phi(z)$. The serial correlation arising from the MA disturbance would typically result in the need for additional lags (beyond p) in a unit root test based on an autoregressive approximation. Taking the temporal aggregation into account provides a more parsimonious representation.

It is possible to detrend the data in continuous time even though observations are available only at discrete intervals of length h. In essence, what is required is estimation of the equation

$$Dy(t) = \frac{\bar{c}}{N} y(t) + \psi' \left(Dz(t) - \frac{\bar{c}}{N} z(t) \right) + e(t) \tag{8}$$

for a suitable choice of $\bar{c}$, where $e(t)$ is a continuous time random disturbance process. This representation is motivated by noting that GLS detrending in discrete time is based on a regression that can be written in the form

$$\Delta y_t - \frac{\bar{c}}{T} y_{t-1} = \psi' \left(\Delta z_t - \frac{\bar{c}}{T} z_{t-1} \right) + e_t,$$

because $y_t^{\bar{\rho}} = \Delta y_t - (\bar{c}/T) y_{t-1}$ and similarly for $z_t^{\bar{\rho}}$. The following Theorem presents the results for the case $\tau = 1$.

Theorem 1. *Let $y(t)$ be generated according to (3) with the observed sequence given by (4). Then GLS detrending in continuous time when $\tau = 1$ is carried out by estimating the equation*

$$Dy(t) = \frac{\bar{c}}{N} y(t) - \psi_0 \frac{\bar{c}}{N} + \psi_1 \left(1 - \frac{\bar{c}}{N} t \right) + e(t), \quad t > 0, \tag{9}$$

where $\bar{c}$ is the detrending parameter. Estimation of (9) is equivalent to the following regression using the discrete time observations:

$$y_{th} - e^{\bar{c}/T} y_{th-h} = \psi_0 \left(1 - e^{\bar{c}/T} \right) + \psi_1 \left(th - e^{\bar{c}/T}(th - h) \right) + e_{th}, \quad t = 1, \ldots, T, \tag{10}$$

where e_{th} is a discrete time random disturbance.

The form of continuous time detrending equation (9) arises because, when $\tau = 1$, we have $z(t) = [1, t]'$ and $Dz(t) = [0, 1]'$. The discrete time equivalent in (10) is of the same form as in the usual discrete time approach except that $e^{\bar{c}/T}$ replaces $1 + \bar{c}/T$ (which are the first two terms of the series expansion of $e^{\bar{c}/T}$) and we are explicitly allowing for an arbitrary sampling interval h; the deterministic terms on the right-hand-side of (10) are clearly of the form $\psi'(z_{th} - e^{\bar{c}/T} z_{th-h})$. Theorem 1 also encompasses the case $\tau = 0$ which arises when $\psi_1 = 0$. If $\hat{\psi}$ denotes the estimator of ψ obtained from (10) then the detrended series is given by $y_t^d = y_t - \hat{\psi}' z_{th}$.

Testing the hypothesis that $\alpha = 0$ in (3) typically requires an estimate of this key parameter. The simple model $Du(t) = \alpha u(t) + \epsilon(t)$ obtained by setting $\phi(z) = \theta(z) = 1$ in (3) is sufficient for highlighting some of the key results. Discrete time data generated by this model satisfy $u_{th} = e^{\alpha h} u_{th-h} + w_{th}$ where w_{th} is white noise with variance $\sigma_\epsilon^2 (e^{2\alpha h} - 1)/(2\alpha)$. Recall that $\alpha = c/N$. For $\tau = -1$ Phillips (1987) shows that the OLS estimator, $\hat{f}$, of $f = e^{\alpha h}$ satisfies

$$T(\hat{f} - f) \Rightarrow \frac{\int_0^1 W_c dW}{\int_0^1 W_c^2}$$

as $T \to \infty$, which is, of course, the distribution in (2). Noting that $\hat{\alpha} = \ln(\hat{f})/h$ enables the following result to be obtained via a Taylor expansion:

$$N(\hat{\alpha} - \alpha) \Rightarrow \frac{\int_0^1 W_c dW}{\int_0^1 W_c^2} \tag{11}$$

as $N \to \infty$; see also Zhou and Yu (2010) who also analyse what happens when $h \to 0$ either simultaneously or when N is fixed. Note that it is data span, N, rather than sample size, T, that is the appropriate normalisation for this distribution. The distribution in (11) also holds under GLS detrending when $\tau = 0$ but when $\tau = 1$ we obtain

$$N(\hat{\alpha} - \alpha) \Rightarrow \frac{\int_0^1 V_c dW}{\int_0^1 V_c^2}, \tag{12}$$

also as in (2).

Although the above results are obtained under a simple model which abstracts from additional serial correlation, it is, nevertheless, the pivotal distributions reported above that unit root test statistics would attempt to achieve (under the null of $c = 0$). Under more general and realistic scenarios the objective would be to estimate the continuous time parameter α on the basis of the discrete time ARMA$(p+1, q)$ representation (6). A common approach to estimating continuous time models based on discrete time observations is to maximise the likelihood function under the assumption of a Gaussian distribution, and a number of approaches are available for the construction of the likelihood function. The approach of Bergstrom (see the articles collected in Bergstrom (1990), for example) is to base the likelihood on the $T \times 1$ vector $w = [w_h, \ldots, w_{Th}]'$ with covariance matrix $\Gamma_w = E(ww')$ which, due to the MA properties of w_{th}, is known to have a sparse Toeplitz form – only the elements on the principal diagonal and the $q - 1$ neighbouring bands are non-zero. If ϕ and θ denote the vectors of AR and MA parameters, respectively, then this approach leads to the form (ignoring a constant)

$$L_w(\alpha, \phi, \theta, \sigma_\epsilon^2) = -\frac{1}{2} \ln |\Gamma_w| - \frac{1}{2} w' \Gamma_w^{-1} w, \tag{13}$$

and Bergstrom (1990) proposes methods to deal with the efficient inversion of the covariance matrix. Provided the serial correlation is dealt with appropriately (for example, the orders p and q are chosen correctly) the normalised estimator $N\hat{\alpha}$ will have the limiting distributions in (11) and (12) above (with $c = 0$), depending on the form of detrending employed, under the null hypothesis that $\alpha = 0$. Such results in likelihood-based approaches to unit root testing have been demonstrated by, for example, Yap and Reinsel (1995) and Shin and

6

Fuller (1998). An alternative to using the statistic $N\hat{\alpha}$ for inference would be to use a likelihood ratio statistic. Denoting unrestricted estimates as $\hat{\theta}$, for example, and restricted estimates under the null of $\alpha = 0$ as $\hat{\theta}_0$, for example, leads to the statistic

$$LR = 2\left[L_w(\hat{\alpha}, \hat{\phi}, \hat{\theta}, \hat{\sigma}_{\epsilon}^2) - L_w(0, \hat{\phi}_0, \hat{\theta}_0, \hat{\sigma}_{\epsilon,0}^2)\right].$$

In the case of no detrending ($\tau = -1$) Yap and Reinsel (1995) have shown that, under the null hypothesis (where $c = 0$ and noting that $W_0 = W$),

$$LR \Rightarrow \frac{\left(\int_0^1 W\,dW\right)^2}{\int_0^1 W^2} \tag{14}$$

as $T \to \infty$. We conjecture that a similar result would arise under GLS detrending with $\tau = 0$ and with W replaced by V_0 when $\tau = 1$; this appears to be borne out by simulations reported in the next section.

