Anxious unit root processes

Jon Michel* Robert M. de Jong†

Abstract

This paper introduces what we will call the "anxious unit root process"; it adds a constant to the standard unit root process whenever the unit root process exceeds a latent bound. The latent bound adjusts whenever such a jump occurs. The process can be viewed as one that generates endogenous structural changes, or as one that is reluctant or eager to go up whenever the latent bound is exceeded. Therefore, this model captures behavior that has been discussed in economic theory and casual economics reporting. The anxious unit root process proposed in this paper is a random sequence that does not depend on sample size. We prove that the anxious unit root process satisfies an invariance principle. A nonstandard limit is obtained in the invariance principle. Therefore, proceeding as if the series satisfies a unit root and using techniques based on the unit root literature will be invalid. We develop a panel test statistic that tests for the null hypothesis of a unit root process against the alternative hypothesis of an anxious unit root process; we then show that we can reject for several macroeconomic aggregates, including the log of GDP. Additionally, we show how correct inference can easily be performed in a cointegration setting when a regressor is an anxious unit root process using a modification of Integrated Modified Ordinary Least Squares.

JEL Classification: C22

1 Introduction

The literature on unit roots in economic time series is venerable and substantial. The importance of unit roots in economic time series was first pointed out by Granger (see,

*Department of Economics, Ohio State University, 304 Arps Hall, Columbus, OH 43210, USA, Email michel.82@osu.edu.

†Department of Economics, Ohio State University, 444 Arps Hall, Columbus, OH 43210, USA, Email de-jong.8@osu.edu.

for example, Granger and Newbold (1974) and Granger (1981)). Dickey and Fuller first found the limit distribution of what came to be known as the Dickey-Fuller tests in Dickey and Fuller (1979). Phillips (1986) considered spurious regressions in the case of weakly dependent innovations; the Phillips-Perron test (Phillips and Perron (1988)) considered unit root testing in the presence of such innovations. Cointegration was introduced in Engle and Granger (1987). Techniques for analyzing cointegrating regressions in the presence of correlated innovations were considered in Saikkonen (1991, "leads and lags" or "DOLS"), Phillips and Hansen (1990, the "fully modified" estimator), and Vogelsang and Wagner (2014, "Integrated Modified Ordinary Least Squares" or "IM-OLS"). These papers form the core of the unit root literature; however, the unit root literature is by now enormous.

Throughout all this literature, the weak convergence of the partial sum process to a multiple of Brownian motion is used. The unit root assumption in time series econometrics has the advantage of analytical tractability and of being easy to grasp. Yet, both economic theory and casual economics reporting suggest that situations where a time series inherently is affected by its proximity to its historical maximum abound. Given this observation, it may be more fruitful to think of such economic time series as behaving inherently differently whenever a certain sufficiently high, endogenously determined level is exceeded.

Models of time series which have regime dependent behavior have been introduced in Caner and Hansen (2001); there, however, the regimes are determined by comparing a stationary sequence's value to a fixed level. A model of this type will be unable to capture behavior which differs only near a historical maximum, as a historical maximum will be non-stationary.

In this paper, we attempt to extend the unit root model in a simple way as to allow different behavior of the time series once an endogenously determined bound has been exceeded, while at the same time retaining analytical tractability. A simple way to capture this type of effect in a unit root type setting is to quantify the notion that $Y_t = Y_{t-1} + \varepsilon_t$ for "historically low" values of Y_{t-1} and $Y_t = c + Y_{t-1} + \varepsilon_t$ for "historically high" values. The jump c can take both positive and negative values, and for the case $c = 0$, the standard unit root process is obtained as a special case. We introduce a process that follows this idea, and show that it converges weakly to a limit that is not Brownian motion when $c \neq 0$. Our "anxious unit root process" does not depend on sample size in any way, and therefore there is no analogy to the literature on near unit root processes (see, for example, Phillips (1987) and Phillips (1988)).

It is easy to envision that the type of behavior modeled by the anxious unit root process can play a role in economic time series. For example, a house price time series may display a certain "nervousness" once a historic height is reached. House prices may accelerate once they reach a historic height. It is also conceivable that house prices instead tend to decrease at such a time. These situations can be modeled by $c > 0$ and $c < 0$ respectively.

The plan for this paper is as follows. Section 2 introduces the data-generating process for the anxious unit root process and discusses its basic properties. In section 3 the weak limit of this process is found. The presence of the jumps changes the invariance principle and a limit process that does not equal standard Brownian motion is obtained. Instead of weak convergence to Brownian motion, in our Theorem 3.2, we show weak convergence to the sum of a scaled Brownian motion plus a multiple of the running maximum of the scaled Brownian motion. In section 4 a panel test is constructed for the null hypothesis of a unit root and the alternative hypothesis of an anxious unit root, we then apply this test to a collection of macroeconomic aggregates. In section 5 we demonstrate how valid inference can be obtained in a cointegration setting using a variant of IM-OLS. All proofs are gathered in the Mathematical Appendix.

2 The model

2.1 Definition and basic ideas

Let ε_t be a weakly dependent sequence of random variables with mean 0 and variance σ^2. The idea of "anxiousness" of our process when it is relatively high is

$$Y_t = \begin{cases} \alpha + Y_{t-1} + \varepsilon_t & \text{if } \alpha + Y_{t-1} + \varepsilon_t \text{ low;} \\ \alpha + c + Y_{t-1} + \varepsilon_t & \text{if } \alpha + Y_{t-1} + \varepsilon_t \text{ high.} \end{cases}$$

This intuitive idea is formalized as follows. The "anxious unit root" model considered in this paper is

$$Y_t = \begin{cases} \alpha + Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} + \varepsilon_t \leq P_{t-1} \\ \alpha + c + Y_{t-1} + \varepsilon_t & \text{if } Y_{t-1} + \varepsilon_t > P_{t-1} \end{cases} \tag{1}$$

and

$$P_t = \begin{cases} P_{t-1} + \alpha & \text{if } Y_{t-1} + \varepsilon_t \leq P_{t-1} \\ \alpha + Y_{t-1} + \varepsilon_t + c + \eta & \text{if } Y_{t-1} + \varepsilon_t > P_{t-1} \end{cases} \tag{2}$$

where we will assume that $\eta \geq 0$, $Y_0 = 0$, and $P_0 = \eta$.
Here P_t is a latent bound on the Y_t process that causes Y_t to behave differently when it is near this bound. Obviously, if $c = 0$, the latent bound P_t is of no consequence to Y_t, and a standard unit root process results.

As long as $Y_{t-1} + \varepsilon_t$ is less than P_{t-1}, the process Y_t will follow a unit root path by Equation (1). In this situation, Equation (2) will leave P_t undisturbed, expect for the drift of α. Whenever $Y_{t-1} + \varepsilon_t$ exceeds P_{t-1}, a constant c is added to Y_t in addition to ε_t. This jump constant c causes the "anxiousness" of our process. This constant c can be zero, positive, or negative, depending on whether the process is indifferent, eager, or reluctant to enter new territory. Through Equation (2), in the case of a jump, P_t adjusts to the new value for Y_t plus a fixed amount η. The nonnegativity of η guarantees that the new bound weakly exceeds the process after a jump. The fact that at any jump time, P_t exceeds Y_t by η motivated our definition of $Y_0 = 0$ and $P_0 = \eta$.

In this paper, we will not consider the situation where P_t and Y_t have different drift rates. This has been analyzed by the authors, however it leads to different asymptotics and requires different methods of proof. We will also not analyze the case where a bound similar to P_t is also present below Y_t. We expect that this case can also be analyzed, however it will also require different methods of proof.

2.2 A representation for Y_t

Before embarking on a formal exposition of our results, we discuss some general ideas in order to facilitate the exposition below. To gain insight into the nature of this data-generating process, we start by recalling that, by Equation (1), for $t \in \mathbb{N}^+$,

$$\Delta Y_t = \alpha + \varepsilon_t + cI(Y_{t-1} + \varepsilon_t > P_{t-1}) \tag{3}$$

implying that

$$Y_t = Y_0 + \alpha t + \sum_{j=1}^{t} \varepsilon_j + c \sum_{j=1}^{t} I(Y_{j-1} + \varepsilon_j > P_{j-1}). \tag{4}$$

Defining $S_t = \sum_{j=1}^{t} \varepsilon_j$ and $N_t = \sum_{j=1}^{t} I(Y_{j-1} + \varepsilon_j > P_{j-1})$, and recalling that $Y_0 = 0$ and defining $S_0 = N_0 = 0$, we now find that

$$Y_t = \alpha t + S_t + cN_t. \tag{5}$$

In the sequel, it will be shown that in general, both S_t and N_t contribute to the asymptotic behavior of $Y_t - \alpha t$.

2.3 Properties of N_t and of the jump times

Since an investigation into the properties of N_t is key to determining the limit behavior of Y_t, we will define what we will refer to as jump times τ_j, as they will turn out to be useful in the analysis of N_t. The jump times τ_j are defined, for $j \in \mathbb{N}^+$, as

$$\tau_j = \min\{t \geq 0 : N_t = j\}, \tag{6}$$

and we define $\tau_0 = 0$.

