Economics Bulletin

Volume 35, Issue 4

# Combining linear and nonlinear unit root tests with an application to PPP.

Jeremy Nguyen
*Swinburne Business School*

Jen-je Su
*Griffith University*

## Abstract

This paper suggests combining the linear Dickey-Fuller and nonlinear Kapetanios-Snell-Shin unit root tests in ways that explicitly acknowledge the underlying uncertainty regarding linearity versus nonlinearity. Simulation results show the proposed combination tests perform well. An empirical example is provided, where purchasing power parity is tested for 29 real exchange rates.

**Citation:** Jeremy Nguyen and Jen-je Su, (2015) "Combining linear and nonlinear unit root tests with an application to PPP.", *Economics Bulletin*, Volume 35, Issue 4, pages 2796-2801
**Contact:** Jeremy Nguyen - jdnguyen@swin.edu.au, Jen-je Su - j.su@griffith.edu.au.
**Submitted:** October 19, 2015. **Published:** December 18, 2015.

## 1. Introduction

An unresolved issue in unit-root testing is whether to use tests derived from a linear setting, such as the Dickey-Fuller test (DF, 1979), or to use tests developed under nonlinear frameworks, such as the one developed by Kapetanios, *et al* (KSS, 2003). In practice, for example in the literature on tests of the purchasing power parity (PPP) hypothesis, linear and nonlinear unit root tests are often applied concurrently, at times yielding conflicting results. Under such circumstances, a naive union-of-rejections (UR) rule: "reject the null hypothesis if either test rejects the null hypothesis" is frequently adopted, either explicitly or implicitly. While such a rule is straightforward and powerful: ultimately, only the more powerful of the two tests is actually employed—it may lead to non-trivial oversizing because of the multiplicity of tests. Despite these issues, the literature has remained quiet on the correct method to combine these unit-root tests.

This article seeks to help address this gap. We begin with two popular unit root tests, DF and KSS, and consider two combination procedures. *First*, we utilize the UR rule, addressing the multiple-test issue through size-adjusted, i.e. scaled, critical values. This UR strategy is adapted from Harvey, *et al*'s (HLT, 2009, 2012) handling of uncertainty regarding trend and initial conditions in unit-root tests. *Second*, we use Fisher's (1932) renowned Chi-squared test statistic to combine DF and KSS; the Fisher test is based on aggregation of *p*-values of the underlying individual tests.

Simulations indicate that these combination tests (UR and Fisher) perform well in terms of size and power. In most instances, the combination tests closely track the power of the more powerful of the two unit root tests; occasionally, the combination tests are even more powerful than the two underlying tests. As an empirical example, we also apply the proposed combination tests to the results of Zhou and Kutan (2011) regarding PPP hypothesis for 29 economies.

## 2. Combining DF and KSS

Let $y_t$ (where $t$ = 1,…,T) be the demeaned or de-trended time series of interest. We refer to the case with demeaned series the "Level" case and the case with de-trended series the "Trend" case. The DF test is based on the t-statistic from the OLS regression

$$\Delta y_t = \lambda y_{t-1} + error \tag{1}$$

with $H_0$: $\lambda = 0$ against $H_1$: $\lambda$ <0. The KSS test is based on the following auxiliary regression:

$$\Delta y_t = \delta y_{t-1}^3 + error\,. \tag{2}$$

where the regressor $y_{t-1}^3$ is obtained from a first-order Taylor series expansion of an exponential smooth transition autoregression (ESTAR) model specified in KSS (2003). The KSS test involves testing $H_0$: $\delta = 0$ against $H_1$: $\delta$ <0. We note that DF and KSS have a common unit-root null hypothesis but different alternative hypotheses: stationary linear autoregression (AR) for DF and stationary nonlinear ESTAR for KSS. Both tests possess power against alternatives of stationarity, both linear and nonlinear.