3. AN ILLUSTRATIVE EXAMPLE

In order to investigate the performance of unit root tests with temporally aggregated data it is necessary to be able to generate the discrete time data $\{y_{th}\}_{t=1}^T$ that satisfy the temporal aggregation scheme exactly. From (3) it is obvious that $y_{th} = d_{th} + u_{th}$ where $d_{th} = d(th)$ and $u_{th} = u(th)$; the task, therefore, is to generate $\{u_{th}\}_{t=1}^T$ to satisfy the underlying continuous time mechanism in (3). Given that $u(t)$ depends on $v(t)$ we consider the following continuous time AR(1) specification for $v(t)$:

$$Dv(t) = \phi v(t) + \epsilon(t), \quad t > 0, \tag{15}$$

where $\phi < 0$ to ensure that $v(t)$ is stationary; in effect we have taken $\phi(z) = z - \phi$ and $\theta(z) = 1$ in (3). This implies that the detrended variable $u(t)$ is the continuous time ARMA(2,1) process

$$D^2 u(t) = (\phi + \alpha)Du(t) - \phi\alpha u(t) + \epsilon(t); \tag{16}$$

under the null hypothesis ($\alpha = 0$) $u(t)$ satisfies $D^2 u(t) = \phi Du(t) + \epsilon(t)$ as it is now $Du(t)$, rather than $u(t)$ itself, that is a stationary continuous time process. An exact discrete time representation for continuous time ARMA(p, q) processes can be found in Chambers and Thornton (2012) whose Corollary 1 is relevant to the (stock) sampling scheme considered here; we are also, in effect, dealing with a special case of the continuous time ARMA(2,1) process analysed in Thornton and Chambers (2013) in which the continuous time moving average parameter is set equal to zero. However, we follow an alternative method of deriving the discrete time representation here that proceeds in two steps – first, the discrete time representation for u_{th} based on the equation $Du(t) = \alpha u(t) + v(t)$ is obtained; secondly, the exact form for the disturbance in this representation is obtained based on (15). A unique feature is that an exact discrete time representation for the integral of a function of a continuous time process is obtained – see the proof of Theorem 2 for details (in particular, the variable ζ_{th} defined in (24)).

Theorem 2. *Let $u(t)$ satisfy $Du(t) = \alpha u(t) + v(t)$ where $v(t)$ is generated according to (15)*

with $v(0)$ given. Then the sequence $u_0, u_h, u_{2h}, \ldots, u_{Th}$, where $u_{th} = u(th)$, satisfies:
(i) when $\alpha \neq 0$,

$$u_{th} = (e^{\alpha h} + e^{\phi h})u_{th-h} - e^{(\alpha+\phi)h}u_{th-2h} + w_{th}, \quad t = 1, \ldots, T, \qquad (17)$$

where w_{th} is an MA(1) process with variance and autocovariance given by, when $\alpha \neq \phi$,

$$\gamma_0 = E(w_{th}^2) = \frac{\sigma_\epsilon^2}{(\phi - \alpha)^2}\left[(e^{2\alpha h} + 1)\left(\frac{e^{2\phi h} - 1}{2\phi}\right) + (e^{2\phi h} + 1)\left(\frac{e^{2\alpha h} - 1}{2\alpha}\right)\right.$$
$$\left. -2\left(\frac{e^{2(\phi+\alpha)h} - 1}{\phi + \alpha}\right)\right],$$

$$\gamma_1 = E(w_{th}w_{th-h}) = \frac{\sigma_\epsilon^2}{(\phi - \alpha)^2}\left[(e^{\phi h} + e^{\alpha h})\left(\frac{e^{(\phi+\alpha)h} - 1}{\phi + \alpha}\right) - e^{\alpha h}\left(\frac{e^{2\phi h} - 1}{2\phi}\right)\right.$$
$$\left. -e^{\phi h}\left(\frac{e^{2\alpha h} - 1}{2\alpha}\right)\right],$$

and when $\alpha = \phi$,

$$\gamma_0 = \sigma_\epsilon^2\left[\frac{e^{4\alpha h} - 1}{4\alpha^3} - \frac{he^{2\alpha h}}{\alpha^2}\right],$$

$$\gamma_1 = \sigma_\epsilon^2\left[\frac{he^{\alpha h}\left(e^{2\alpha h} + 1\right)}{4\alpha^2} - \frac{e^{\alpha h}\left(e^{2\alpha h} - 1\right)}{4\alpha^3}\right];$$

(ii) when $\alpha = 0$,

$$\Delta_h u_{th} = e^{\phi h}\Delta_h u_{th-h} + w_{th}, \quad t = 1, \ldots, T, \qquad (18)$$

where $\Delta_h = 1 - L^h$ and the autocovariances of w_{th} are given by

$$\gamma_0 = E(w_{th}^2) = \frac{\sigma_\epsilon^2}{\phi^2}\left[h(e^{2\phi h} + 1) - \left(\frac{e^{2\phi h} - 1}{\phi}\right)\right],$$

$$\gamma_1 = E(w_{th}w_{th-h}) = \frac{\sigma_\epsilon^2}{\phi^2}\left[\left(\frac{e^{2\phi h} - 1}{2\phi}\right) - he^{\phi h}\right].$$

Theorem 2 demonstrates that u_{th} is an ARMA(2,1) process under the maintained hypothesis while Δu_{th} is ARMA(1,1) under the null. Note that the variance and autocovariance of w_{th} are functions of all the parameters of the continuous time model, and these restrictions need to be taken into account in estimation in order to account properly for the temporal aggregation in the observed variable. Moreover it can be shown that the autocovariance, γ_1, is strictly positive; see Lemma 1 in the Appendix. Although one representation of the log-likelihood was provided in (13) we follow here an alternative approach based on an explicit MA(1) representation of w_{th} given below.

Proposition 1. Let w_{th} be the disturbance process defined in Theorem 2 with the stated variance (γ_0) and autocovariance (γ_1). Then w_{th} has the representation

$$w_{th} = \eta_{th} + \theta\eta_{th-h}, \qquad (19)$$

where $0 < \theta = (\gamma_0 - d)/(2\gamma_1) < 1$, $d = \sqrt{\gamma_0^2 - 4\gamma_1^2}$, and η_{th} is a white noise process with mean zero and variance $\sigma_\eta^2 = \gamma_1/\theta$.

