Re-visiting Unit Root Testing Strategies when Presence of Deterministic Trend is Uncertain

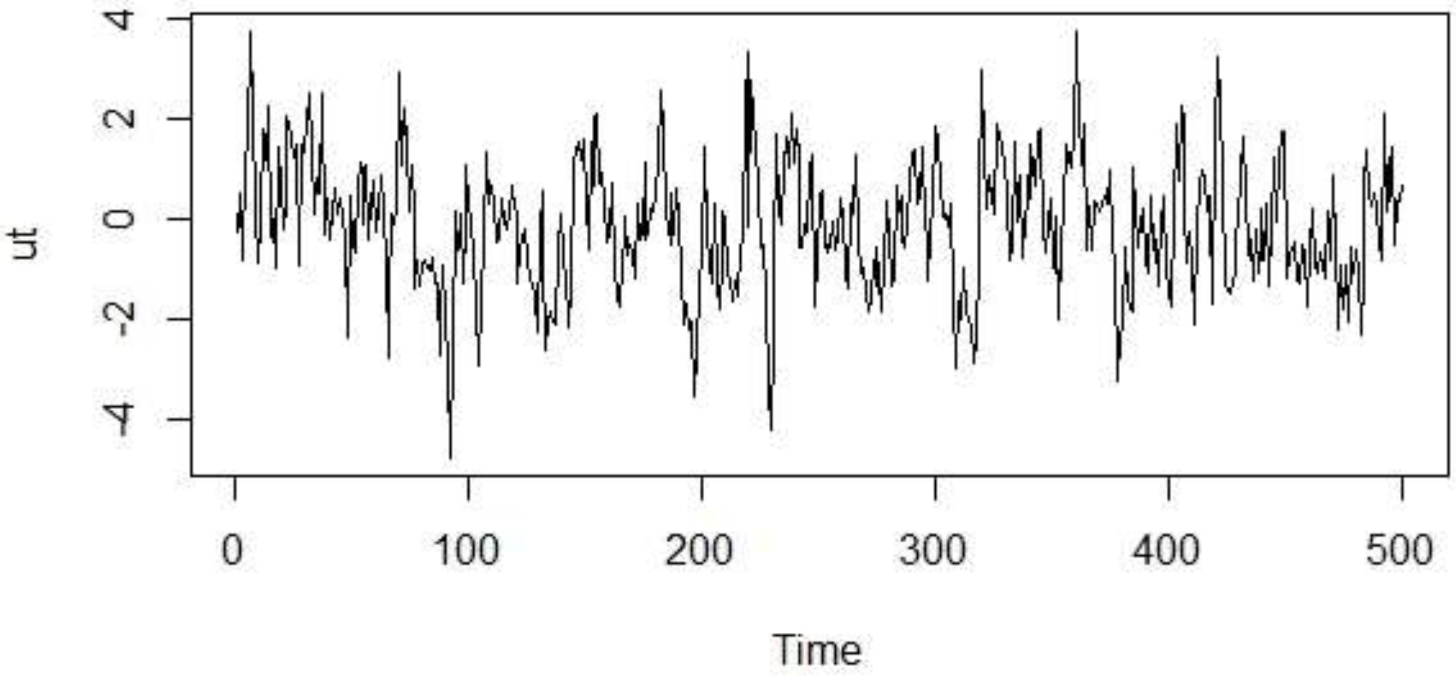
Tapan Kar

IFMR GSB, Krea University

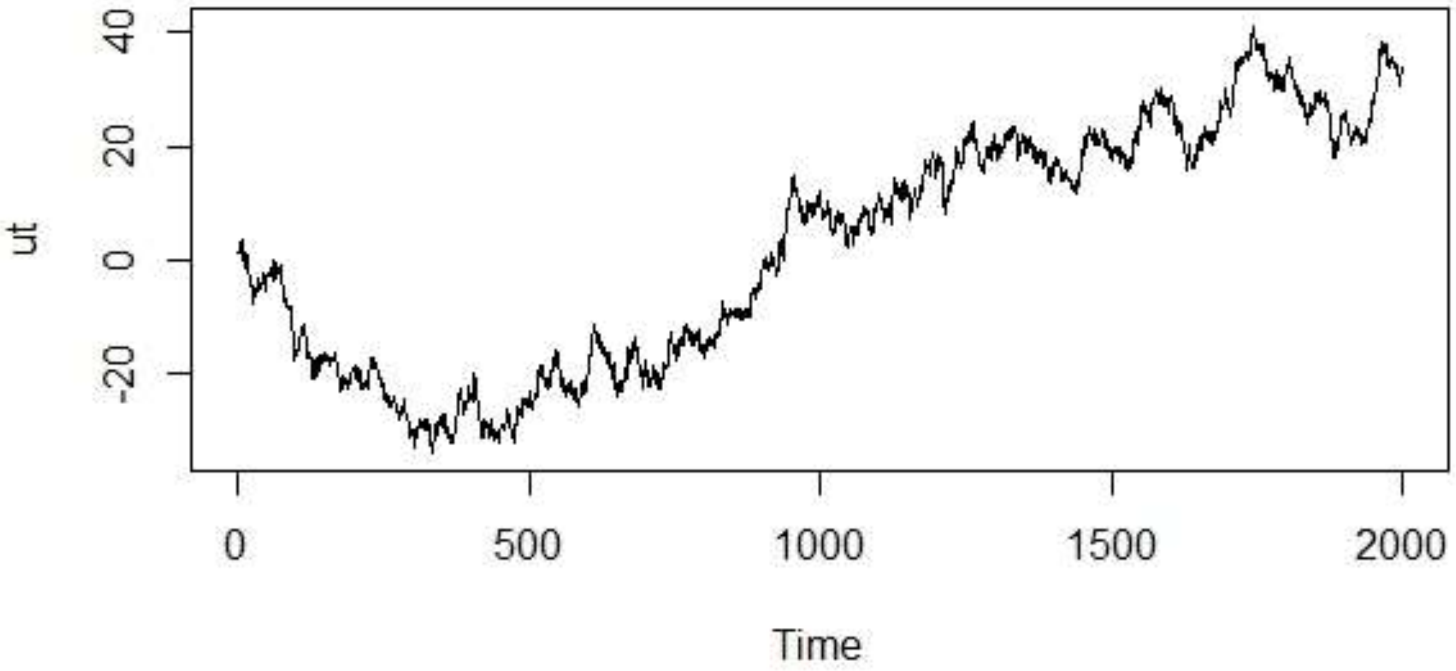
Stationarity

- Stationarity is the foundation of time series analysis.

- A time series $\{u_t\}$ is said to be weakly stationary if the mean of $\{u_t\}$ and covariance between u_t and u_{t-r} are time invariant, that is, $E(u_t) = \mu$, which is a constant (considered as a long run mean) and $Cov(u_t, u_{t-r}) = \gamma_{r,}$ which only depends on the lag r. So, for the case of weak stationarity, the first two moments are finite.



4
2
0
-2
-4
ut
0
100
200
300
400
500
Time


40
20
0
-20
ut
0
500
1000
1500
2000
Time

Properties of AR(1) Model

- Consider the $AR(1)$ model

$u_t = \alpha u_{t-1} + \varepsilon_t$, ε_t is a white noise process with variance σ_ε^2 . The process has the property

$E(u_t) = 0$, $Var(u_t) = \frac{\sigma_\varepsilon^2}{(1-\alpha^2)}$ and $Cor(u_t, u_{t-r}) = \alpha^r$.

- What will happen when $\alpha = 1$?

- when $\alpha = 1$ the 1-step ahead forecast of the $AR(1)$ model is

$$\widehat{u_t}(1) = E(u_{t+1} | u_t, u_{t-1}, \dots \dots) = u_{t.}$$

- Similarly, for any forecast horizon l

$$\widehat{u_t}(l) = u_t.$$

- forecast error defined by $e(l)$

$$e(l) = \varepsilon_{t+l} + \varepsilon_{t+l-1} + \cdots + \varepsilon_{t+1.}$$

- $Var[e(l)] = l\sigma_\varepsilon^2$ so, the variance of the forecast error will diverge to infinity.

- The above results clearly show that when $\alpha = 1$ the future observation is not predictable, and the forecast error will diverge to infinity as the forecast horizon will tend to infinity.

- Can $|\alpha|$ be greater than 1? (AR(1) model will become unstable.)

- Desired Hypothesis is H_0: α = 1 vs H_1: $|\alpha|$ < 1.

- Why is this study called unit root test ? (since at $\alpha = 1$, the root of the equation

$(1 - \alpha z) = 0$ is 1).