Note that with this definition, the τ_j are not necessarily proper, that is, $P(\tau_j < \infty)$ does not necessarily equal 1. From this definition of N_t, it follows that N_t is a counting process. We also note that for any $m \geq 0$ and any t such that $\tau_m \leq t < \tau_{m+1}$ we have that $N_t = N_{\tau_m}$. Therefore, N_t is flat between jump times. Furthermore, it follows from our definitions that for $m \geq 0$, $m = N_{\tau_m}$. Given these definitions, it also follows that $N_t = \max\{m \geq 0 : \tau_m \leq t\}$. Since $I(N_t \leq n) = I(\max\{m \geq 0 : \tau_m \leq t\} \leq n)$ by definition, it also follows that

$$I(N_t \leq n) = I(\tau_n \geq t). \tag{7}$$

Furthermore, at any jump time τ_j, $j \geq 1$, we have $Y_{\tau_j - 1} + \varepsilon_{\tau_j} > P_{\tau_j - 1}$ and therefore

$$Y_{\tau_j} = \alpha + Y_{\tau_j - 1} + \varepsilon_{\tau_j} + c \quad \text{and} \quad P_{\tau_j} = \alpha + Y_{\tau_j - 1} + \varepsilon_{\tau_j} + c + \eta. \tag{8}$$

Together, these two equations imply

$$P_{\tau_j} - Y_{\tau_j} = \eta. \tag{9}$$

Therefore, at any jump point τ_j , P_{τ_j} exceeds Y_{τ_j} by η, and the next jump point after $t = \tau_j$ occurs at the first time when $Y_t - Y_{\tau_j} = S_t - S_{\tau_j}$ exceeds η. This implies that the τ_j are proper random variables if $S_t - S_{\tau_j}$ will eventually exceed η with probability 1. If the ε_t are i.i.d. and $E|\varepsilon_t| < \infty$, this property holds because the random walk is recurrent.

The jump times are key to unlocking some of the properties of N_t. This is because the τ_j are themselves a random walk in the case where ε_t is i.i.d.. Defining the time between jumps $\Delta\tau_j = \tau_j - \tau_{j-1}$, the main feature of these random variables is that they are i.i.d.:

Lemma 2.1. *Let ε_t be an i.i.d. sequence of random variables and assume that $E|\varepsilon_t| < \infty$. Then $\Delta\tau_j$ is a sequence of i.i.d. random variables and $E\tau_1 = \infty$.*

3 The invariance principle

From this section onwards we will assume that the ε_t are i.i.d. with mean 0 and variance σ^2. We will later argue, in section 6, that the same weak limit, up to a different variance term, can be obtained when the errors are assumed to be Markov. In this section we show an invariance principle by showing that under regularity conditions, $t^{-1/2}N_t$ is $O_p(1)$ and asymptotically equivalent to a multiple of $t^{-1/2}M_t$, where $M_t = \max_{1\leq j\leq t} S_j$. In order to derive this result, we introduce the following notation. We define the overshoots e_j as

$$e_j = \sum_{i=\tau_{j-1}+1}^{\tau_{j-1}+\Delta\tau_j} \varepsilon_i - \eta. \tag{10}$$

In words, e_j is the amount by which Y_t attempted to go past the boundary at the time of the j^{th} jump. We will show that these overshoots are i.i.d. and positive in the following lemma.

Lemma 3.1. $(e_j)_{j\geq 1}$ *is an i.i.d. sequence of positive random variables. If* $Var(\varepsilon_t) < \infty$ *then* $Ee_j < \infty$.

Using the previous lemmas we can show an invariance principle for $\frac{Y_{[rT]}-\alpha[rT]}{\sqrt{T}}$.

Theorem 3.2. *If* $E|\varepsilon_t|^p < \infty$ *for some* $p > 4$*, then*

$$\frac{Y_{[rT]} - [rT]\alpha}{\sqrt{T}} \Rightarrow \sigma(W(r) + \frac{c}{\eta + E(e_j)} \max_{s\in[0,r]} W(s)) = \sigma(W(r) + \tilde{c}M(r), \tag{11}$$

where $\tilde{c} = \frac{c}{\eta+E(e_j)}$ *and* $M(r) = \max_{s\in[0,r]} W(s)$.

The above limiting process is distinct from the standard Brownian motion found with unit roots. In addition, this process is not found in the limiting distributions of any standard time series models.
Careful inspection of the proof reveals that the i.i.d. assumption is only required to show a strong law of large numbers for e_j (Lemma 7.2) and tightness in $\mathcal{D}[0,1]$ for $N_{[rT]}/\sqrt{T}$. As such these could instead have been taken as the assumptions and these are likely to hold in more general cases where the error terms are not assumed to be i.i.d.

4 Hypothesis testing in a panel setting

4.1 Testing

We now consider the issue of testing whether a time series follows a unit root or an anxious unit root.

A pure time series test for the anxious unit root process is difficult to construct for this model. Since intuitively the only information on c is obtained from the N_T observations for which the boundary is exceeded, and because estimators are typically estimated root-n consistently and because $N_T = O_p(\sqrt{T})$, we suspect that an estimator of c necessarily converges no better than with rate $T^{1/4}$.

We will therefore consider testing in a panel setting. Below, we assume that Y_t^i follows an anxious unit root process with parameters η^i, α^i, c^i and $Y_0^i = \omega^i, P_0^i = \omega^i + \eta$. In addition, assume that ε_t^i has cdf F_i which is assumed to be such that $E\varepsilon_t^i = 0$ and $\text{Var}(\varepsilon_t^i) = (\sigma^i)^2$.

Let $c^i = c$. For each i, this gives $\tilde{c}^i = \frac{c}{\eta^i + E(e_j^i)}$. We therefore have , by Theorem 3.2 , that for each i

$$\frac{Y^i_{[rT]} - \alpha^i[rT] - Y^i_0}{\sigma^i\sqrt{T}} \Rightarrow W^i(r) + \tilde{c}^i M^i(r) \text{ as } T \to \infty \tag{12}$$

where $W^i(\cdot)$ is a standard Brownian motion and $M^i(r) = \max_{s\in[0,r]} W^i(s)$. We assume independence across i. We are interested in testing the null hypothesis $H_0 : c^i = 0, \ \forall i$. We observe that $\text{sign}(\tilde{c}^i) = \text{sign}(\frac{c}{\eta + E(e_j^i)}) = \text{sign}(c)$, due to the fact that $\eta + E(e_j^i) > 0$. Therefore, we can instead consider the equivalent null hypothesis $H_0 : \tilde{c}^i = 0, \ \forall i$. In order to construct a test for H_0 we will first find consistent estimators for α^i and σ^i. These will be $\hat{\alpha}^i = T^{-1}\sum_{t=1}^T \Delta Y_t^i$ and $(\hat{\sigma}^i)^2 = T^{-1}\sum_{t=1}^T(\Delta Y_t^i - \hat{\alpha}^i)^2$. These will be shown to be consistent estimators in the following lemma.

Lemma 4.1. *If $E|\varepsilon_t^i|^p < \infty$ for some $p > 4$, it follows that $\hat{\alpha}^i \xrightarrow{P} \alpha^i$ as $T \to \infty$ and $\hat{\sigma}^i \xrightarrow{P} \sigma^i$ as $T \to \infty$.*

We propose the test statistic,

$$\mathcal{J}_{N,T} = N^{-1/2}\sqrt{12}\sum_{i=1}^{N}(T^{-3/2}(\hat{\sigma}^i)^{-1}\sum_{t=1}^{T}(Y_t^i - Y_0^i - \hat{\alpha}^i t)). \tag{13}$$

The behavior of this test statistic is characterized in the following theorem.

Theorem 4.2. *If Equation (12) holds for each i, $E|\varepsilon_t^i|^p < \infty$ for some $p > 4$, and Y_t^i are independent across i, then*

1. *If H_0 holds, then $\mathcal{J}_{N,T} \xrightarrow{d} N(0,1)$ as $T \to \infty$.*
2. *If H_0 does not hold and $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \tilde{c}^i \neq 0$, then $\mathcal{J}_{N,T}/\sqrt{N} \xrightarrow{P} \sqrt{(1/18\pi)}\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \tilde{c}^i$. as $T \to \infty$ followed by $N \to \infty$.*

We note that we use the sequential asymptotics of Phillips and Moon (1999) for considering the behavior of $\mathcal{J}_{N,T}$ under the alternative hypothesis.

Note that in the case where Y_i^t is independent across i and a trend stationary process, for a fixed N, $\mathcal{J}_{N,T} = O_P(T^{-1})$. Therefore, our test will not reject H_0 in this situation.