Note that Proposition 1 is valid both when $\alpha \neq 0$ and when $\alpha = 0$. An alternative but equivalent representation for σ_η^2 is $\sigma_\eta^2 = \gamma_0/(1 + \theta^2)$, and this expression can be shown to be equal to the one in Proposition 1 by some straightforward algebra.[4] In view of Proposition 1 we can derive the Gaussian likelihood by taking $\eta_{th} \sim N(0, \sigma_\eta^2)$, resulting in (apart from a constant)

$$L_\eta(\alpha, \phi, \sigma_\epsilon^2) = -\frac{T}{2}\ln \sigma_\eta^2 - \frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}\eta_{th}^2. \tag{20}$$

For given values of $(\alpha, \phi, \sigma_\epsilon^2)$ the residuals can be computed using the formula

$$\eta_{th} = u_{th} - (e^{\alpha h} + e^{\phi h})u_{th-h} + e^{(\alpha+\phi)h}u_{th-2h} - \theta\eta_{th-h}.$$

The two likelihood-based statistics are, then, the normalised estimator of α, $N\hat{\alpha}$, and the likelihood ratio statistic $LR = L_\eta(\hat{\alpha}, \hat{\phi}, \hat{\sigma}_\epsilon^2) - L(0, \hat{\phi}_0, \hat{\sigma}_{\epsilon0}^2)$, where $\hat{\phi}_0$ and $\hat{\sigma}_{\epsilon0}^2$ denote the estimated values of ϕ and σ_ϵ^2, respectively, when the restriction $\alpha = 0$ is imposed.

4. SIMULATION RESULTS

We explore three tests for the existence of a unit root in the observed series y_{th} based on the model in section 3. Two of the tests take into account the fact that the data are temporally aggregated, and both are based on the Gaussian likelihood function. The first is the normalised estimator of α (normalised by span, rather than sample size), $N\hat{\alpha}$, based on the exact discrete time representation in Theorem 2, the second is the likelihood ratio statistic, LR. In order to examine the specific role that temporal aggregation is playing when carrying out tests using aggregated data, the third test is based on a discrete time ARMA(2,1) model that ignores the temporal aggregation. The model, based on (1), is given by

$$y_t = d_t + u_t, \quad u_t = \rho u_{t-1} + v_t, \quad v_t = \kappa v_{t-1} + \epsilon_t,$$

and the statistic considered is $T(\hat{\rho} - 1)$, where $\hat{\rho}$ denotes the estimator of ρ obtained by maximising the Gaussian likelihood function.[5]

In the simulations two values of the detrending parameter are considered, $\tau = 0, 1$, these being the most relevant from an empirical point of view, while two values of the stationary continuous time parameter are used, $\phi = \{-0.5, -0.25\}$. In view of the asymptotic distribution for $\hat{\alpha}$ being dependent on N rather than on T or h, we fix $h = 1$ and consider two values for the span, $N = \{120, 240\}$ (so that $T = \{120, 240\}$ as well). These values can be interpreted as corresponding to 10 and 20 years of monthly data or 30 and 60 years of quarterly data. Nine values of the local-to-unity parameter, c, are considered, covering the

---

[4]It would, in principle, be possible to extend Proposition 1 to higher-order MA processes although the complexity of deriving the solution increases as the MA order increases.

[5]An earlier version of this paper used a statistic based on a misspecified autoregression in which the lag length was chosen by minimising a modified information criterion proposed by Ng and Perron (2001). The statistics used in the current paper allow a more direct investigation of the effects of accounting for temporal aggregation.

interval -20 to 0 inclusive in steps of length 2.5. A total of 10,000 replications were carried out for each of the four combinations of ϕ and N. Asymptotic critical values for the test statistics under the null at the 5% significance level were obtained for $N\hat{\alpha}$ from the (exact) values reported in Perron (1989) for $\tau = 0$ (equal to -8.038) and Chambers (2013) for $\tau = 1$ (-16.594), while for the LR statistic they were obtained by simulation using one million replications of 10,000 observations yielding the values 4.133 (for $\tau = 0$) and 8.118 ($\tau = 1$).[6]

Table 1 reports the values of some of the discrete time parameters that correspond to the various continuous time parameters. Although the same values of c are used for each value of N it is worth noting that the values of $\alpha = c/N$ are much further away from zero when $N = 120$ than when $N = 240$ so it might be expected that unit (zero) root tests would be more powerful in the former case for each value of c. The coefficient $e^{\phi h}$ that appears in the discrete time representation in Theorem 2[7] is larger for smaller (absolute) values of ϕ; in fact, $e^{\phi h} = 0.6065$ when $\phi = -0.5$ and $e^{\phi h} = 0.7788$ when $\phi = -0.25$. These differences can also be expected to have an impact on the performance of the tests.

The size and power properties of the three test statistics are reported in Tables 2–5. The first two contain the results for $\tau = 0$, with Table 2 containing the size and raw power of the tests and Table 3 containing the size-adjusted power. Tables 4 and 5 report the same information for the case $\tau = 1$. A test can have high power due to it having a larger than nominal size and so the size-adjusted power results are also reported so as to put the comparison of the tests on the same empirical 5% level. Taking Table 2 first, it is apparent that the size of the LR test is closest to the nominal 5% level except when $N = 240$ and $\phi = -0.25$, a situation where both continuous time-based tests are under-sized. Otherwise the size of $N\hat{\alpha}$ is slightly better than $T(\hat{\rho} - 1)$. There is a fiarly clear ranking in terms of raw power, however, with $N\hat{\alpha}$ tending to have the highest power, followed by $T(\hat{\rho} - 1)$ and then LR, although the poorer performance of LR is undoubtedly due it not suffering from the size distortions of the other two tests. The size-adjusted power results in Table 3 put all three tests on the same empirical 5% size footing and show that both continuous time-based tests dominate $T(\hat{\rho} - 1)$ with $N\hat{\alpha}$ having highest size-adjusted power for smaller values of c i.e. those values of α closest to the null of $\alpha = 0$. For the case of $\tau = 1$ detrending Table 4 shows that the sizes of the tests tend to be slightly larger than for the $\tau = 0$ case but again it is the LR test that is closest to the nominal 5% level. In terms of raw power the ranking tends to be $N\hat{\alpha} > T(\hat{\rho} - 1) > LR$ but for size-adjusted power Table 5 indicates the ranking $N\hat{\alpha} > LR > T(\hat{\rho} - 1)$ although $T(\hat{\rho} - 1)$ does better than LR when $N = 120$ and $\phi = -0.25$.

In summary the evidence, at least based on these simulations, appears to be that taking temporal aggregation into account can result in tests with better size and power properties than tests that ignore the temporal aggregation restrictions, even though the latter are based on an ARMA model with the correct orders.

---

[6]Yap and Reinsel (1995) report the value 8.16 but used a degrees of freedom adjustment in calculating their corresponding statistic.

[7]This is also the AR(1) coefficient for ζ_{th}, the disturbance in the representation (24) of u_{th}, given in the Appendix.