Existing Unit Root Tests

- Unit root test has a vast literature. In the last three decades, this topic has been studied in great detail. Here I have mentioned few of them.
- Dickey and Fuller (1979), Phillips and Perron (1988) , Schmidt and Phillips (1992), Kawiatkowski *et al*. (1992), Max test of Leybourne (1995), Pantula. *et al*. (1994), ERS (1996) and Ng and Perron (2001) etc.
- What are the issues with them? (Power of the test is very low especially when the autoregressive coefficient α is close to one.)
- What is power of a test ?
- Following Chan and Wei (1987) and Phillips (1987) we reparametrize by

$$\frac{C}{N} = \alpha - 1,\ N \text{ is the sample size.}$$

- Many Macroeconomic variables, for example, Gross National Product (GNP), Gross Domestic Product (GDP), and Consumer Price Index (CPI) exhibit a linear deterministic trend while Interest Rate (IR) and Real Exchange Rate (RER) may or may not show a linear deterministic trend. When testing for the unit root in practice, one must choose the model with drift or the model with time trend to compute the test statistic.

The DGP

We assume the following data generating process (DGP) for our analysis

$$y_t = a_0 + at + u_t, \quad u_t = \alpha u_{t-1} + \varepsilon_t.$$

The initial condition u_0 satisfies $u_0 = 0$. In case of no time trend $a = 0$.

The problem of low power gets further compounded when there is uncertainty about whether or not a linear deterministic trend is present in the data.

Assumption 1 : The errors ε_t are independently and identically distributed with expectation zero and finite variance σ_ε^2.

- Marsh (2007) argues that if deterministic trend is not present in the data and the test statistic is computed by including the time trend in the model, then the power of the test will be low as compared to that of the test with drift only model.
- So, an efficient rule is: if the deterministic trend is present in the data, one must include a deterministic trend term in the model to compute the test statistic; and if it is absent, one must use only the drift term in the model to compute the test statistic.
- However, whether or not a deterministic trend is present in the data is unknown. To address this issue, it is crucial to develop a strategy, which helps practitioners to ensure that they are using the appropriate test for their data and minimize the risk of power loss.

- Ayat and Burridge (2000), Bunzel and Vogelsang (2005), and Harvey, Leybourne, and Taylor (2007) and Harvey, Leybourne, and Taylor (2009) have studied this important problem.
- Harvey *et al*. (2009) showed that their union-based rejection strategy enjoys fairly good small sample performance and accurately takes care of the size of the test among all the procedures proposed by other researchers.

A Proposal to Address the Power Issue and uncertainty over trend

- For estimating the parameters of the DGP, We assume that the error process follows Johnson SU distribution (Kar and Bhattacharyya 2022)

- We use M-estimation technique for parameter estimation.

- We propose a new strategy using M-estimation.

Johnson SU distribution

The probability density function (PDF) of Johnson SU distribution is given by

$$g(x) = \frac{\delta}{\lambda\sqrt{2\pi}} R\left(\frac{x-\xi}{\lambda}\right) \exp\left\{-\frac{1}{2}\left[\gamma + \delta V\left(\frac{x-\xi}{\lambda}\right)\right]^2\right\}, -\infty < x < \infty$$

$$R(y) = \frac{1}{\sqrt{(y^2+1)}} \text{ and } V(y) = \log\left[y + \sqrt{(y^2+1)}\right].$$

The mean is defined by $\mu = \xi - \lambda\Theta^{\frac{1}{2}}\sin h(\Phi)$, where $\Theta = e^{\delta^{-2}}$ and $\Phi = \frac{\gamma}{\delta}$.

When mean is zero $\xi = \lambda\Theta^{1/2}\sin h(\Phi)$.

CONT...............

- It is a distribution with four parameters.

- What is the advantage with Johnson SU distribution ?

- The parameters γ and δ control skewness and kurtosis. Accordingly, the distribution is positively (negatively) skewed as γ is negative (positive). Increasing δ, holding γ constant, reduces the kurtosis. Johnson SU distribution can capture a wide range of shapes depending on its parameter values.

Derivation of test statistic

For the DGP with no time trend, the negative of log-likelihood function evaluated at $\frac{C}{N}$, conditional on the first observation is

$$ML^{drift} = \sum_{t=2}^{N} h(\, y_t - y_{t-1} - \frac{C}{N} y_{t-1} - l)$$

Where $h = -log(g)$ (g = Johnson SU pdf) and $l = -\, a_0 \frac{C}{N}$.