To reach a joint test decision, we suggest two combination tests based on DF and KSS. First, following HLT (2009, 2012), we consider the following UR decision rule:

$$\textbf{\textit{(UR strategy)}} \quad \textit{Reject } H_0 \textit{ if } UR = \min\left\{ DF, \left( \frac{cv_{DF}^{\alpha}}{cv_{KSS}^{\alpha}} \right) KSS \right\} < \tau^{\alpha} cv_{DF}^{\alpha} \equiv cv_{UR}^{\alpha}. \tag{3}$$

where $\alpha$ is a given significance level, $\tau^{\alpha}$ is a scaling constant for size adjustment, and $cv_{DF}^{\alpha}$ and $cv_{KSS}^{\alpha}$ are critical values of *DF* and *KSS*, respectively. Second, we use Fisher's Chi-squared test, aggregating the *p*-values of *DF* and *KSS* ($p_{DF}$ and $p_{KSS}$, respectively) as

$$\textbf{\textit{(Fisher's Chi-squared strategy)}} \quad F_p = -2\left[\ln\left(p_{DF}\right) + \ln\left(p_{KSS}\right)\right]. \tag{4}$$

Both *UR* and $F_p$ have well-defined asymptotic null distributions as *DF* and *KSS* converge jointly. However, unlike the standard Fisher test, which assumes independence across underlying individual tests, we cannot use $\chi^2(4)$ as the null distribution for $F_p$ because *DF* and *KSS* are correlated—we find via simulation that the correlation coefficient between the *p*-values of *DF* and *KSS* is about 0.82 (Level) and 0.74 (Trend).

In Table I, we tabulate the critical values of *UR* and $F_p$ for different sample sizes, obtained via simulation using GAUSS with 100,000 replications. For the UR test, $|cv_{UR}^{\alpha}| > |cv_{DF}^{\alpha}|$ -- i.e., $\tau^{\alpha}$ is larger than 1, implying that the UR test would be oversized if the usual critical values were used. The critical values of $F_p$ are larger than the corresponding critical values of $\chi^2(4)$ reflecting that *DF* and *KSS* are correlated.

**Table I: Critical values for various levels of significance (α)**

(a) Level

| | T=100 | | T=200 | | T=500 | |
|---|---|---|---|---|---|---|
| α | *UR* | $F_p$ | *UR* | $F_p$ | *UR* | $F_p$ |
| 10% | -2.776 | 8.950 | -2.764 | 8.896 | -2.745 | 8.959 |
| 5% | -3.089 | 11.392 | -3.066 | 11.398 | -3.041 | 11.395 |
| 1% | -3.700 | 17.054 | -3.650 | 17.116 | -3.617 | 17.169 |

(b) Trend

| | T=100 | | T=200 | | T=500 | |
|---|---|---|---|---|---|---|
| α | *UR* | $F_p$ | *UR* | $F_p$ | *UR* | $F_p$ |
| 10% | -3.393 | 8.768 | -3.355 | 8.813 | -3.331 | 8.849 |
| 5% | -3.695 | 11.217 | -3.638 | 11.244 | -3.613 | 11.380 |
| 1% | -4.299 | 16.844 | -4.221 | 16.937 | -4.158 | 16.956 |

## 3. Simulation Results

We examine performance of the combination tests through Monte Carlo experiments based on the following DGPs:

$$\text{(AR)} \qquad y_t = \rho y_{t-1} + u_t, \tag{5}$$

$$\text{(ESTAR)} \qquad y_t = \left\{1+\gamma\left[1-\exp\left(-\theta y_{t-1}^2\right)\right]\right\} y_{t-1} + u_t. \tag{6}$$

We set T=200; $\rho$=1, 0.95, 0.9, 0.85 for the AR model; $\gamma$=-0.1, -0.5 and $\theta$=0.01, 0.05, 0.1 for the ESTAR model; and $\{u_t\}_{t=1}^T$ is drawn from i.i.d. N(0,1). Simulations are performed in GAUSS, using 20,000 replications at 5% significance. Relevant critical values are presented in Table I, test results are reported in Table II, for AR alternatives, and Table III, for ESTAR alternatives.