4.2 Testing for anxious unit roots in a panel of aggregate time series

In this section we test several panels of aggregate time series for the presence of an anxious unit root and we calculate the $\mathcal{J}_{N,T}$ test statistic from Theorem (4.2). Each of the series we test are macro aggregate time series that have previously been treated as time series with unit roots in the literature. Table 1 gives test statistic values along with the data source. We used two sources for our data

1. Macroeconomic aggregates from the Penn World Table;

2. Macroeconomic aggregates that were used in recent Journal of Applied Econometrics articles on panel unit root testing or cointegration.

The data used was obtained from the Penn World Tables which can be found at `http://cid.econ.ucdavis.edu/pwt.html`. In order to limit the cross-sectional dependence, 24 countries were selected so as to decrease geographic proximity amongst the group. The selected countries were Australia, Canada, Switzerland, China, Germany, Spain, Finland, United Kingdom, Hungary, Iceland, Israel, Italy, Japan, Mexico, Malaysia, Nepal, Pakistan, Peru, Philippines, Saudi Arabia, Singapore, Thailand, Turkey, and Venezuela. We used the log of real GDP, the log of real consumption, exchange rate, and the log of the share of government consumption. The Penn World table refers to these as $rgdpna$, $rconna$, xr, and csh_g. The time period used was 1970-2014 except for the exchange rate, for which 1990-2014 was used due to data availability. See Feenstra, Inklaar, and Timmer (2015) for information on the Penn World Tables.
The data from the Journal of Applied Econometrics articles can be found at their data archieve, `http://qed.econ.queensu.ca/jae/`.

Table 1: Panel testing for anxious unit roots

Time Series	$\mathcal{J}_{N,T}$	Data Source
Log Real GDP	5.64	Penn World Tables
Log Real Consumption	6.32	Penn World Tables
Real Exchange Rate	3.81	Penn World Tables
log Share of Government Consumption	1.90	Penn World Tables
Interest Rate	-0.13	Westerlund 2008
CPI	-0.27	Westerlund and Hess 2011
Inflation	-2.08	Gengenbach et al 2016

Table 1 suggests that endogenous changes in the drift a a feature of numerous macroeconomic time series such as log(GDP), log(Consumption), Real Exchange Rates, and Inflation. The small test statistic values for CPI and Inflation can be explained by either stationary-like behavior or a situation where c is small of 0 (i.e., the unit root situation.) Given that $\mathcal{J}_{N,T}$ is roughly $N(0,1)$ under the null of a unit root, the very small values of $\mathcal{J}_{N,T}$ for these two series suggests that stationary-like behavior is likely the cause. It is unclear how to interpret the value of 1.9 of $\mathcal{J}_{N,T}$ for log Share of Government Consumption. The above result for log GDP casts doubt on the widespread practice of using log GDP as a regressor in cointegration analysis.

5 Cointegration

This section proposes a valid method of inference in a cointegration setting when a regressor follows an anxious unit root process. The model of interest for this section is

$$Y_t = \beta X_t + u_t^*, \tag{14}$$

where $X_t = S_t + cN_t$ is an anxious unit root without drift, i.e. $\alpha = 0$, and with $\Delta S_t = \varepsilon_t$. Note that u_t^* is not assumed to be uncorrelated with ε_t. We will assume that $\tilde{c} > -1$. This guarantees that $Y_{[rT]}/\sqrt{T}$ converges to a process that is not always negative. In this cointegration setting there are three commonly used methods of estimating β: Fully Modified (Phillips and Hansen 1990); Dynamic Ordinary Least Squares (Saikkonen 1991); and Integrated Modified Ordinary Least Squares (Vogelsang and Wagner 2014). Additional work can show that all three of these estimators are consistent in the anxious unit root setting, however their standard errors will be invalid. While Fully Modified and Dynamic Ordinary Least Squares can be adapted to this setting, this requires an estimate of $\tilde{c}$ and therefore we instead consider Integrated Modified Ordinary Least Squares (IM-OLS). IM-OLS proceeds by summing Equation (14), which gives

$$S_t^Y = \beta S_t^X + S_t^{u^*} \tag{15}$$

where $S_t^Y = \sum_{j=1}^t Y_j$ and S_t^X and $S_t^{u^*}$ are defined analogously. We will assume that $T^{-1/2}(S_{[rT]}, S_{[sT]}^{u^*}) \Rightarrow \tilde{W}(r,s)$, where $\tilde{W}$ is a two-dimensional scaled Brownian motion. As seen in section 3 of Vogelsang and Wanger (2014), there is some γ such that if $S_t^u = S_t^{u^*} - \gamma S_t$ then $T^{-1/2}(S_{[rT]}, S_{[sT]}^u) \Rightarrow (\sigma W(r), U(s))$, where $W(r)$ is Brownian motion and $U(r)$ is a scaled Brownian motion independent of $W(r)$. We rewrite Equation (15) as

$$S_t^Y = \beta S_t^X + \gamma X_t + (-\gamma c)N_t + S_t^u. \tag{16}$$

IM-OLS is based on a regression of S_t^X and X_t on S_t^Y in a setting where N_t is absent. If N_t were observed in the data then it could simply be added to the regression equation; however, N_t is not observed in the data. This issue is fixed by observing that asymptotically the running maximum of X_t is a multiple of N_t. This is seen in the following lemma.

Lemma 5.1. *Let* $M_{[rT]}^X = \max_{j \leq [rT]} X_j$. *If* $E|\varepsilon_t|^p < \infty$ *for some* $p > 4$ *and* $\tilde{c} > -1$, *then*

$$M_{[rT]}^X / \sqrt{T} \Rightarrow (1+\tilde{c})M(r). \tag{17}$$

In addition, for any $r \in (0,1]$

$$N_{[rT]}/M_{[rT]}^X \xrightarrow{P} \frac{\tilde{c}}{c(1+\tilde{c})}. \tag{18}$$

Therefore, we have that $S_t^Y = \beta S_t^X + \gamma X_t + \theta_t M_t + S_t^u$, where $\theta_t = -\gamma c \frac{N_t}{M_t^X} I(M_t^X > 0)$. By Lemma 5.1, $\theta_{[rT]} \Rightarrow \theta := -\gamma \frac{\tilde{c}}{1+\tilde{c}}$ on $[\delta, 1]$ for any $\delta > 0$. Our estimator will then be to regress S_t^Y on S_t^X, X_t, and M_t^X. This gives

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \\ \hat{\theta} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^T (S_t^X)^2 & \sum_{t=1}^T S_t^X X_t & \sum_{t=1}^T S_t^X M_t^X \\ \sum_{t=1}^T X_t S_t^X & \sum_{t=1}^T X_t^2 & \sum_{t=1}^T X_t M_t^X \\ \sum_{t=1}^T S_t^X M_t^X & \sum_{t=1}^T X_t M_t^X & \sum_{t=1}^T (M_t^X)^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T S_t^X S_t^Y \\ \sum_{t=1}^T X_t S_t^Y \\ \sum_{t=1}^T M_t^X S_t^Y \end{pmatrix}. \tag{19}$$

The fact that θ_t is random will cause complications in the analysis of this estimator, however the asymptotics will behave as though $\theta_t = \theta$ everywhere. We characterize the convergence of this estimator in the following theorem.

Theorem 5.2. *If* $(\frac{S_{[rT]}}{\sigma\sqrt{T}}, \frac{S_t^U}{\sqrt{T}}) \Rightarrow (W(r), U(r))$, $E|\varepsilon_t|^p < \infty$ *for some* $p > 4$, *and* $c/\eta > -1$, *then*

$$\begin{pmatrix} T(\hat{\beta} - \beta) \\ \hat{\gamma} - \gamma \\ \hat{\theta} - \theta \end{pmatrix} \xrightarrow{d} \tag{20}$$

$$\begin{pmatrix} \int_{[0,1]} \int_{[0,s]} X(r)^2 dr ds & \int_{[0,1]} (\int_{[0,s]} X(r)dr) X(s) ds & (1+\tilde{c}) \int_{[0,1]} (\int_{[0,s]} X(r)dr) M^X(s) ds \\ \int_{[0,1]} (\int_{[0,s]} X(r)dr) X(s) ds & \int_{[0,1]} X(r)^2 dr & (1+\tilde{c}) \int_{[0,1]} X(r) M^X(r) dr \\ (1+\tilde{c}) \int_{[0,1]} (\int_{[0,s]} X(r)dr) M^X(s) ds & (1+\tilde{c}) \int_{[0,1]} X(r) M^X(r) dr & (1+\tilde{c})^2 \int_{[0,1]} M^X(r)^2 dr \end{pmatrix}^{-1}$$

$$
\times \begin{pmatrix} \int_{[0,1]}(\int_{[0,s]} X(r)dr)U(s)ds \\ \int_{[0,1]} X(r)U(r)dr \\ (1+\tilde{c})\int_{[0,1]} M^X(r)U(r)dr \end{pmatrix}, \tag{21}
$$

where $M^X(r) = \max_{s\in[0,r]} \sigma W(s)$, $X(r) = \sigma(W(r) + \tilde{c}M(r))$, and $U(r)$ is a scaled Brownian motion independent of $W(r)$.