- The optimal $\bar{C},$ is then given by

$$\bar{C} = argmin_{(C,l)}\left(ML^{drift}\right)$$

- For the DGP with time trend, the negative of log-likelihood function (using Johnson SU as a reference density) evaluated at $\frac{C}{N}$, conditional on the first observation, is
- $ML^{\text{trend}} = \sum_{t=2}^{N} h\left(y_t - y_{t-1} - \frac{C}{N} y_{t-1} - b_1 - \, b_2 \frac{(t-1)}{N}\right)$
- where $h = \log(g)$, $b_1 = a - \left(\frac{a_0 C}{N}\right)$ and $b_2 = \, -Ca$

CONT……………….

Model with drift the test statistic is

$$t^{drift} = \frac{\sum(y_{t-1} - \overline{Y}\,)\,\psi(\widehat{\varepsilon_t})}{\sigma_\psi[\,\sum(\,y_{t-1} - \overline{Y})^2]^{\frac{1}{2}}}$$

$\psi(x)$ is the first derivative of $h(x)$. $\overline{Y}$ is the sample mean. $\widehat{\varepsilon_t} = (y_t - y_{t-1} - \omega_0)$, be the vector of residuals, where ω_0 is defined by

- $\omega_0 = \text{argmin}_{a_0} \sum_{t=2}^{N} h(y_t - y_{t-1} - a_0)$

Model with trend the test statistic is $t^{trend} = \frac{\sum_{t=2}^{N}(r_t)\psi(\widehat{\varepsilon_t})}{\widehat{\sigma_\psi}[\,\sum r_t{}^2]^{\frac{1}{2}}}$

- $\widehat{\varepsilon_t} = \left(y_t - y_{t-1} - \omega_0 - \frac{\omega_1 t}{N}\right)$
- $(\omega_0, \omega_1) = \text{argmin}_{a_0,a} \sum_{t=2}^{N} h\left(y_t - y_{t-1} - a_0 - \frac{at}{N}\right)$
- Where r_t is the residual obtained from a least square regression of y_{t-1} on $(1, t)$.

Asymptotic Distribution

- Null hypothesis will be rejected for small value of t^{drift}.

- Our task is to find the asymptotic distribution of t^{drift} and t^{trend}
- Further, I define the following:
- $\varphi(x)$ is the second derivative of $h(x)$.

$\omega = E[\varphi(\varepsilon_t)], \sigma_\varepsilon^2 = \text{Var}(\varepsilon_t), \rho = \text{Corr}\big(\varepsilon_t, \psi(\varepsilon_t)\big), \ \sigma_\psi^2 = \text{Var}\,[\psi(\varepsilon_t)]$, and

$\vartheta = \frac{\sigma_\varepsilon \omega}{\sigma_\psi}$.

- **Assumption 2.** The function $h(x)$ is continuously differentiable, and its second and higher order derivatives are bounded.
- **Assumption 3.** $E[\psi(\varepsilon_t)] = 0$.

- **Theorem 1:** Following Rothenberg and Stock (1996) and Xiao (2001), the asymptotic distribution of t^{drift} will converge to

$$Ƚ_C \equiv \rho \frac{T_C}{\sqrt{R_C}} + \sqrt{(1-\rho^2)} \left(\frac{\int_0^1 D_C(r) dW_1(r)}{\sqrt{R_C}} \right) + \vartheta C \sqrt{R_C}$$

- $D_C(r) = W_C(r) - \int_0^1 W_c(s)ds, T_C = \int_0^1 D_C\,(r)dW(r)$ and $R_C = \int_0^1 D_C(r)^2\,dr$.
- Let W_0 be a standard Brownian motion defined on $[0, 1]$ and $W_C(.)$ be a related diffusion process
- $W_c(t) = \int_0^t \exp\big(c(t-s)\big)\,dW_0(s)$

- **Theorem 2**: $t^{trend} \equiv \rho \frac{T_C}{\sqrt{R_C}} + \sqrt{(1-\rho^2)} \left(\frac{\int_0^1 D_C(r) dW_1(r)}{\sqrt{R_C}} \right) + \vartheta C \sqrt{R_C}$, where
- $D_C(r) = W_C(r) - 2\int_0^1 \big(2 - 3s - r(3-6s)\big) W_C(s) ds, T_C = \int_0^1 D_C(r) dW(r)$, and $R_C = \int_0^1 D_C(r)^2 \, dr$

CONT…….