**Table II: Power against stationary linear AR processes ($\alpha$ = 5%, T=200)**

| | Level | | | | Trend | | | |
|---|---|---|---|---|---|---|---|---|
| $\rho$ | *DF* | *KSS* | *UR* | $F_p$ | *DF* | *KSS* | *UR* | $F_p$ |
| 1 | 0.049 | 0.052 | 0.050 | 0.051 | 0.049 | 0.050 | 0.050 | 0.050 |
| 0.95 | 0.320 | 0.263 | 0.299 | 0.315 | 0.187 | 0.151 | 0.174 | 0.170 |
| 0.9 | 0.871 | 0.605 | 0.818 | 0.812 | 0.636 | 0.410 | 0.576 | 0.567 |
| 0.85 | 0.996 | 0.828 | 0.989 | 0.983 | 0.953 | 0.669 | 0.921 | 0.905 |

The first rows of Table II indicate that individual and combination tests are all correctly-sized. The remaining rows show that, as expected, if $\rho$<1, DF is more powerful than KSS and the combination tests have higher power than KSS, but lower than DF. UR and $F_p$ have similar power. The power disadvantages of UR and $F_p$ relative to DF are small in most instances.

**Table III: Power against stationary nonlinear ESTAR processes ($\alpha$ = 5%, T=200)**

| | Level | | | | Trend | | | |
|---|---|---|---|---|---|---|---|---|
| ($\gamma$,$\theta$) | *DF* | *KSS* | *UR* | $F_p$ | *DF* | *KSS* | *UR* | $F_p$ |
| (-0.1,0.01) | 0.142 | 0.158 | 0.153 | 0.156 | 0.102 | 0.103 | 0.101 | 0.108 |
| (-0.1,0.05) | 0.374 | 0.430 | 0.430 | 0.458 | 0.212 | 0.228 | 0.231 | 0.241 |
| (-0.1,0.1) | 0.575 | 0.555 | 0.600 | 0.638 | 0.328 | 0.321 | 0.346 | 0.375 |
| (-0.5,0.01) | 0.591 | 0.726 | 0.735 | 0.775 | 0.342 | 0.445 | 0.433 | 0.456 |
| (-0.5,0.05) | 1.000 | 0.997 | 1.000 | 1.000 | 0.994 | 0.976 | 0.997 | 0.998 |
| (-0.5,0.1) | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 |

Table III presents results under the stationary ESTAR alternative. KSS is generally more powerful than DF, with some exceptions. The combination tests are either nearly as powerful as the more powerful of DF and KSS, or even slightly more powerful. The power advantage gained by the combination tests stems from the fact that *DF* and *KSS* are not perfectly correlated and thus may complement one another in some cases. Of the two combination tests, $F_p$ tends to be more powerful than UR. This is mainly due to the rejection frequencies of DF and of KSS being, in all cases, not far apart.

## 4. Empirical Example

**Table IV: Empirical Results**

| Country | *ADF* | *KSS* | *UR* | $F_p$ |
|---|---|---|---|---|
| **EU Countries** | | | | |
| Austria | -2.73 | -2.43 | -2.73 | 8.98 |
| Belgium | -2.80 | -2.14 | -2.80 | 8.22 |
| Denmark | -2.62 | -1.79 | -2.62 | 6.31 |
| Finland | -2.78 | -2.91 | -2.84 | 11.43* |
| France | -2.97* | -2.73 | -2.97 | 11.49* |
| Germany | -2.94* | -2.45 | -2.94 | 10.10 |
| Greece | -2.35 | -1.99 | -2.35 | 5.75 |
| Ireland | -2.45 | -2.60 | -2.54 | 5.16 |
| Italy | -2.73 | -3.22* | -3.14* | 12.98* |
| Luxembourg | -2.75 | -2.43 | -2.75 | 9.06 |
| Netherlands | -2.97* | -2.48 | -2.97 | 10.37 |
| Portugal | -1.99 | -1.74 | -1.99 | 3.83 |
| Spain | -2.91* | -2.84 | -2.91 | 11.75* |
| Sweden | -2.76 | -3.12* | -3.05 | 12.51* |
| UK | -2.96* | -2.15 | -2.96 | 9.05 |
| **Non-EU Industrial Countries** | | | | |
| Australia | -2.19 | -1.69 | -2.19 | 4.38 |
| Canada | -2.10 | -1.84 | -2.10 | 4.43 |
| Japan | -2.54 | -2.37 | -2.54 | 7.86 |
| New Zealand | -3.26* | -3.20* | -3.26* | 15.61* |
| Norway | -3.10* | -2.46 | -3.10** | 10.98 |
| Switzerland | -2.93* | -2.47 | -2.93 | 10.13 |
| **Asian Countries** | | | | |
| Hong Kong | -2.83 | -3.73** | -3.64* | 16.88* |
| Indonesia | -1.43 | -3.77** | -3.68* | 12.17* |
| Korea | -2.55 | -4.89** | -4.77** | 36.74** |
| Malaysia | -1.76 | -2.44 | -2.38 | 5.49 |
| Philippines | -1.54 | -0.83 | -1.54 | 1.62 |
| Singapore | -3.14* | -3.70** | -3.61* | 18.06** |
| Sri Lanka | -4.67** | -4.10** | -4.67** | 50.66** |
| Thailand | -1.49 | -2.11 | -2.06 | 3.66 |