5. AN EMPIRICAL EXAMPLE

We apply the continuous time unit root tests to monthly time series data on the producer price index for the UK and the US over the period February 1996 to March 2014, yielding a total of 218 observations. Letting $p(t)$ denote the unobserved price index in continuous time, and $p_t = p(t)$ denote the observed monthly counterpart (for $t = 1, \ldots, 218$), the continuous time model is defined by the following three equations:

$$\ln p(t) = \psi_0 + \psi_1 t + u(t), \tag{21}$$
$$Du(t) = \alpha u(t) + v(t), \tag{22}$$
$$Dv(t) = \phi v(t) + \epsilon(t). \tag{23}$$

These equations imply that the detrended variable $u(t) = \ln p(t) - \psi_0 - \psi_1 t$ satisfies the second-order stochastic differential equation given by

$$D^2 u(t) = (\phi + \alpha)Du(t) - \phi \alpha u(t) + \epsilon(t).$$

The objective is to test the null hypothesis that $\alpha = 0$ based on the discrete time observations $\ln p_t$.

Defining $u_t^d = \ln p_t - \hat{\psi}_0 - \hat{\psi}_1 t$ to be the GLS-detrended variable in accordance with Theorem 1, it follows from Theorem 2 that u_t^d follows the ARMA(2,1) process

$$u_t^d = (e^\alpha + e^\phi)u_{t-1}^d - e^{(\alpha+\phi)}u_{t-2}^d + \eta_t + \theta\eta_{t-1}, \quad t = 3, \ldots, T,$$

where use has also been made of the result in Proposition 1. Under the null hypothesis ($\alpha = 0$) this becomes an ARMA(1,1) in first differences of the form

$$\Delta u_t^d = e^\phi \Delta u_{t-1}^d + \eta_t + \theta\eta_{t-1}, \quad t = 3, \ldots, T.$$

Note that two observations are required for initial conditions, resulting in an effective sample size of $T - 2 = 216$ in this application. There is no loss in generality by taking $h = 1$ here (with span $N = T = 216$) as no mixed frequency comparisons are being made. The same results for the test statistics are obtained when setting $h = 1/12$, for example, and defining the span to be $N = Th = 18$.

The results of the continuous time unit root tests (i.e. the statistics $N\hat{\alpha}$ and LR) are contained in Table 6, which also contains the estimated unrestricted and restricted continuous time models (upon which the test statistics are based) in the lower panel. Both tests are unable to reject the null hypothesis that $\alpha = 0$ i.e. that there is a zero root in continuous time (and a unit root in discrete time).[8] The implication is that the continuous time variable $Du(t)$ (the derivative of the detrended variable $u(t)$) satisfies the stationary process

$$D[Du(t)] = \phi Du(t) + \epsilon(t).$$

Indeed, the estimates of ϕ are significantly negative ruling out the possibility of a second

---

[8] For comparison, the statistic $T(\hat{\rho} - 1)$ obtained from an unrestricted discrete time ARMA(2,1) model yielded a value of -1.6200 for the UK and -12.1392 for the US, neither being able to reject the null that $\rho = 1$.

zero (unit) root in these two series, implying that producer price inflation is not itself an integrated process. The statistics S_4 and S_{12} reported in Table 6 are portmanteau-type statistics suggested by Bergstrom (1990, p.160) that have an approximate chi-squared distribution with 4 and 12 degrees of freedom, respectively (these correspond to the number of lags of residuals used in their construction); the Table reports the corresponding marginal probability values of the statistics. The null hypothesis can be interpreted as the model being adequately specified, at least in terms of capturing the dynamic features of the data. For the UK price series neither statistic is able to reject the null but there is some evidence of misspecification for the US series.

6. CONCLUDING COMMENTS

This paper has attempted to obtain some evidence as to whether there are gains to be made by taking into account the restrictions implied on discretely-observed data by the process of temporal aggregation when constructing unit root tests. A general framework has been proposed which, effectively, enables GLS detrending to be carried out in continuous time prior to estimating the continuous time parameter of interest in a discrete time ARMA model that accounts for the temporal aggregation. Simulations suggest that this approach leads to test statistics with better size and power properties than a test, based on an ARMA model of the correct orders, that ignores the temporal aggregation.

While the results reported here are encouraging, they also call for further work. For example, extending the framework to allow for a flow variable would be a logical next step. This would result in an additional order of moving average that may further enhance the advantages of accounting for the temporal aggregation. It would also be interesting to examine the effects in simulations of more general continuous time ARMA processes, such as in (3), using the exact discrete time models of Chambers and Thornton (2012), which would have the potential effect of increasing both the AR and MA orders compared to the simulations conducted here. Of course, all of the above could be analysed in a more detailed theoretical study to derive the asymptotic properties of the test statistics than has been carried out here and would be an important avenue to pursue.[9]

It is perhaps appropriate at this stage to return to the contributions of John Nankervis mentioned in the introductory paragraph to this paper. In the concluding section of DeJong, Nankervis, Savin and Whiteman (1992b, p.342) the authors, when reflecting on the simulation results they have obtained, state: "Given the slimness of this reed on which unit root testing now stands, efforts directed toward developing tests with higher power are in order." A great deal of work has been carried out since those articles were published in order to achieve this aim, and it is hoped that the current contribution provides another small step along this path.

---

[9]Indeed, the author has ongoing work on this topic with Roderick McCrorie and Michael Thornton that will be reported in the future. The current paper can be regarded as a prequel to this work.

APPENDIX:

Proof of Theorem 1. Let $\tilde{z}(t) = [-\bar{c}/N, 1 - (\bar{c}/N)t]'$ and $\psi = [\psi_0, \psi_1]'$. Then (9) can be written

$$Dy(t) = \frac{\bar{c}}{N}y(t) + \psi'\tilde{z}(t) + e(t)$$

whose solution (which is unique in the mean square sense) is given by

$$y(th) = e^{(\bar{c}/N)th}y(0) + \int_0^{th} e^{(\bar{c}/N)(th-s)}\psi'\tilde{z}(s)ds + \int_0^{th} e^{(\bar{c}/N)(th-s)}e(s)ds, \quad t > 0.$$

This solution can be used to show that

$$y(th) = e^{(\bar{c}/N)h}y(th - h) + \int_{th-h}^{th} e^{(\bar{c}/N)(th-s)}\psi'\tilde{z}(s)ds + e_{th}, \quad t = 1, \ldots, T,$$

where $e_{th} = \int_{th-h}^{th} e^{(\bar{c}/N)(th-s)}e(s)ds$. Evaluating the deterministic integral requires the components

$$\int_{th-h}^{th} e^{(\bar{c}/N)(th-s)}\psi_0\frac{\bar{c}}{N}ds = \psi_0\frac{\bar{c}}{N}\int_0^h e^{r\bar{c}/N}dr = \psi_0\left(e^{(\bar{c}/N)h} - 1\right),$$

$$\int_{th-h}^{th} e^{(\bar{c}/N)(th-s)}\psi_1 ds = \psi_1\int_0^h e^{r\bar{c}/N}dr = \frac{\psi_1}{\bar{c}/N}\left(e^{(\bar{c}/N)h} - 1\right),$$

and

$$
\begin{aligned}
\int_{th-h}^{th} e^{(\bar{c}/N)(th-s)}\psi_1\frac{\bar{c}}{N}sds &= \psi_1\frac{\bar{c}}{N}\int_0^h e^{r\bar{c}/N}(th - r)dr \\
&= \psi_1\frac{\bar{c}}{N}th\int_0^h e^{r\bar{c}/N}dr - \psi_1\frac{\bar{c}}{N}\int_0^h e^{r\bar{c}/N}rdr \\
&= \psi_1 th\left(e^{(\bar{c}/N)h} - 1\right) - \psi_1\left[e^{(\bar{c}/N)h}\left(h - \frac{1}{\bar{c}/N}\right) + \frac{1}{\bar{c}/N}\right].
\end{aligned}
$$

Combining these deterministic terms, and noting that the observations are $y_{th} = y(th)$ and that $h/N = 1/T$ implies that $e^{(\bar{c}/N)h} = e^{\bar{c}/T}$, results in (10) as required.