- T_C, R_C and D_C are functionals of Brownian motions.

- Where is the source of power improvement ? (Since the alternative hypothesis is one-(left) sided, the rejection zone is on the left tail of the distribution of the test statistic.)

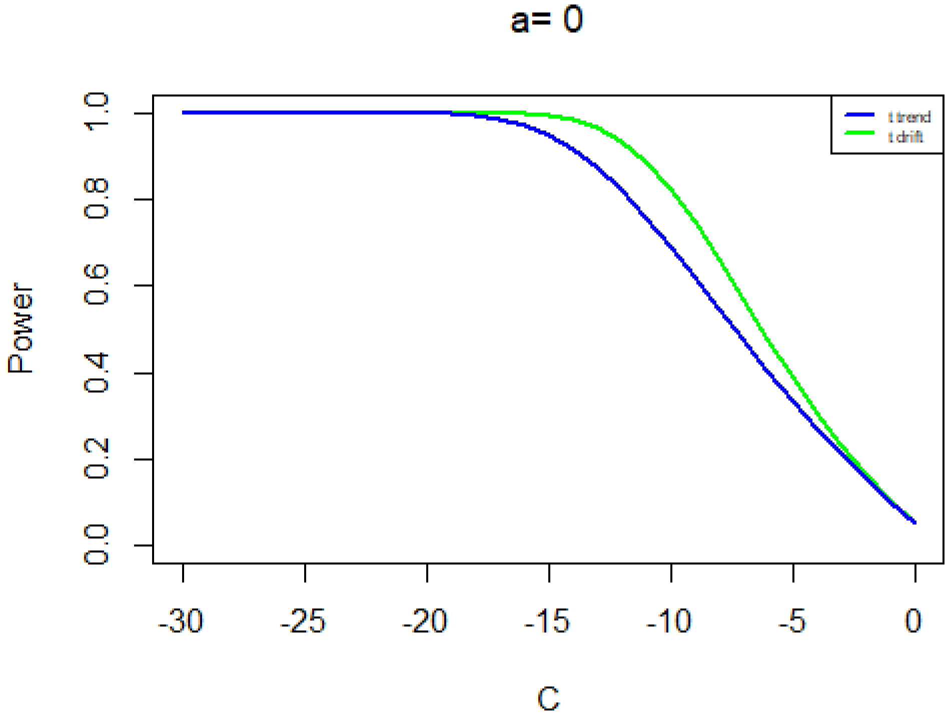
Calculation of Critical value

- According to the asymptotic distribution the critical value (i.e. under null C= 0) is dependent on ρ.
- Following Thompson (2004b), I approximate the $Q(\rho)$% critical value by a third order polynomial in $(1-\rho)$ given below.

$$Q(\rho) = A_0 + A_1(1-\rho) + A_2(1-\rho)^2 + A_3(1-\rho)^3$$

- ρ has to be estimated from the data set.

Without trend



With trend

- **Lemma 1:** y_t follows the process defined in DGP with $a = K\sigma_\varepsilon$. Assumptions 1-3 hold, then, under $H_0, t^{drift} \Rightarrow N(0,1)$
- Therefore, asymptotic size of the t^{drift} test under fixed trend $(a = K\sigma_\varepsilon)$ is defined by $Pr(Z < cri^{drift})$.
- For $\rho = 0$, the critical value of t^{drift} at 5% significance level is -1.64, resulting in an asymptotic size of 0.05. Similarly, for $\rho = 0.5$ and $\rho = 1$, the critical values of t^{drift} are -2.34 and -2.85, respectively, with corresponding asymptotic sizes of 0.009 and 0.002.

Proposed Strategy

- Our decision rule is Reject H_0 if $\{t^{drift} < \ cri^{drift} \ \ or \ \ t^{trend} <$

- By continuous mapping theorem, S converges weakly to T_S, where
- $T_S \equiv \ min\left(Ƒ_C^{drift}, \frac{Ƒ_C^{trend}\ cri^{drift}}{cri^{trend}}\right)$
- $Ƒ_C^{drift}$ and $Ƒ_C^{trend}$ are the limiting distributions of t^{drift} and t^{trend}
- However, this method rejects the null when any of the tests rejects it. So, Bonferroni bound could be invoked, and we need to calculate the critical value of T_S at the significance level $2q$.
- So, we have used a scaling constant $\lambda_q > 1$ to the critical values cri^{drift} and cri^{trend} to keep the overall asymptotic size of the test statistic at q. The test statistic, say U, is defined below.