Theorem 5.2 implies that standard asymptotic inference is valid for $\hat{\beta}$ because the limit distribution is a mixed normal. Note that the resulting estimation procedure does not require an estimation of $\tilde{c}$. Theorem 5.2 illustrates that for series which may have an anxious unit root, such as log GDP (as seen in section 4.2), a modified version of a cointegration analysis is easy to implement. Additionally, the mixed asymptotic normality result continues to hold in the pure unit root case of $c = 0$. In that case $\tilde{c} = 0$ and $\theta = 0$, yet inference on β using Theorem 5.2 is unaffected.

6 Dependent errors

Throughout sections 3 through 5 we have assumed that the ε_t are i.i.d. In this section, we will briefly discuss how this assumption could be relaxed. Without the i.i.d. assumption Lemma 3.1 will no longer hold. However, if we assume that ε_t has the Markov property then we can see that $E_j \equiv (e_j, \varepsilon_{\tau_j})$ is a Markov chain. This is due to the strong Markov property that is obtained since $\Delta\tau_j < \infty$ a.s. Therefore, with a great deal of additional work and some additional technical assumptions on ε_t it is likely that e_j will still satisfy a strong law of large numbers, as needed in the proof of Theorem 3.2.

Inspection of the proofs also reveals that the tightness proof in Theorem 3.2 used the i.i.d. assumption, this, also, appears as though it will still hold under the assumption that ε_t satisfies the Markov property. Additional work would be required to rigorously show this, however. Given that the above holds, then Theorem 3.2 would be replaced by the following,

$$
\frac{Y_{[rT]} - \alpha[rT]}{\sqrt{T}} \Rightarrow \sigma(W(r) + \tilde{c}M(r)), \tag{22}
$$

where $\sigma^2 = \text{Var}(\varepsilon_t) + 2\sum_{t=1}^{\infty} \text{cov}(\varepsilon_1, \varepsilon_{1+t})$. We note that this will not affect the cointegration analysis of the previous section and as such adding the maximum will still be required for valid cointegration inference.

7 Conclusion

This paper introduces a new time series model that adds endogenous jumps to a standard unit root process whenever the unit root process is near a historically high level, as one might suspect to be present in series such as log GDP, house prices, inflation, exchange rate, and unemployoment rate. While distinct from the unit root model, this model is still analytically tractable, and gives rise to a novel limiting process. A simple panel testing procedure is proposed and applied to panels of aggregate time series. Evidence for an anxious unit root is found in several series, among which is log GDP. We show that cointegration analysis in the presence of an anxious unit root can be conducted with a simple modification of Vogelsang and Wagner's (2014) IM-OLS procedure.

Mathematical Appendix

Proof of Lemma 2.1:

We will first show that the $\Delta\tau_j$ all have the same distribution. This is because

$$P(\Delta\tau_j > t) = P(\max_{1\le s\le t} \sum_{i=\tau_{j-1}+1}^{\tau_{j-1}+s} \varepsilon_i \le \eta) = \sum_{l=1}^{\infty} P(\max_{1\le s\le t} \sum_{i=\tau_{j-1}+1}^{\tau_{j-1}+s} \varepsilon_i \le \eta, \tau_{j-1} = l)$$

$$= \sum_{l=1}^{\infty} P(\max_{1\le s\le t} \sum_{i=l+1}^{l+s} \varepsilon_i \le \eta, \tau_{j-1} = l), \tag{23}$$

and the two events in the last probability are independent. Therefore, the last probability can be written as

$$\sum_{l=1}^{\infty} P(\max_{1\le s\le t} \sum_{i=l+1}^{l+s} \varepsilon_i \le \eta) P(\tau_{j-1} = l) = \sum_{l=1}^{\infty} P(\max_{1\le s\le t} \sum_{i=1}^{s} \varepsilon_i \le \eta) P(\tau_{j-1} = l)$$

$$= P(\max_{1\le s\le t} \sum_{i=1}^{s} \varepsilon_i \le \eta), \tag{24}$$

which shows that the $\Delta\tau_j$ all have the same distribution. We will now show that they are pairwise independent, joint independence follows by a similar argument. Let $j < k$, $j, k \in \mathbb{N}$ and $s, t \in \mathbb{N}$. Then

$$P(\Delta\tau_j > t, \Delta\tau_k > s) = P(\Delta\tau_j > t, \max_{1\le s\le t} \sum_{i=\tau_{k-1}+1}^{\tau_{k-1}+s} \varepsilon_i \le \eta)$$

$$= \sum_{l=1}^{\infty} P(\Delta\tau_j > t, \tau_{k-1} = l, \max_{1\le s\le t} \sum_{i=l+1}^{l+s} \varepsilon_i \le \eta) = \sum_{l=1}^{\infty} P(\Delta\tau_j > t, \tau_{k-1} = l) P(\max_{1\le s\le t} \sum_{i=l+1}^{l+s} \varepsilon_i \le \eta)$$

$$= \sum_{l=1}^{\infty} P(\Delta\tau_j > t, \tau_{k-1} = l) P(\max_{1\le s\le t} \sum_{i=1}^{s} \varepsilon_i \le \eta)$$

$$= P(\Delta\tau_j > t) P(\max_{1\le s\le t} \sum_{i=1}^{s} \varepsilon_i \le \eta) = P(\Delta\tau_j > t) P(\Delta\tau_k > s).$$

Therefore, it follows that the $\Delta\tau_j$ are i.i.d.
For $E\tau_1$ we note that

$$E\tau_1 = \sum_{n=1}^{\infty} P(\tau_1 \geq n) = \sum_{n=1}^{\infty} P(M_{n-1} \leq \eta) \geq \sum_{n=1}^{\infty} P(M_{n-1} \leq 0), \tag{25}$$

and the last expression is infinite by Theorem 1a of section XII in Feller (1971), found on page 415 of volume II. □

Proof of Lemma 3.1:

The fact that $e_j \geq 0$ follows from the definition of τ_j. To show the i.i.d. property we first show identical distributions of e_j. This holds because

$$P(e_j \geq x - \eta) = P(\sum_{t=\tau_{j-1}+1}^{\tau_{j-1}+\Delta\tau_j} \varepsilon_t \geq x) = \sum_{k=1}^{\infty} P(\sum_{t=\tau_{j-1}+1}^{\tau_{j-1}+\Delta\tau_j} \varepsilon_t \geq x, \tau_{j-1} = k) \tag{26}$$

$$= \sum_{k=1}^{\infty} P(\sum_{t=k+1}^{k+\Delta\tau_j} \varepsilon_t \geq x, \tau_{j-1} = k). \tag{27}$$

By independence and Lemma (2.1) this is equal to

$$, \sum_{k=1}^{\infty} P(\sum_{t=1}^{\Delta\tau_j} \varepsilon_t \geq x) P(\tau_{j-1} = k) = P(\sum_{t=1}^{\Delta\tau_1} \varepsilon_t \geq x). \tag{28}$$

We now show pairwise independence of the e_j, joint independence follows by a similar argument. Let $j > m$. Then

$$P(e_m \geq y - \eta, e_j \geq x - \eta) = P(\sum_{t=\tau_{m-1}+1}^{\tau_m} \varepsilon_t \geq y, \sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \geq x) \tag{29}$$

$$= \sum_{k=1}^{\infty} P(\sum_{t=\tau_{m-1}+1}^{\tau_m} \varepsilon_t \geq y, \sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \geq x, \tau_{j-1} = k) \tag{30}$$

$$= \sum_{k=1}^{\infty} P(\sum_{t=\tau_{m-1}+1}^{\tau_m} \varepsilon_t \geq y, \sum_{s=k+1}^{k+\Delta\tau_j} \varepsilon_s \geq x, \tau_{j-1} = k). \tag{31}$$

By independence and Lemma (2.1), it follows that the previous line equals

$$\sum_{k=1}^{\infty} P(\sum_{t=\tau_{m-1}+1}^{\tau_m} \varepsilon_t \geq y, \tau_{j-1} = k) P(\sum_{s=k+1}^{k+\Delta\tau_j} \varepsilon_s \geq x) \tag{32}$$

$$= P(\sum_{s=1}^{\Delta\tau_1} \varepsilon_s \geq x) \sum_{k=1}^{\infty} P(\sum_{t=\tau_{m-1}+1}^{\tau_m} \varepsilon_t \geq y, \tau_{j-1} = k) = P(e_1 \geq x - \eta) P(e_m \geq y - \eta) \tag{33}$$

$$= P(e_m \geq y - \eta) P(e_j \geq x - \eta). \tag{34}$$

Therefore, the e_j are independent.
We now show that $Ee_j < \infty$ if $\mathrm{Var}(\varepsilon_t) < \infty$. We first define, for $j \in \mathbb{N}^+$, $\tilde{\tau}_j = \inf\{t > \tilde{\tau}_{j-1} : S_t > S_{\tilde{\tau}_{j-1}}\}$, $\tilde{e}_j = S_{\tilde{\tau}_j} - S_{\tilde{\tau}_{j-1}}$, and $\tilde{N} = \inf\{n \geq 1 : \sum_{j=1}^{n} \tilde{e}_j > \eta\}$ along with $\tilde{\tau}_0 = 0$. By a similar argument as Lemma 3.1 we can show that $\tilde{e_j}$ are a sequence of i.i.d. random variables. In addition, is well known (see page 249 of Doney (1980), for example) that if $\mathrm{Var}(\varepsilon_t) < \infty$ then $E\tilde{e}_j < \infty$. $\tilde{N}$ is a stopping time with respect to the natural filtration generated by $\tilde{e}_j$, and therefore if $E\tilde{N} < \infty$, by Wald's first identity, it follows that

$$ES_{\tau_1} = \eta + Ee_1 = E \sum_{j=1}^{\tilde{N}} \tilde{e}_j = E\tilde{N} E\tilde{e}_j < \infty. \tag{35}$$