Proof of Theorem 2. (i) $\alpha \neq 0$: It is well known that u_{th} satisfies

$$u_{th} = e^{\alpha h}u_{th-h} + \zeta_{th}, \quad \zeta_{th} = \int_{th-h}^{th} e^{\alpha(th-r)}v(r)dr; \tag{24}$$

see, for example, Theorem 1 of Bergstrom (1984) which has been generalised above to allow for an arbitrary sampling interval h. The objective is to derive the law of motion for ζ_{th}, given that $v(t)$ is assumed to satisfy (15). The solution of (15) yields

$$v(th) = e^{\phi h}v(th - h) + \int_{th-h}^{th} e^{\phi(th-s)}\epsilon(s)ds;$$

substituting this into the definition of ζ_{th} yields

$$\zeta_{th} = e^{\phi h}\int_{th-h}^{th} e^{\alpha(th-r)}v(r - h)dr + w_{th}, \quad w_{th} = \int_{th-h}^{th} e^{\alpha(th-r)}\left(\int_{r-h}^r e^{\phi(r-s)}\epsilon(s)ds\right)dr.$$

13

The first integral on the right-hand-side can be written (using the substitution $s = r - h$)

$$\int_{th-h}^{th} e^{\alpha(th-r)} v(r-h) dr = \int_{th-2h}^{th-h} e^{\alpha(th-h-s)} v(s) ds = e_{th-h}$$

and so $\zeta_{th} = e^{\phi h} \zeta_{th-h} + w_{th}$. Applying the operator $(1 - e^{\phi h} L^h)$ to u_{th} in (24) yields the law of motion for u_{th} given in the Theorem.

The double integral defining w_{th} can be expressed as the sum of two single integrals. Taking, first, the case $\alpha \neq \phi$ we obtain

$$
\begin{aligned}
w_{th} &= \int_{th-h}^{th} \left(\int_s^{th} e^{\alpha(th-r)} e^{\phi(r-s)} dr \right) \epsilon(s) ds + \int_{th-2h}^{th-h} \left(\int_{th-h}^{s+h} e^{\alpha(th-r)} e^{\phi(r-s)} dr \right) \epsilon(s) ds \\
&= \int_{th-h}^{th} \pi_1(th - s) \epsilon(s) ds + \int_{th-2h}^{th-h} \pi_2(th - h - s) \epsilon(s) ds,
\end{aligned}
$$

where $\pi_1(r) = (e^{\phi r} - e^{\alpha r})/(\phi - \alpha)$ and $\pi_2(r) = (e^{\phi h} e^{\alpha r} - e^{\alpha h} e^{\phi r})/(\phi - \alpha)$; see, for example, McCrorie (2000) for details of this procedure. When $\alpha = \phi$ we find that

$$w_{th} = \int_{th-h}^{th} \left(\int_s^{th} e^{\alpha(th-s)} dr \right) \epsilon(s) ds + \int_{th-2h}^{th-h} \left(\int_{th-h}^{s+h} e^{\alpha(th-s)} dr \right) \epsilon(s) ds$$

but where, now, $\pi_1(r) = r e^{\alpha r}$ and $\pi_2(r) = e^{\alpha h}(h-r)e^{\alpha r}$. Clearly, w_{th} is an MA(1) process in both cases with autocovariances given by

$$E(w_{th}^2) = \sigma_\epsilon^2 \int_0^h \pi_1(r)^2 dr + \sigma_\epsilon^2 \int_0^h \pi_2(r)^2 dr, \quad E(w_{th} w_{th-h}) = \sigma_\epsilon^2 \int_0^h \pi_1(r) \pi_2(r) dr$$

woth $E(w_{th} w_{th-jh}) = 0$ for $j \geq 2$. The expressions in the Theorem are obtained by evaluating these integrals.

(ii) $\alpha = 0$: In this case the law of motion for u_{th} follows from that in part (i) by setting $\alpha = 0$. The autocovariances of the resulting w_{th} process can be obtained by taking the limits of those in part (i) as $\alpha \to 0$ or by setting $\alpha = 0$ in the expression for w_{th} and evaluating the integrals; we follow the second option and find that

$$
\begin{aligned}
w_{th} &= \int_{th-h}^{th} \left(\int_s^{th} e^{\phi(r-s)} dr \right) \epsilon(s) ds + \int_{th-2h}^{th-h} \left(\int_{th-h}^{s+h} e^{\phi(r-s)} dr \right) \epsilon(s) ds \\
&= \int_{th-h}^{th} \pi_{10}(th - s) \epsilon(s) ds + \int_{th-2h}^{th-h} \pi_{20}(th - h - s) \epsilon(s) ds,
\end{aligned}
$$

where $\pi_{10}(r) = (e^{\phi r} - 1)/\phi$ and $\pi_{20}(r) = (e^{\phi h} - e^{\phi r})/\phi$. The resulting autocovariances are

$$E(w_{th}^2) = \sigma_\epsilon^2 \int_0^h \pi_{10}(r)^2 dr + \sigma_\epsilon^2 \int_0^h \pi_{20}(r)^2 dr, \quad E(w_{th} w_{th-h}) = \sigma_\epsilon^2 \int_0^h \pi_{10}(r) \pi_{20}(r) dr,$$

while $E(w_{th} w_{th-jh}) = 0$ for $j \geq 2$. Evaluation of these integrals results in the stated expressions.