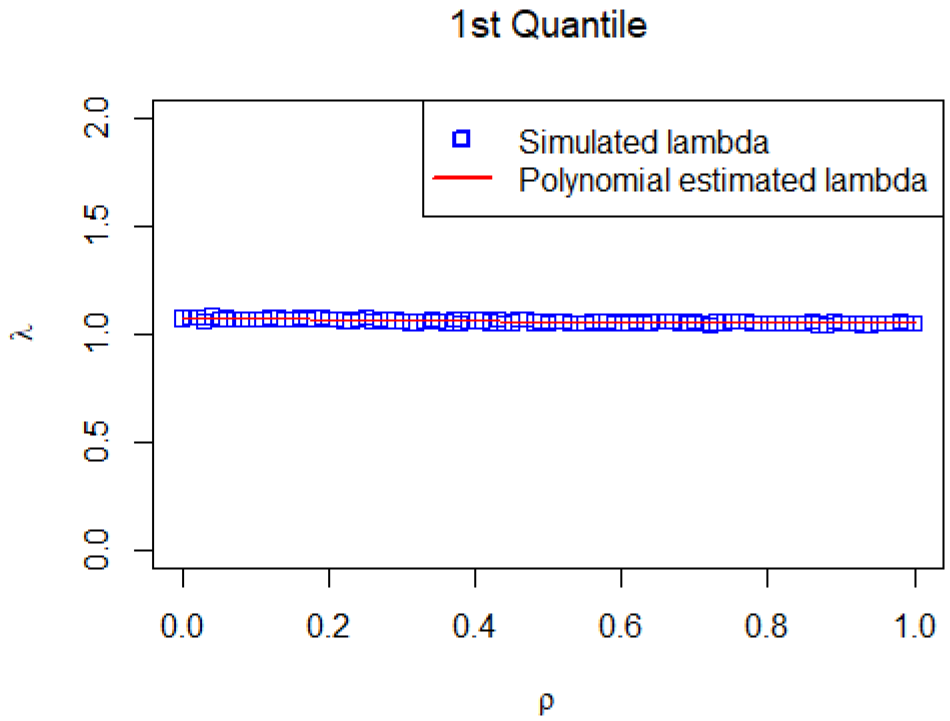
- $U = \ t^{drift} 1_{drift}\left(t^{drift} < \ \lambda_q \ cri^{drift}\right) + \ t^{trend} 1_{trend}\left(t^{trend} <\right.$

- First, we simulate the distribution of $T_S \equiv \; min\left(Ƒ_0^{drift}, \frac{Ƒ_0^{trend} \; cri^{drift}}{cri^{trend}}\right)$ (as under null C is zero) and compute the qth quartile of T_S, let us denote it by cri^{T_S}.

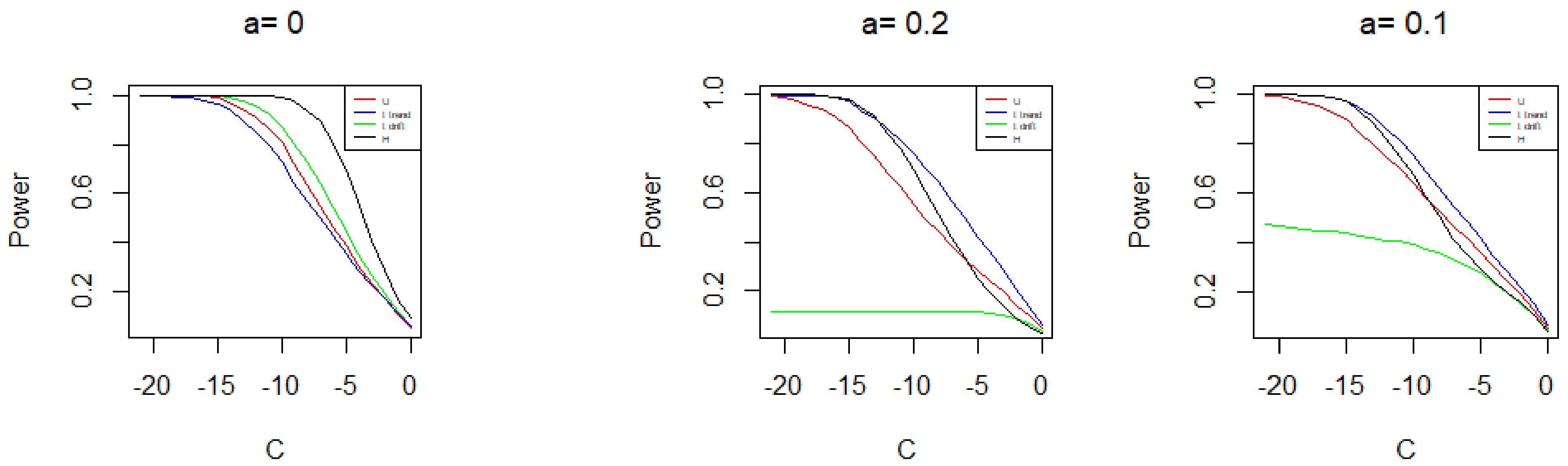
- Then the desired λ_q is estimated by $\lambda_q = \left(\frac{cri^{T_S}}{cri^{drift}}\right)$.
- To apply our method, we have to perform simulations to estimate the λ_q. Therefore, a computationally convenient approach for the practitioner has been presented
- We approximate λ_q by a third order polynomial in $(1-\rho)$ by
- $\lambda_q = \; \widehat{A\_0} + \; A_1(1-\rho) + A_2(1-\rho)^2 + \; A_3(1-\rho)^3$