We now will show that $E\tilde{N} < \infty$ to complete the argument. We first note that

$$E\tilde{e}_1 \geq E\varepsilon_1 I(\tilde{\tau}_1 = 1) = E\varepsilon_1 I(\varepsilon_1 > 0) > 0. \tag{36}$$

Therefore, there exists some $K > 0$ such that $P(\tilde{e}_1 > K) = p > 0$. We now can use this property, Markov's inequality, and the property that $\tilde{e}_j \geq 0$ to observe that

$$P(\tilde{N} > n) = P(\sum_{j=1}^{n} \tilde{e}_j \leq \eta) \leq P(\sum_{j=1}^{n} \tilde{e}_j I(\tilde{e}_j > K) \leq \eta) \tag{37}$$

$$\leq P(\sum_{j=1}^{n} KI(\tilde{e}_j > K) \leq \eta) = P(\exp(-\sum_{j=1}^{n} I(\tilde{e}_j > K)) > \exp(-\eta/K)) \tag{38}$$

$$\leq E(\exp(-\sum_{j=1}^{n} I(\tilde{e}_j > K)))/\exp(-\eta/K) = (p\exp(-1) + (1-p))^n \exp(\eta/K). \tag{39}$$

Since $p \in (0, 1]$ we have that $|p\exp(-1) + (1-p)| < 1$ and therefore $P(\tilde{N} > n) = O(\gamma^n)$ for some γ such that $|\gamma| < 1$. Therefore, $E\tilde{N} = \sum_{j=1}^{\infty} P(\tilde{N} \geq j) < \infty$. $\square$

The following lemma is crucial for relating $t^{-1/2}N_t$ to $t^{-1/2}M_t$:

Lemma 7.1. *Define*

$$R_t = M_t - N_t\eta - \sum_{j=1}^{N_t} e_j. \tag{40}$$

Then $0 \leq R_t \leq \eta$ *for every* t.

Proof of Lemma 7.1:

From Equations (10) and (40), it follows that

$$R_t = M_t - N_t\eta - \sum_{j=1}^{N_t}(\sum_{i=\tau_{j-1}+1}^{\tau_j} \varepsilon_i - \eta) \tag{41}$$

$$= M_t - N_t\eta - \sum_{j=1}^{N_t}\sum_{i=\tau_{j-1}+1}^{\tau_j} \varepsilon_i + N_t\eta = M_t - S_{\tau_{N_t}}. \tag{42}$$

We will now show that this is bounded between 0 and η.
First we show that for every $j \geq 1$ we have that $S_{\tau_j} = M_{\tau_j}$. We then proceed by induction. For $j = 1$, $S_{\tau_1} = M_{\tau_1}$ is due to the fact that τ_1 is the smallest value of t for which $S_t > \eta$, and as such it must be the maximum of $\{S_1, ..., S_{\tau_1}\}$.
Now assume that $S_{\tau_j} = M_{\tau_j}$ is true for some $j \geq 1$. Then τ_{j+1} is the smallest value of t for which $\sum_{i=\tau_j+1}^{t} \varepsilon_i > \eta$, and therefore $S_{\tau_{j+1}} - S_{\tau_j} = S_{\tau_{j+1}} - M_{\tau_j} > \eta$. In addition, since τ_{j+1} is the first time that $S_t - S_{\tau_j}$ exceeds η, we have $S_{\tau_{j+1}} - S_{\tau_j} = M_{\tau_{j+1}} - S_{\tau_j}$. This then gives us that $S_{\tau_{j+1}} = M_{\tau_{j+1}}$.
We now return to $M_t - S_{\tau_{N_t}}$. Since $S_{\tau_j} = M_{\tau_j}$ by the above argument, it follows that $S_{\tau_{N_t}} = M_{\tau_{N_t}}$, and therefore $M_t - S_{\tau_{N_t}} = M_t - M_{\tau_{N_t}}$.

Now,

$$M_t - M_{\tau_{N_t}} = \sum_{i=\tau_{N_t}+1}^{K} \varepsilon_i \tag{43}$$

for some $K \in \{\tau_{N_t}+1, ..., t\}$. Yet, if there existed a value of t for which $S_t - S_{\tau_{N_t}} > \eta$, we would have had an additional jump between τ_{N_t} and t, which would contradict the definition of τ_{N_t}. Thus, $0 \leq R_t < \eta$ as claimed. □

We now characterize the behavior of $\sum_{j=1}^{N_t} e_j$ in the following lemma.

Lemma 7.2. *If $Ee_j < \infty$ then,*

$$\frac{1}{N_t}\sum_{j=1}^{N_t} e_j \xrightarrow{as} E(e_j). \tag{44}$$

Proof of Lemma 7.2:

We first show that $N_t \xrightarrow{p} \infty$. Fix $k \in \mathbb{N}$. We have that

$$P(N_t \geq k) = P(\tau_k \leq t) = P(\sum_{i=1}^{k} \Delta\tau_i \leq t). \tag{45}$$

Thus if we take the limit at $t \to \infty$ we find

$$\lim_{t\to\infty} P(N_t \geq k) = \lim_{t\to\infty} P(\sum_{i=1}^{k} \Delta\tau_i \leq t) = 1 \tag{46}$$

because $\Delta\tau_i$ is proper and i.i.d. by Lemma 2.1. In addition, N_t is nondecreasing in t, and therefore $N_t \xrightarrow{as} \infty$. In addition, note that

$$\frac{1}{n}\sum_{j=1}^{n} e_j \xrightarrow{as} E(e_j) \tag{47}$$

by Kolmogorov's strong law of large numbers because e_j is i.i.d. and $E|e_j| < \infty$ by assumption. Therefore, since $\frac{1}{n}\sum_{i=1}^{n} e_i \xrightarrow{as} E(e_j)$ and $N_t \xrightarrow{as} \infty$, it follows that

$$\frac{1}{N_t}\sum_{j=1}^{N_t} e_j \xrightarrow{as} E(e_j). \tag{48}$$

This completes the proof. □

Proof of Theorem 3.2:

For this result we need to show convergence of finite dimensional distributions and tightness in $D[0,1]$. We first show convergence of finite dimensional distributions. That is, for every finite collection $r_1, .., r_n \in (0,1]$ we will show that

$$(\frac{Y_{[r_1T]} - \alpha[r_1T]}{\sqrt{T}}, ..., \frac{Y_{[r_nT]} - \alpha[r_nT]}{\sqrt{T}}) \xrightarrow{d}$$

$$\sigma(W(r_1) + \frac{c}{\eta + E(e_j)} \max_{s\in[0,r_1]} W(s), ..., W(r_n) + \frac{c}{\eta + E(e_j)} \max_{s\in[0,r_n]} W(s)).$$

By definition,

$$R_t = M_t - N_t\eta - \sum_{j=1}^{N_t} e_j, \tag{49}$$

and therefore

$$R_{[rT]} = M_{[rT]} - N_{[rT]}\eta - N_{[rT]}(N_{[rT]}^{-1} \sum_{j=1}^{N_{[rT]}} e_j), \tag{50}$$

which gives us

$$N_{[rT]} = \frac{M_{[rT]} - R_{[rT]}}{\eta + \frac{1}{N_{[rT]}} \sum_{j=1}^{N_{[rT]}} e_j}. \tag{51}$$

Since $Y_{[rT]} = \alpha[rT] + S_{[rT]} + cN_{[rT]}$, it follows that

$$Y_{[rT]} = \alpha[rT] + S_{[rT]} + c(\frac{M_{[rT]} - R_{[rT]}}{\eta + \frac{1}{N_{[rT]}} \sum_{j=1}^{N_{[rT]}} e_j}). \tag{52}$$

Therefore,

$$\frac{Y_{[rT]} - \alpha[rT]}{\sqrt{T}} = \frac{S_{[rT]}}{\sqrt{T}} + \frac{c}{\eta + \frac{1}{N_{[rT]}} \sum_{j=1}^{N_{[rT]}} e_j} \frac{M_{[rT]}}{\sqrt{T}} - \frac{c}{\eta + \frac{1}{N_{[rT]}} \sum_{j=1}^{N_{[rT]}} e_j} \frac{R_{[rT]}}{\sqrt{T}}. \tag{53}$$

By Lemma 7.1, $T^{-1/2} \sup_{r\in[0,1]} |R_{[rT]}| \leq T^{-1/2}\eta$. Also, because $e_j \geq 0$, $|\frac{c}{\eta + \frac{1}{N_{[rT]}} \sum_{j=1}^{N_{[rT]}} e_j}| \leq |c|/\eta$, and therefore, the last term in Equation (53) is asymptotically uniformly small.