Proof of Proposition 1. From the representation (19) we obtain

$$E(w_{th}^2) = \gamma_0 = \sigma_\eta^2(1 + \theta^2), \quad E(w_{th}w_{th-h}) = \gamma_1 = \theta\sigma_\eta^2.$$

The second expression implies $\sigma_\eta^2 = \gamma_1/\theta$ and so it remains to determine θ. Substituting into the first expression for σ_η^2 gives $\gamma_0 = (\gamma_1/\theta)(1 + \theta^2)$, which is a quadratic in θ and can be written

$$\gamma_1\theta^2 - \gamma_0\theta + \gamma_1 = 0.$$

Let $d = \sqrt{\gamma_0^2 - 4\gamma_1^2}$. We first show that $d > 0$ by noting that it can be written in the form $d = \sqrt{(\gamma_0 - 2\gamma_1)(\gamma_0 + 2\gamma_1)}$. Lemma 2 (below) establishes that $(\gamma_0 - 2\gamma_1) > 0$ and $\gamma_1 > 0$, the latter implying that $(\gamma_0 + 2\gamma_1) > 0$, hence the term under the square-root sign is positive. The roots of the quadratic of interest are

$$r_1 = \frac{\gamma_0 - d}{2\gamma_1}, \quad r_2 = \frac{\gamma_0 + d}{2\gamma_1}.$$

We shall demonstrate that these roots are reciprocals of each other, and that the smallest (r_1) is the invertible root. If $r_1 = r_2^{-1}$ then it must be the case that

$$\frac{\gamma_0 - d}{2\gamma_1} = \frac{2\gamma_1}{\gamma_0 + d},$$

which implies that $(\gamma_0 - d)(\gamma_0 + d) = 4\gamma_1^2$ i.e. $\gamma_0^2 - d^2 = 4\gamma_1^2$. But this is clearly satisfied by the definition of d and hence the roots are reciprocals of each other. Next, we show that $0 < r_1 < 1$. Let $x = (\gamma_0/\gamma_1) - 2$ and note that

$$r_1 = \frac{1}{2}\left(\frac{\gamma_0}{\gamma_1} - \frac{d}{\gamma_1}\right) = \frac{1}{2}\left(x + 2 - \sqrt{x(x + 4)}\right)$$

where we have used the result that

$$\frac{d}{\gamma_1} = \sqrt{\frac{\gamma_0^2 - 4\gamma_1^2}{\gamma_1^2}} = \sqrt{\left(\frac{\gamma_0^2}{\gamma_1^2} - 4\right)} = \sqrt{\left(\frac{\gamma_0}{\gamma_1} - 2\right)\left(\frac{\gamma_0}{\gamma_1} + 2\right)} = \sqrt{x(x + 4)}.$$

Lemma 1 (below) implies that $x = (\gamma_0/\gamma_1) - 2 > 0$ and so we need to consider r_1 as the function of x defined above for $x > 0$. Doing so establishes that r_1 is a positive monotonically decreasing function of x with a maximum point $r_1 = 1$ at $x = 0$ and, hence, r_1 denotes the invertible root.

Lemma 1. *For γ_0 and γ_1 defined in Theorem 2, $\gamma_1 > 0$ and $\gamma_0 - 2\gamma_1 > 0$.*

Proof of Lemma 1. We begine with the case where $\alpha \neq 0$, $\alpha \neq \phi$. It is convenient to write γ_1 in the form

$$\gamma_1 = \frac{\sigma_\epsilon^2}{(\phi - \alpha)^2}\left\{e^{\alpha h}\left[\left(\frac{e^{(\phi+\alpha)h} - 1}{\phi + \alpha}\right) - \left(\frac{e^{2\phi h} - 1}{2\phi}\right)\right]\right.$$
$$\left. + e^{\phi h}\left[\left(\frac{e^{(\phi+\alpha)h} - 1}{\phi + \alpha}\right) - \left(\frac{e^{2\alpha h} - 1}{2\alpha}\right)\right]\right\}.$$

We can ignore the positive scaling factor outside the braces which means that we need to show that

$$e^{\alpha h}\left[\left(\frac{e^{(\phi+\alpha)h}-1}{\phi+\alpha}\right)-\left(\frac{e^{2\phi h}-1}{2\phi}\right)\right] > e^{\phi h}\left[\left(\frac{e^{2\alpha h}-1}{2\alpha}\right)-\left(\frac{e^{(\phi+\alpha)h}-1}{\phi+\alpha}\right)\right].$$

It is convenient to multiply through by $e^{-(\phi+\alpha)h}$ and to express both sides in terms of the common denominator $M = 2\phi\alpha(\phi+\alpha)$; the first term becomes

$$
\begin{aligned}
T_1 &= \frac{1}{M}\left[2\phi\alpha(e^{\alpha h}-e^{-\phi h})-\alpha(\phi+\alpha)(e^{\phi h}-e^{-\phi h})\right]\\
&= \frac{1}{M}\left[2\phi\alpha(e^{\alpha h}-e^{\phi h})+[2\phi\alpha-\alpha(\phi+\alpha)](e^{\phi h}-e^{-\phi h})\right]\\
&= \frac{1}{M}\left[2\phi\alpha(e^{\alpha h}-e^{\phi h})+2\alpha(\phi-\alpha)\sinh(\phi h)\right],
\end{aligned}
$$

while applying similar operations to the second term yields

$$T_2 = \frac{1}{M}\left[2\phi\alpha(e^{\alpha h}-e^{\phi h})+2\phi(\phi-\alpha)\sinh(\alpha h)\right].$$

Consider the difference

$$T_1 - T_2 = \frac{2(\phi-\alpha)}{M}\left[\alpha\sinh(\phi h)-\phi\sinh(\alpha h)\right]$$

and note that $\alpha\sinh(\phi h)-\phi\sinh(\alpha h) \lessgtr 0$ is equivalent to $\sinh(\phi h)/\phi \lessgtr \sinh(\alpha h)/\alpha$. For $x < 0$ the function $\sinh(x)/x$ is a decreasing function of x, enabling us to consider the following cases:

(a) $\phi < \alpha < 0$: here, $\phi - \alpha < 0$, $M < 0$ and $\sinh(\phi h)/\phi > \sinh(\alpha h)/\alpha$, resulting in $T_1 - T_2 > 0$.

(b) $\alpha < \phi < 0$: in this case, $\phi - \alpha > 0$, $M < 0$ and $\sinh(\phi h)/\phi < \sinh(\alpha h)/\alpha$, resulting in $T_1 - T_2 > 0$.

(c) $\alpha = \phi < 0$:

This establishes that $\gamma_1 > 0$ for all combinations of α and ϕ when $\alpha \neq 0$ and $\alpha \neq \phi$; a similar approach can be used to show that the same result holds when $\alpha = \phi$.