- To perform the polynomial regression, first we simulate T_S for each $\rho = \rho_0, \rho_1, \dots \dots . \rho_{100}$ and compute $\lambda_{q,i}$ (at significance level q) by the procedure described above. The regression study is defined by
- $\widehat{A} = argmin_{(A_0,A_1,A_2,A_3)} \sum_{i=0}^{100} \big(\widehat{\lambda_{q,i}} - A_0 - A_1(1-\rho_i) -$

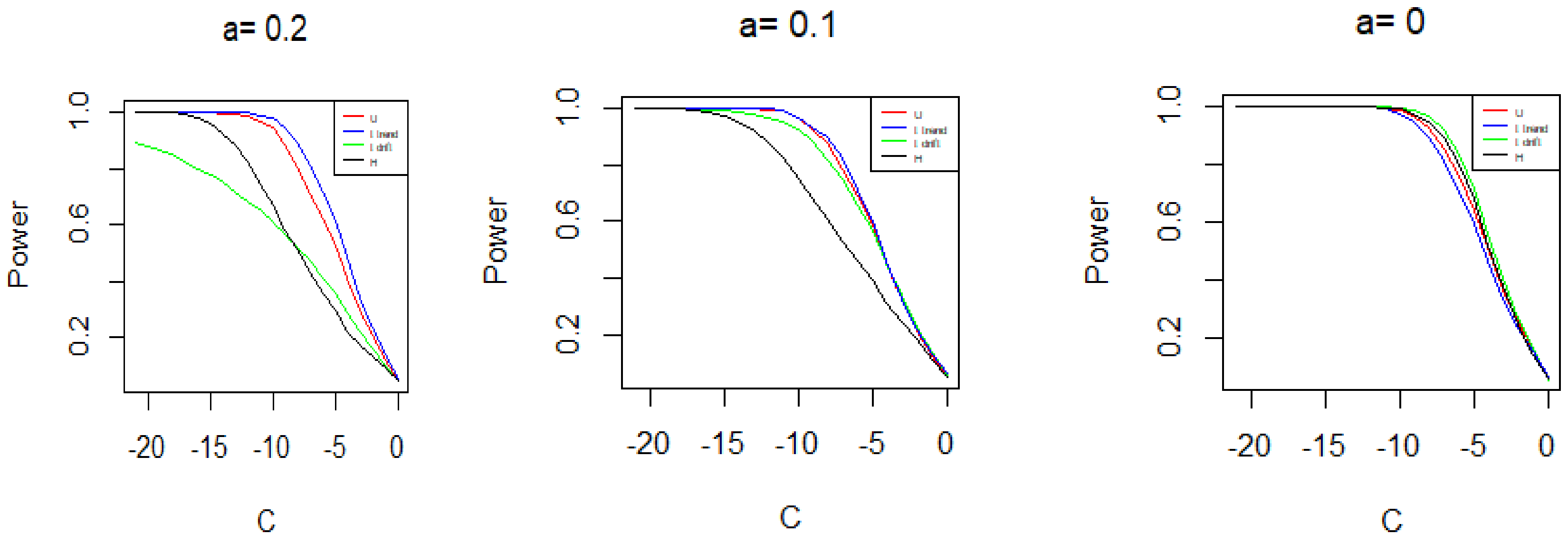
Quantile (%)	A_0	A_1	A_2	A_3
1	1.051	-0.008	0.052	-0.020
5	1.066	0.040	-0.032	0.047
10	1.079	0.062	-0.053	0.077