Now, by Slutsky's theorem, the functional central limit theorem for $T^{-1/2}S_{[rT]}$ and the Continuous Mapping Theorem we have that

$$\frac{Y_{[rT]} - \alpha[rT]}{\sqrt{T}} \Rightarrow \sigma(W(r) + \frac{c}{\eta + E(e_j)} \max_{s \leq r} W(s)). \tag{54}$$

Since this argument extends to any $r_1, .., r_n \in (0, 1]$, the result now follows. Note that for $r = 0$, the finite dimensional convergence holds trivially.
We now show tightness. Since $\frac{Y_{[rT]}-\alpha[rT]}{\sqrt{T}} = \frac{S_{[rT]}}{\sqrt{T}} + c\frac{N_{[rT]}}{\sqrt{T}}$, and because it is well known that $\frac{S_{[rT]}}{\sqrt{T}}$ is tight under our assumptions (see e.g. Davidson (1994)), we will only need to show tightness for $N_{[rT]}/\sqrt{T}$. We define $N_T(r) = N_{[rT]}$. By Theorem 28.12 of Davidson (1994), this entails showing that for all $\eta > 0$,

$$\lim_{\delta \to 0} \limsup_{T \to \infty} P(\inf_{\Pi_\delta} \max_{1 \leq i \leq k} \max_{r,s \in [t_i, t_{i+1})} |N_T(r) - N_T(s)| \geq \eta/\sqrt{T}) = 0 \tag{55}$$

where Π_δ is the set of all partitions of $[0, 1]$ where each interval is of size $\geq \delta$.
We now define a random partition. We let $l_T = [(1/2)(\eta\sqrt{T} - 1)]$. We let

$$K_T = \max\{k \in \mathbb{N}^+ : kl_T \leq N_T\}. \tag{56}$$

We define the random partition as $\mathcal{P} = \{p_0, p_1, ..., p_{K_T}\}$ by $p_k = T^{-1}\tau_{kl_T}$ for $k < K_T$ and $p_{K_T} = 1$. We have that

$$P(\inf_{\Pi_\delta} \max_{1 \leq i \leq k} \max_{r,s \in [t_i, t_{i+1})} |N_T(r) - N_T(s)| \geq \eta/\sqrt{T}) \tag{57}$$

$$= P(\inf_{\Pi_\delta} \max_{1 \leq i \leq k} \max_{r,s \in [t_i, t_{i+1})} |N_T(r) - N_T(s)| \geq \eta\sqrt{T}, \mathcal{P} \in \Pi_\delta) \tag{58}$$

$$+P(\inf_{\Pi_\delta} \max_{1 \leq i \leq k} \max_{r,s \in [t_i, t_{i+1})} |N_T(r) - N_T(s)| \geq \eta\sqrt{T}, \mathcal{P} \notin \Pi_\delta)$$

$$\leq P(\max_{1 \leq i \leq K} \max_{r,s \in [p_i, p_{i+1})} |N_T(r) - N_T(s)| \geq \eta\sqrt{T}) + P(\mathcal{P} \notin \Pi_\delta). \tag{59}$$

Since N_t is monotone we know that the maximum difference inside each interval will have to occur at the end points, and therefore the above simplifies to

$$P(\max_{1 \leq i \leq k} N_T(p_{i+1}^-) - N_T(p_i) \geq \eta\sqrt{T}) + P(\mathcal{P} \notin \Pi_\delta), \tag{60}$$

where $N_T(p_{i+1}^-) := \lim_{s\uparrow p_{i+1}} N_T(s)$. By the definition of τ_j in Equation (6) we observe that $N_{\tau_j} = j$ and that $N_{\tau_j - 1} = j-1$. This implies that $N_T(p_{i+1}^-) - N_T(p_i) = l_T - 1$ for $i \in \{0, ..., K_T - 2\}$ and that $N_T(p_{K_T}^-) - N_T(p_{K_T-1}) < 2l_T$. Therefore, 60 is bounded from above by

$$P(2l_T \geq \eta\sqrt{T}) + P(\mathcal{P} \notin \Pi_\delta) = P(\mathcal{P} \notin \Pi_\delta).$$

We now find an upper bound for $P(\mathcal{P} \notin \Pi_\delta)$. We have that $\mathcal{P} \notin \Pi_\delta$ if and only if $\min_k(p_{k+1} - p_k) < \delta$. Therefore,

$$P(\mathcal{P} \notin \Pi_\delta) = P(\min_{0\leq k\leq K_T-1} \tau_{(k+1)l_T} - \tau_{kl_T} < \delta T, T - \tau_{(K_T-1)l_T} < \delta T). \tag{61}$$

Since $\tau_{K_T l_T} \leq T$ this is bounded above by

$$P(\min_{0\leq k\leq K_T} \tau_{(k+1)l_T} - \tau_{kl_T} < \delta T) \tag{62}$$

We let $\gamma \in \mathbb{N}^+$. The above probability is equal to

$$P(\min_{0\leq k\leq K_T} \tau_{(k+1)l_T} - \tau_{kl_T} < \delta T, K_T \leq \gamma) + P(\min_{0\leq k\leq K_T} \tau_{(k+1)l_T} - \tau_{kl_T} < \delta T, K_T > \gamma). \tag{63}$$

We bound this from above by the following

$$P(\min_{0\leq k\leq \gamma} \tau_{(k+1)l_T} - \tau_{kl_T} < \delta T) + P(K_T > \gamma) \tag{64}$$

$$= 1 - P(\tau_{(k+1)l_T} - \tau_{kl_T} \geq \delta T, \forall k \in \{0, .., \gamma\}) + P(K_T > \gamma). \tag{65}$$

Since the $\Delta\tau_j$ are i.i.d., as seen in Lemma 2.1, this is equal to

$$1 - P(\tau_{l_T} \geq \delta T)^\gamma + P(K_T > \gamma) = 1 - P(\sum_{i=1}^{l_T} \Delta\tau_i \geq \delta T)^\gamma + P(K_T > \gamma) \tag{66}$$

$$= 1 - P(N_{[\delta T]} \leq l_T)^\gamma + P(K_T > \gamma) \leq 1 - P(N_{[\delta T]} \leq l_T)^\gamma + P(N_T/l_T > \gamma), \tag{67}$$

where the last line follows from the fact that $K_T l_T \leq N_T$. Therefore, using the above upper bound, the finite dimensional convergence of $N_{[rT]}/\sqrt{T}$, and that $\lim_{T\to\infty} l_t/\sqrt{T} = (1/2)\eta$ it follows that for all $\gamma > 0$,

$$\limsup_{\delta\to 0} \limsup_{T\to\infty} P(\inf_{\Pi_\delta} \max_{1\leq i\leq k} N_T(t_{i+1}^-) - N_T(t_i) \geq \eta\sqrt{T}) \tag{68}$$

$$\leq \limsup_{\delta\to 0}(1 - P(\frac{\sigma M(\delta)}{\eta + E(e_j)} \leq (1/2)\eta)^{\gamma} + P(\frac{\sigma M(1)}{\eta + E(e_j)} \geq (1/2)\gamma\eta)) \tag{69}$$

$$= P(\frac{\sigma M(1)}{\eta + E(e_j)} \geq (1/2)\gamma\eta). \tag{70}$$

Taking the limit as $\gamma \to \infty$, we obtain

$$\limsup_{\delta\to 0}\limsup_{T\to\infty} P(\inf_{\Pi_\delta}\max_{1\leq i\leq k} N_T(t_{i+1}^-) - N_T(t_i) \geq \eta\sqrt{T}) = 0. \tag{71}$$

This completes the proof that $N_{[rT]}/\sqrt{T}$ is tight in $D[0,1]$. □

Proof of Lemma 4.1:

For $\hat{\alpha}^i$ we have, by Theorem 3.2,

$$\hat{\alpha}^i = T^{-1}(Y_T^i - Y_0^i) = T^{-1}(\alpha^i T + S_T^i + cN_T^i - \omega^i) = \alpha^i + O_P(T^{-1/2}). \tag{72}$$

Therefore, $\hat{\alpha}^i \xrightarrow{P} \alpha^i$ as $T \to \infty$. For $(\hat{\sigma}^i)^2$ we observe that

$$T^{-1}\sum_{t=1}^{T}(\Delta Y_t^i - \hat{\alpha}^i)^2 = T^{-1}\sum_{t=1}^{T}(\alpha^i - \hat{\alpha}^i + \varepsilon_t^i + c\Delta N_t^i)^2 \tag{73}$$

$$= T^{-1}\sum_{t=1}^{T}[(\alpha^i - \hat{\alpha}^i)^2 + 2(\alpha^i - \hat{\alpha}^i)(\varepsilon_t^i + c\Delta N_t^i) + (\varepsilon_t^i + c\Delta N_t^i)^2 \tag{74}$$

$$= (\alpha^i - \hat{\alpha}^i)^2 + 2(\alpha^i - \hat{\alpha}^i)Y_T^i/T + T^{-1}\sum_{t=1}^{T}(\varepsilon_t^i + c\Delta N_t^i)^2. \tag{75}$$