To show that $\gamma_0 - 2\gamma_1 > 0$ we can normalise the difference by the common denominator M as above and multiply through by $e^{-(\phi+\alpha)h}$ to give

$$
\begin{aligned}
\gamma_0 - 2\gamma_1 &= \frac{1}{M}\Big[\alpha(\phi+\alpha)e^{-\alpha h}(e^{\alpha h}+1)^2(e^{\phi h}-e^{-\phi h})-4\phi\alpha(e^{(\phi+\alpha)h}-e^{-(\phi+\alpha)h})\\
&\quad + \phi(\phi+\alpha)e^{-\phi h}(e^{\phi h}+1)^2(e^{\alpha h}-e^{-\alpha h})-4\phi\alpha(e^{-\phi h}+e^{-\alpha h})(e^{(\phi+\alpha)h}-1)\Big].
\end{aligned}
$$

Noting that (after expanding the term in brackets and simplifying)

$$e^{-\alpha h}(e^{\alpha h}+1)^2 = 2[1+\cosh(\alpha h)]$$

(and similarly for the term with α replaced by ϕ), and that

$$(e^{-\phi h}+e^{-\alpha h})(e^{(\phi+\alpha)h}-1) = 2\sinh(\phi h)+2\sinh(\alpha h),$$

the difference can be written (with some further simplification)

$$\gamma_0 - 2\gamma_1 = 4(\alpha - \phi)\left\{\alpha \sinh(\phi h)\left[1 + \cosh(\alpha h)\right] - \phi \sinh(\alpha h)\left[1 + \cosh(\phi h)\right]\right\}.$$

The term in braces is $\lesseqgtr 0$ according to whether

$$r(\phi) = \frac{\sinh(\phi h)}{\phi[1 + \cosh(\phi h)]} \lesseqgtr \frac{\sinh(\alpha h)}{\alpha[1 + \cosh(\alpha h)]} = r(\alpha).$$

For $x < 0$ the function $\sinh(x)/[x(1+\cosh(x))]$ is a positive monotonically increasing function and so we can consider the following cases:

(a) $\phi < \alpha < 0$: here, $\alpha - \phi > 0$, $M < 0$ and $r(\phi) < r(\alpha)$, resulting in $\gamma_0 - 2\gamma_1 > 0$.

(b) $\alpha < \phi < 0$: in this case, $\alpha - \phi < 0$, $M < 0$ and $r(\phi) > r(\alpha)$, resulting in $\gamma_0 - 2\gamma_1 > 0$.

(c) $\alpha = \phi < 0$:

This establishes that $\gamma_0 - 2\gamma_1 > 0$ for all combinations of α and ϕ when $\alpha \neq 0$ and $\alpha \neq \phi$; a similar approach can be used to show that the same result holds when $\alpha = \phi$.

When $\alpha = 0$, from Theorem 2 we have

$$\gamma_1 = \frac{\sigma_\epsilon^2}{\phi^2}\left[\left(\frac{e^{2\phi h} - 1}{2\phi}\right) - he^{\phi h}\right].$$

As the scaling factor is positive we focus on

$$\left(\frac{e^{2\phi h} - 1}{2\phi}\right) - he^{\phi h} = e^{\phi h}\left(\frac{e^{\phi h} - e^{-\phi h}}{2\phi}\right) - he^{\phi h} = e^{\phi h}\left(\frac{\sinh(\phi h)}{\phi} - h\right) > 0$$

for $\phi < 0$ as $\sinh(x)/x > 1$ for $x < 0$.

REFERENCES

Bergstrom, A. R. (1984) Continuous time stochastic models and issues of aggregation over time. In *Handbook of Econometrics*, vol. 2 (eds Z. Griliches and M. D. Intriligator), Amsterdam, North Holland.

Bergstrom, A. R. (1990) *Continuous Time Econometric Modelling*. Oxford, Oxford University Press.

Chambers, M.J. (2013) The calculation of some limiting distributions arising in near-integrated models with GLS detrending. *University of Essex Department of Economics Discussion Paper No. 727*.

Chambers, M. J. and Thornton, M. A. (2012) Discrete time representation of continuous time ARMA processes. *Econometric Theory* 28, 219–238.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992a) Integration versus trend stationarity in time series. *Econometrica* 60, 423–433.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992b) The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics* 53, 323–343.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

McCrorie, J. R. (2000) Deriving the exact discrete analog of a continuous time system. *Econometric Theory* 16, 998–1015.

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

Perron, P. (1989) The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* 5, 241–255.

Phillips, P. C. B. (1987) Towards a unified theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P. C. B. (1991) Error correction and long run equilibrium in continuous time. *Econometrica* 59, 967–980.

Phillips, P. C. B. and Perron, P. (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Said, S. E. and Dickey, D. A. (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599–607.

Shin, D. W. and Fuller, W. A. (1998) Unit root tests based on unconditional maximum likelihood estimation for the autoregressive moving average. *Journal of Time Series Analysis* 19, 591–599.

Thornton, M. A. and Chambers, M. J. (2013) Continuous-time autoregressive moving average processes in discrete time: representation and embeddability. *Journal of Time Series Analysis* 34, 552–561.

Weiss, A. A. (1984) Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics* 26, 271–281.

Working, H. (1960) Note on the correlation of first differences of averages in a random chain.*Econometrica* 28, 916–918.

Yap, S. F. and Reinsel, G. C. (1995) Results on estimation and testing for a unit root in the nonstationary autoregressive moving-average model. *Journal of Time Series Analysis* 16, 339–353.

Zhou, Q. and Yu, J. (2010) Asymptotic distributions of the least squares estimator for diffusion processes. *SMU Economics and Statistics Working Paper No. 20–2010*.

Table 1. Discrete time parameter values

$h = 1,\ \phi = -0.5,\ e^{\phi h} = 0.6065$					
c	-20	-15	-10	-5	0
$N = 120$					
α	-0.1667	-0.1250	-0.0833	-0.0417	0.0000
$e^{\alpha h}$	0.8465	0.8825	0.9200	0.9592	1.0000
γ_0	0.3519	0.3664	0.3817	0.3977	0.4146
γ_1	0.0868	0.0904	0.0942	0.0982	0.1024
θ	0.2637	0.2639	0.2640	0.2641	0.2641
σ_η^2	0.3290	0.3426	0.3568	0.3718	0.3875
$N = 240$					
α	-0.0833	-0.0625	-0.0417	-0.0208	0.0000
$e^{\alpha h}$	0.9200	0.9394	0.9592	0.9794	1.0000
γ_0	0.3817	0.3896	0.3977	0.4060	0.4146
γ_1	0.0942	0.0962	0.0982	0.1003	0.1024
θ	0.2640	0.2641	0.2641	0.2641	0.2641
σ_η^2	0.3568	0.3642	0.3718	0.3795	0.3875
$h = 1,\ \phi = -0.25,\ e^{\phi h} = 0.7788$					
c	-20	-15	-10	-5	0
$N = 120$					
α	-0.1667	-0.1250	-0.0833	-0.0417	0.0000
$e^{\alpha h}$	0.8465	0.8825	0.9200	0.9592	1.0000
γ_0	0.4435	0.4618	0.4810	0.5012	0.5225
γ_1	0.1104	0.1150	0.1198	0.1249	0.1302
θ	0.2666	0.2667	0.2669	0.2670	0.2670
σ_η^2	0.4141	0.4311	0.4490	0.4679	0.4877
$N = 240$					
α	-0.0833	-0.0625	-0.0417	-0.0208	0.0000
$e^{\alpha h}$	0.9200	0.9394	0.9592	0.9794	1.0000
γ_0	0.4810	0.4910	0.5012	0.5117	0.5225
γ_1	0.1198	0.1223	0.1249	0.1275	0.1302
θ	0.2669	0.2669	0.2670	0.2670	0.2670
σ_η^2	0.4490	0.4583	0.4679	0.4777	0.4877