1st Quantile
Simulated lambda
Polynomial estimated lambda
λ
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
ρ

Small sample study with size 100 and error follows normal distribution

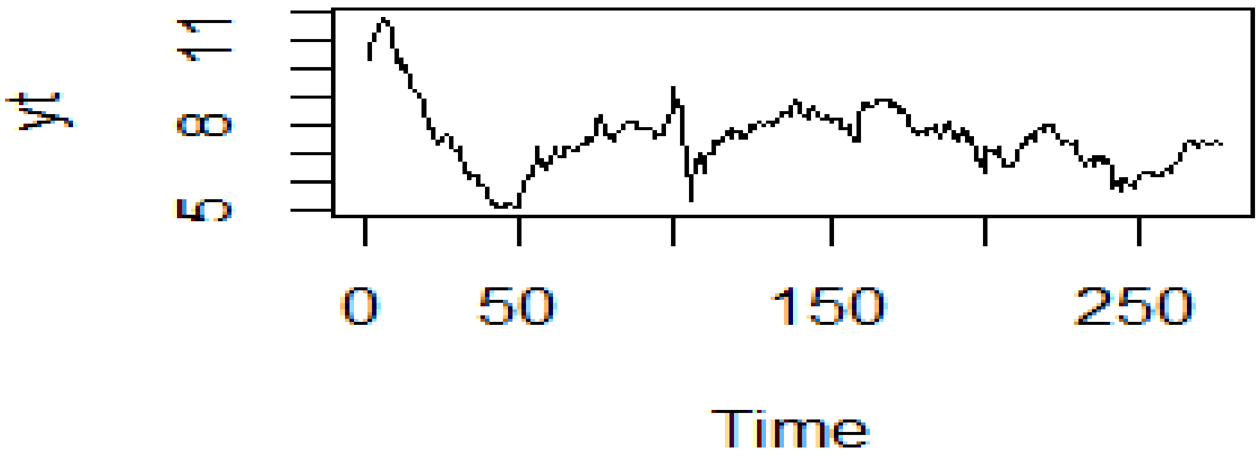


Small sample study with size 100 and error follows t-distribution



Empirical Evidence

- We have applied the proposed method on India's nominal monthly interest rate from January 2000 to March 2023. Data source RBI.



Series	N	Kurtosis	Skewness	Jarque-Bera
Nominal interest rate	276	4.36	-0.70	0.00

Unit root results

Series	U	t^{drift}	t^{trend}	H
Nominal interest rate	Reject	Reject	Not able to reject	Not able to reject

Contribution

- Using Johnson SU distribution I have proposed a unit root test when uncertainty over trend which outperforms other test for asymmetric and heavy tailed innovation.
- The proposed method dominates others especially for the case with asymmetric error process.

Reference

- Ayat, Leila, and Peter Burridge, 2000, Unit root tests in the presence of uncertainty about the non-stochastic trend, *Journal of Econometrics* 95, 71–96.
- Bunzel, Helle, and Timothy J. Vogelsang, 2005, Powerful Trend Function Tests That Are Robust to Strong Serial Correlation, with an Application to the Prebisch: Singer Hypothesis, *Journal of Business & Economic Statistics* 23, 381–394.
- Harvey, David I., Stephen J. Leybourne, and A. M. Robert Taylor, 2009, Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition, *Econometric Theory* 25, 587–636.
- Thompson, Samuel B., 2004, Robust Tests of the Unit Root Hypothesis Should Not Be “Modified,” *Econometric Theory* 20, 360–381.

- Xiao, Zhijie, 2001, Likelihood-Based Inference in Trending Time Series with a Root near Unity, *Econometric Theory* 17, 1082–1112.
- Kar, Tapan, and Malay Bhattacharyya, 2022, Can Choice of Reference Density Improve Power of M-Estimation Based Unit Root Tests?, *Journal of Mathematical Finance* 12, 340–355.

THANK YOU