The first two terms in Equation (75) converge in probability to 0 as $T \to \infty$ since both $\alpha^i - \hat{\alpha}^i$ and Y_T^i/T go to 0 in probability as $T \to \infty$. Therefore, we look at the last term.

$$T^{-1}\sum_{t=1}^{T}[(\varepsilon_t^i)^2 + 2c\varepsilon_t^i\Delta N_t^i + c^2\Delta N_t^i] \tag{76}$$

$$= T^{-1}[\sum_{t=1}^{T}(\varepsilon_t^i)^2 + 2c\sum_{j=1}^{T}\varepsilon_t^i \Delta N_t^i + c^2 N_T^i)] \tag{77}$$

We have that $T^{-1}N_T^i = O_P(T^{-1/2})$ by Theorem 3.2. For the middle term we have that

$$|T^{-1}\sum_{t=1}^{t}\varepsilon_t^i \Delta N_t^i| \leq T^{-1}N_T^i \max_{1\leq t\leq T}|\varepsilon_t^i| = O_P(T^{-1/2}\max_{1\leq t\leq T}|\varepsilon_t^i|) = o_p(1), \tag{78}$$

where $T^{-1/2}\max_{1\leq t\leq T}|\varepsilon_t^i| = o_P(1)$ since $E|\varepsilon_t^i|^p < \infty$ for some $p > 4$. Therefore, it follows that

$$(\hat{\sigma}^i)^2 = T^{-1}\sum_{t=1}^{T}(\varepsilon_t^i)^2 + O_P(T^{-1/2}) + o_P(1), \tag{79}$$

therefore $\hat{\sigma}^i \xrightarrow{P} \sigma^i$ as $T \to \infty$. □

Lemma 7.3. *For each i we let $\mathcal{S}_T^i = T^{-3/2}(\hat{\sigma}^i)^{-1}\sum_{t=1}^{T}(Y_t^i - \hat{\alpha}^i t - Y_0^i)$. Under the assumptions of section 4, we have that*

$$\mathcal{S}_T^i \xrightarrow{d} \int_0^1 (W^i(r) + \tilde{c}^i M^i(r) - rW^i(1) + r\tilde{c}^i M^i(r))dr, \quad \textit{as } T \to \infty. \tag{80}$$

In addition, under H_0 $\tilde{c}^i = 0$, we have, for each i,

$$\mathcal{S}_T^i \xrightarrow{d} N(0, 1/12), \quad \textit{as } T \to \infty. \tag{81}$$

Proof of Lemma 7.3 :

We first note that for each i we have that

$$\frac{Y_{[rT]}^i - \hat{\alpha}^i - Y_0^i[rT]}{\sqrt{T}} = \frac{Y_{[rT]}^i - \alpha^i[rT] - Y_0^i}{\sqrt{T}} + r\sqrt{T}(\alpha^i - \hat{\alpha}^i) \tag{82}$$

$$= \frac{Y_{[rT]}^i - \alpha^i[rT] - Y_0^i}{\sqrt{T}} + r\sqrt{T}(\alpha^i - \frac{Y_T^i - Y_0^i}{T}) \tag{83}$$

$$= \frac{Y^i_{[rT]} - \alpha^i[rT] - Y^i_0}{\sqrt{T}} - r\frac{Y^i_T - \alpha^i T - Y^i_0}{\sqrt{T}} \Rightarrow \sigma^i(W^i(r) + \tilde{c}^i M^i(r) - rW^i(1) + r\tilde{c}^i M^i(1)), \quad (84)$$

where the above convergence follows from Theorem 3.2. Therefore an application of Theorem 3.2 and the Continuous Mapping Theorem gives the desired convergence.
Under H_0 we have that $\tilde{c}^i = 0$. Therefore, by the above result, we have that

$$\mathcal{S}^i_T \xrightarrow{d} \int_0^1 (W^i(r) - rW^i(1))dr. \quad (85)$$

It is clear that $\mathcal{S}^i_T$ is asymptotically normal. We first note that

$$E(\int_0^1 (W^i(r) - rW^i(1))dr) = 0, \quad (86)$$

since $E(W^i(r)) = 0$ for all r. Also,

$$E(\int_0^1 W^i(r) - rW^i(1)dr)^2 = E(\int_0^1 W^i(r)dr - W^i(1)/2)^2 \quad (87)$$

$$= E[(\int_0^1 W^i(r)dr)^2 - W^i(1)\int_0^1 W^i(r)dr + W^i(1)^2/4] \quad (88)$$

$$= 1/3 - \int_0^1 EW(1)W(r)dr + 1/4 \quad (89)$$

The last equality is due to the fact that $\int_0^1 W^i(r)dr \overset{d}{\sim} N(0, 1/3)$ and $W^i(1) \overset{d}{\sim} N(0, 1)$. Therefore, the above is equal to

$$1/3 - \int_0^1 rdr + 1/4 = 1/3 - 1/2 + 1/4 = 1/12. \quad (90)$$

This gives us that

$$\mathcal{S}^i_T \xrightarrow{d} N(0, 1/12), \quad \text{as T} \to \infty. \quad (91)$$

□

Lemma 7.4. *We have that*

$$E(\int_0^1 (W^i(r) + \tilde{c}^i M^i(r) - rW^i(1) + r\tilde{c}^i M^i(r))dr) = \tilde{c}^i\sqrt{1/(18\pi)}. \quad (92)$$

Proof of Lemma 7.4 :

$$E[\int_0^1 [(W^i(r) - rW^i(1)) + \tilde{c}^i(M^i(r) - rM^i(1))]dr] = \tilde{c}^i E[\int_0^1 (M^i(r) - rM^i(1))dr] \quad (93)$$

$$= \tilde{c}^i \int_0^1 EM^i(r)dr - (\tilde{c}^i/2)EM^i(1). \quad (94)$$

It can easily be seen by the reflection principle for Brownian motion that $EM^i(r) = \sqrt{2r/\pi}$. Therefore, the above equals

$$\tilde{c}^i \int_0^1 \sqrt{2r/\pi}dr - (\tilde{c}^i/2)\sqrt{2/\pi} = \tilde{c}^i\sqrt{2/\pi}(2/3 - 1/2) = \tilde{c}^i\sqrt{1/(18\pi)}. \quad (95)$$

□

Proof of Theorem 4.2 :

We observe that

$$\mathcal{J}_{N,T} = N^{-1/2}\sqrt{12}\sum_{i=1}^{N} \mathcal{S}_T^i, \quad (96)$$

where $\mathcal{S}_T^i$ is defined in Lemma 7.3. Under H_0 this is asymptotically a sum of independent $N(0,1)$ random variables by assumption and Lemma 7.3, therefore $\mathcal{J}_{N,T} \xrightarrow{d} N(0,1)$ as $T \to \infty$.

Under H_A the result follows from Lemma 7.4 and the weak law of large numbers. □

Proof of Lemma 5.1:

Since

$$\frac{M^X_{[rT]}}{\sqrt{T}} = \max_{0\le j\le [rT]} \frac{X_j}{\sqrt{T}}, \quad (97)$$

the Continuous Mapping Theorem and Theorem 3.2 imply that

$$\frac{M^X_{[rT]}}{\sqrt{T}} \Rightarrow \max_{s\in[0,r]} \sigma(W(s) + \tilde{c}M(s)). \quad (98)$$

By assumption $\tilde{c} > -1$ and therefore

$$\max_{s\in[0,r]} \sigma(W(s) + \tilde{c}M(s)) \leq \max_{s\in[0,r]} \sigma(M(s) + \tilde{c}M(s)) = \max_{s\in[0,r]} (1+\tilde{c})\sigma M(s) \tag{99}$$

$$= (1+\tilde{c})\sigma M(r). \tag{100}$$

In addition, if we let $s^* \in [0, r]$ be such that $W(s^*) = M(r)$ then we can observe that

$$\max_{s\in[0,r]} \sigma(W(s) + \tilde{c}M(s)) \geq \sigma(W(s^*) + \tilde{c}M(s^*) = (1+\tilde{c})\sigma M(r). \tag{101}$$

Therefore, $\max_{s\in[0,r]}(W(s) + \tilde{c}M(s)) = (1+\tilde{c})M(r)$.
We can also see, by the definition of R_t in Lemma 7.1, that

$$T^{-1/2}(M^X_{[rT]}, cN_{[rT]}) = T^{-1/2}(M^X_{[rT]}, c\frac{M^X_{[rT]} - r_{[rT]}}{\eta + \frac{1}{N_{[rT]}}\sum_{j=1}^{N_{[rT]}} e_j}). \tag{102}$$