*Notes*:
(i) Figures with * and ** indicate significance at the 5% and 1% levels, respectively.
(ii) Figures for *ADF* and *KSS* are taken from *ADF1s* and *KSS1s* of Zhou and Kutan (2011, Table III). The US dollar is the numeraire in calculating the real exchange rates.

The two combination tests, UR and $F_p$, are employed to augment the findings of Zhou and Kutan (2011). Using DF and KSS, these authors examined the PPP hypothesis, testing 29 bilateral real exchange rates for stationarity from 1973Q1 to 2009Q2. We calculate the values of *UR* and $F_p$ based on the $ADF1_{\text{sig}}$ and $KSS1_{\text{sig}}$ results reported in Table III of Zhou and Kutan (2011), where "1" stands for the Level case and "sig" refers to the sequential testing procedure to determine the augmented lag order.

Our results are reported in Table IV. At the 5% level, the ADF test rejects the unit root null for 10 real exchange rates, while KSS rejects 8. Of the combination tests, UR results in 8 rejections while $F_p$ rejects 11. Rejections from UR include all but one (Sweden) of the rejections from KSS, plus one (Norway) of the 7 rejections made only by ADF. $F_p$ produces more rejections than any other test, including all of the 8 rejections made by KSS, and 5 of the 10 rejections made by ADF (note that 3 countries are in both the ADF and KSS lists). Interestingly, while neither ADF nor KSS is able to reject the non-stationarity null for Finland's real exchange rate, $F_p$ is able to do so. On the other hand, $F_p$ fails to reject the unit-root null for 5 European countries in the ADF list: Germany, Netherlands, UK, Norway and Switzerland. The 5 relevant real exchange rates may have conformed, over the study period, more closely to the linear AR process than the nonlinear ESTAR process, thus favouring the linear-based ADF.

## 5. Conclusion

In this paper, we suggest two combination unit root tests to formally combine the linear DF and nonlinear KSS tests, in recognition of the fundamental uncertainty regarding whether the underlying process is in fact linear or nonlinear. One of the suggested tests, the UR test, follows the union-of-rejections approach suggested by HLT (2009, 2011) and the other, the Chi-squared $F_p$ test, derives from Fisher (1932). Simulations indicate that both combination tests are potentially useful: both are correct-sized and both possess good power, with $F_p$ and UR often outperforming both DF and KSS under the stationary ESTAR alternative. As for comparisons between the two combination tests, our results suggest that, under the stationary ESTAR alternative, Fisher's Chi-squared test is often more powerful.

## References


Bayer, C. and C. Hanck (2013) "Combining non-cointegration tests" *Journal of Time Series Analysis* **34**, 83-95.

Dickey, D. and W. Fuller (1979) "Distribution of the estimators for autoregressive time series with a unit root" *Journal of the American Statistical Association* **74**, 427-431.

Fisher, R. (1932) *Statistical Methods for Research Workers*, Oliver and Boyd: London..

Harvey, D.I., S.J. Leybourne, and A.M.R. Taylor (2009) "Unit root testing in practice: dealing with uncertainty over the trend and initial condition" *Econometric Theory* **25**, 587-636.

Harvey, D.I., S.J. Leybourne, A.M.R. Taylor (2012) "Testing for unit roots in the presence of uncertainty over both the trend and initial condition" *Journal of Econometrics* **169**, 188-195.

Kapetanios, G., A. Snell and Y. Shin (2003) "Testing for unit root in the nonlinear STAR framework" *Journal of Econometrics* **112**, 359-379.

Zhou, S. and A.M. Kutan (2011) "Is the evidence for PPP reliable? A sustainability examination of the stationarity of real exchange rates" *Journal of Banking and Finance* **35**, 2479-2490.