Table 2. Size and power of tests: $\tau = 0$

			$h = 1, \phi = -0.5$			
	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	11.2	9.8	5.8	7.0	6.4	5.0
-2.5	25.0	28.3	14.4	19.5	20.7	13.3
-5.0	42.8	48.1	27.1	37.3	43.1	27.7
-7.5	61.0	67.3	44.0	56.8	63.2	45.9
-10.0	74.1	80.3	58.3	73.2	78.7	63.1
-12.5	83.1	87.2	71.6	86.2	88.9	79.2
-15.0	90.0	91.8	82.4	92.5	95.4	87.9
-17.5	93.1	95.3	87.6	96.9	98.0	93.9
-20.0	96.0	97.3	92.2	98.4	99.4	96.7
			$h = 1, \phi = -0.25$			
	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	13.6	12.9	5.0	5.0	4.6	2.8
-2.5	32.4	43.4	10.7	22.1	33.4	12.7
-5.0	49.2	57.5	22.5	39.5	46.2	27.0
-7.5	62.7	66.7	36.1	57.6	66.1	42.5
-10.0	71.1	74.4	48.9	72.0	79.9	57.6
-12.5	77.3	80.8	58.2	82.3	87.3	70.7
-15.0	83.8	86.3	67.1	89.2	93.1	80.3
-17.5	87.7	89.5	74.9	93.4	95.3	86.7
-20.0	89.9	92.1	80.0	96.1	97.3	92.3

Table 3. Size and size-adjusted power of tests: $\tau = 0$

$h = 1$, $\phi = -0.5$

	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	11.2	9.8	5.8	7.0	6.4	5.0
−2.5	12.0	18.1	12.3	14.3	17.5	13.3
−5.0	21.9	30.1	23.8	29.3	38.0	27.7
−7.5	33.9	46.6	39.1	46.6	57.9	46.0
−10.0	45.7	60.9	53.8	63.4	73.3	63.1
−12.5	55.8	69.0	67.8	77.4	85.4	79.4
−15.0	64.4	74.6	78.1	86.9	93.0	88.0
−17.5	71.2	81.7	85.0	93.1	96.5	93.9
−20.0	75.9	87.0	90.0	96.7	98.8	96.7

$h = 1$, $\phi = -0.25$

	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	13.6	12.9	5.0	5.0	4.6	2.8
−2.5	14.2	34.6	11.0	22.0	33.8	21.2
−5.0	24.7	39.3	22.7	39.4	47.8	41.0
−7.5	39.7	46.6	36.5	57.5	68.1	59.3
−10.0	50.9	52.1	49.4	72.0	81.7	74.4
−12.5	59.3	58.2	58.4	82.3	88.6	84.2
−15.0	65.9	65.7	67.5	89.2	93.6	91.4
−17.5	72.0	72.9	75.1	93.4	95.9	94.4
−20.0	76.4	76.4	80.2	96.1	97.4	97.0

Table 4. Size and power of tests: $\tau = 1$

	$h = 1, \phi = -0.5$					
	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	13.9	12.4	3.2	11.2	10.1	5.6
-2.5	26.7	31.0	6.4	14.2	18.7	7.5
-5.0	35.4	34.3	10.3	21.1	24.5	11.4
-7.5	43.6	46.5	15.8	31.4	37.0	18.6
-10.0	54.6	58.2	23.5	41.1	46.8	28.5
-12.5	62.3	64.6	32.2	52.8	59.2	38.6
-15.0	68.6	70.4	42.0	63.3	70.2	50.2
-17.5	72.7	76.2	49.2	73.2	80.2	61.0
-20.0	78.7	80.5	58.3	80.4	87.5	70.3

	$h = 1, \phi = -0.25$					
	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	12.8	15.7	1.3	18.2	16.5	4.8
-2.5	18.1	30.1	2.2	22.3	31.6	6.6
-5.0	27.6	47.1	3.6	31.0	50.6	10.2
-7.5	38.5	49.3	5.4	41.8	53.5	17.3
-10.0	50.6	56.3	10.0	50.8	62.2	25.3
-12.5	60.1	65.1	15.4	57.5	67.0	31.8
-15.0	65.2	70.0	20.4	66.8	72.6	42.2
-17.5	72.9	76.7	27.1	73.6	78.6	49.9
-20.0	77.7	80.0	32.2	77.1	81.3	58.2

Table 5. Size and size-adjusted power of tests: $\tau = 1$

$h = 1, \phi = -0.5$

	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	13.9	12.4	3.2	11.2	10.1	5.6
-2.5	10.2	13.1	10.4	5.9	10.9	6.9
-5.0	14.7	18.4	15.8	9.9	15.7	11.0
-7.5	21.5	26.7	23.7	14.8	22.3	17.2
-10.0	29.6	33.8	33.5	20.7	29.5	26.5
-12.5	37.2	42.9	43.5	26.5	38.2	35.8
-15.0	47.0	48.4	53.7	35.6	50.8	47.6
-17.5	53.5	57.6	60.5	44.1	59.6	58.3
-20.0	61.9	64.7	69.9	53.0	69.1	67.7

$h = 1, \phi = -0.25$

	$N = T = 120$			$N = T = 240$		
c	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR	$T(\hat{\rho} - 1)$	$N\hat{\alpha}$	LR
0.0	12.8	15.7	1.3	18.2	16.5	4.8
-2.5	7.3	11.8	7.5	6.7	27.5	7.0
-5.0	12.6	17.9	13.3	10.9	34.5	10.7
-7.5	20.3	30.7	19.9	16.9	40.8	17.6
-10.0	28.8	34.4	28.0	25.1	42.5	25.9
-12.5	39.5	42.7	37.9	30.6	43.9	32.7
-15.0	46.0	47.8	44.8	41.4	48.0	43.0
-17.5	55.9	57.9	51.7	48.4	49.8	51.7
-20.0	63.8	67.6	57.4	53.8	52.9	59.3

Table 6. Unit root tests: UK and US producer prices,
April 1996 – March 2014

	UK		US	
	Statistic	Critical value	Statistic	Critical value
$N\hat{\alpha}$	-1.1448	-16.594	-6.6312	-16.594
LR	0.4116	8.118	3.2550	8.118

	Continuous time model estimates			
Parameter	Unrestricted	Restricted	Unrestricted	Restricted
α	-0.0053 (0.0084)	0.0000	-0.0307 (0.0177)	0.0000
ϕ	-1.1726 (0.1944)	-1.1854 (0.1942)	-2.0486 (0.4113)	-2.1460 (0.4270)
σ_ϵ	0.0056 (0.0005)	0.0056 (0.0005)	0.0312 (0.0042)	0.0319 (0.0044)
$\ln L$	1185.6476	1185.4418	879.3238	877.6963
S_4	0.5150	0.5479	0.0285	0.0572
S_{12}	0.2510	0.2854	0.0124	0.0098

Numbers in parentheses are standard errors; entries for S_4 and S_{12} are probability values.