Therefore, by Lemma 7.1,

$$T^{-1/2}(M^X_{[rT]}, cN_{[rT]} \xrightarrow{d} \sigma((1+\tilde{c})M(r), \tilde{c}M(r)). \tag{103}$$

Since $M(r) > 0$ $a.s.$ for any $r > 0$ we have, by the Continuous Mapping Theorem, that

$$M^X_{[rT]}/N_{[rT]} \xrightarrow{P} c(1+\tilde{c})/\tilde{c}. \tag{104}$$

□

Lemma 7.5. *If $\tilde{c} > -1$ and $E|\varepsilon_t|^p < \infty$ for some $p > 4$, then*

$$T^{-2}\sum_{t=1}^{T}(\theta_t - \theta)M^X_t S^u_t \xrightarrow{P} 0. \tag{105}$$

Proof of Lemma 7.5 :

We observe that

$$M^X_t = \max_{j\in\{1,..,t\}}(S_j + cN_j) = \max_{j\in\{1,..,t\}}(N_j\eta + cN_j + \sum_{i=1}^{N_j} e_i + \sum_{i=\tau_{N_j}+1}^{j} \varepsilon_i) \geq N_t(\eta + c), \tag{106}$$

where the above inequality comes from selecting $j = \tau_{N_t}$ and the fact that $e_i \geq 0$. Therefore,

$$N_t/M_t^X \leq (\eta + c)^{-1} < \infty, \tag{107}$$

where the last inequality follows from the assumption that $\tilde{c} > -1$. This implies that $|\theta_t| = |\gamma c \frac{N_t}{M_t^X} I(M_t^X > 0)| \leq |\gamma c(\eta + c)^{-1}|$. We now consider $T^{-2} \sum_{t=1}^{T} (\theta_t - \theta) M_t^X S_t^u$.

$$|T^{-2} \sum_{t=1}^{T} (\theta_t - \theta) M_t^X S_t^u| \leq \frac{M_T^X}{\sqrt{T}} \frac{\max_{t \in \{1,..,T\}} S_t^u}{\sqrt{T}} T^{-1} \sum_{t=1}^{T} |\theta_t - \theta| \tag{108}$$

$$= \frac{M_T^X}{\sqrt{T}} \frac{\max_{t \in \{1,..,T\}} S_t^u}{\sqrt{T}} \int_0^1 |\theta_{[rT]} - \theta| dr. \tag{109}$$

By Lemma 5.1 and Theorem 3.2 we have that $\frac{M_T^X}{\sqrt{T}} \frac{\max_{t \in \{1,..,T\}} S_t^u}{\sqrt{T}} \xrightarrow{d} (1+\tilde{c}) M(1) \max_{s \in [0,1]} U(s) = O_P(1)$. In addition, for any $\delta > 0$ we have that

$$\int_0^1 |\theta_{[rT]} - \theta| dr = \int_0^\delta |\theta_{[rT]} - \theta| dr + \int_\delta^1 |\theta_{[rT]} - \theta| dr \tag{110}$$

$$\leq \delta(|\theta| + |\gamma c(\eta + c^{-1})|) + \int_\delta^1 |\theta_{[rT]} - \theta| dr \xrightarrow{P} \delta(|\theta| + |\gamma c(\eta + c^{-1})|). \tag{111}$$

The above bound holds for any δ, therefore taking letting δ go to 0 gives that $\int_0^1 |\theta_{[rT]} - \theta| dr = o_P(1)$. Therefore, $T^{-2} \sum_{t=1}^{T} (\theta_t - \theta) M_t^X S_t^u = o_P(1)$. This completes the lemma. $\square$

Proof of Theorem 5.2:

We have that

$$\begin{pmatrix} T(\hat{\beta} - \beta) \\ \hat{\gamma} - \gamma \\ \hat{\theta} - \theta \end{pmatrix} = \begin{pmatrix} T^{-4} \sum_{t=1}^{T} (S_t^X)^2 & T^{-3} \sum_{t=1}^{T} S_t^X X_t & T^{-3} \sum_{t=1}^{T} S_t^X M_t^X \\ T^{-3} \sum_{t=1}^{T} X_t S_t^X & T^{-2} \sum_{t=1}^{T} X_t^2 & T^{-2} \sum_{t=1}^{T} X_t M_t^X \\ T^{-3} \sum_{t=1}^{T} S_t^X M_t^X & T^{-2} \sum_{t=1}^{T} X_t M_t^X & T^{-2} \sum_{t=1}^{T} (M_t^X)^2 \end{pmatrix}^{-1} \times \tag{112}$$

$$\left(\begin{pmatrix} T^{-3} \sum_{t=1}^{T} S_t^X S_t^u \\ T^{-2} \sum_{t=1}^{T} X_t S_t^u \\ T^{-2} \sum_{t=1}^{T} M_t^X S_t^u \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ T^{-2} \sum_{t=1}^{T} (\theta_t - \theta) M_t^X S_t^u \end{pmatrix} \right).$$

By Lemma 7.5, Lemma 5.1, and Theorem 3.2 the above converges in distribution to

$$\begin{pmatrix} \int_{[0,1]}\int_{[0,s]} X(r)^2 dr ds & \int_{[0,1]}(\int_{[0,s]} X(r)dr)X(s)ds & (1+\tilde{c})\int_{[0,1]}(\int_{[0,s]} X(r)dr)M^X(s)ds \\ \int_{[0,1]}(\int_{[0,s]} X(r)dr)X(s)ds & \int_{[0,1]} X(r)^2 dr & (1+\tilde{c})\int_{[0,1]} X(r)M^X(r)dr \\ (1+\tilde{c})\int_{[0,1]}(\int_{[0,s]} X(r)dr)M^X(s)ds & (1+\tilde{c})\int_{[0,1]} X(r)M^X(r)dr & (1+\tilde{c})^2\int_{[0,1]} M^X(r)^2 dr \end{pmatrix}^{-1} \tag{113}$$

$$\times \begin{pmatrix} \int_{[0,1]}(\int_{[0,s]} X(r)dr)U(s)ds \\ \int_{[0,1]} X(r)U(r)dr \\ (1+\tilde{c})\int_{[0,1]} M^X(r)U(r)dr \end{pmatrix}.$$

□

References

[1] M. Caner and B.E. Hansen. *Threshold autoregression with a unit root.* Econometrica 69 (6), 2001.

[2] J. Davidson. *Stochastic Limit Theory: An Introduction for Econometricicans.* Oxford University Press, 1994.

[3] D.A. Dickey and W.A. Fuller. *Distribution of the estimators for autoregressive time series with a unit root.* Journal of the American Statistical Association, 74, 1979.

[4] R.A. Doney. *Moments of ladder heights in random walks.* Journal of Applied Probability, 17, 1980.

[5] R.F. Engle and C.W.J. Granger. *Co-integration and error correction: Representation, estimation, and testing.* Econometrica, 55, 1987.

[6] R.C. Feenstra, R. Inklaar, and M.P. Timmer. *The next generation of the Penn World Table.* American Economic Review, 105(10), 2015.

[7] W. Feller. *An Introduction to Probability Theory and Its Applications.* Wiley, 1968.

[8] C. Gengenbach, J.P. Urbain, and J. Westerlund. *Error correction testing in panels with common stochastic trends.* Journal of Applied Econometrics, 31(6), 2016.

[9] C.W.J. Granger. *Some properties of time series data and their use in econometric model specification.* Journal of Econometrics, 16, 1981.

[10] C.W.J. Granger and P. Newbold. *Spurious regressions in econometrics.* Journal of Econometrics, 2, 1974.

[11] P.C.B. Phillips. *Understanding spurious regressions in econometrics.* Journal of Econometrics, 33, 1986.

[12] P.C.B. Phillips. *Towards a unified asymptotic theory for autoregressions.* Biometrika, 74, 1987.

[13] P.C.B. Phillips. *Regression theory for near-integrated time series.* Econometrica, 56, 1988.

[14] P.C.B. Phillips and B.E. Hansen. *Statistical inference in instrumental variables regression with I(1) processes.* The Review of Economic Studies, 57 (1), 1990.

[15] P.C.B. Phillips and H.R. Moon. *Linear regression limit theory for nonstationary panel data.* Econometrica 67(5), 1999.

[16] P.C.B. Phillips and P. Perron. *Testing for unit roots in time series regression.* Biometrika, 75, 1988.

[17] P. Saikkonen. *Asymptotically efficent estimation of cointegration regressions.* Econometric Theory, 7(1), 1991.

[18] T.J. Vogelsang and M. Wagner. *Integrated modified OLS estimation and fixed-b inference for cointegrating regressions.* Journal of Econometrics, 178(2), 2014.

[19] J. Westerlund. *Panel cointegration tests of the Fisher Effect.* Journal of Applied Econometrics, 23(2), 2008.

[20] J. Westerlund and H. Wolfgang. *A new poolability test for cointegrated panels.* Journal of Applied Econometrics, 26(1), 2011.